

# A Strategy for Openness

## Enhancing E-Records Access in New York State Part III-B: Results of Request for Public Comments

*Submitted to:*

*The Honorable David A. Paterson, Governor  
The Honorable Joseph L. Bruno, Temporary President of the Senate  
The Honorable Sheldon Silver, Speaker of the Assembly*



### Executive Co-Sponsors

**Dr. Melodie Mayberry-Stewart**  
New York State Chief Information Officer and Director,  
New York State Office for Technology and Office of the  
New York State Chief Information Officer

**Christine Ward**  
New York State Department of Education  
Assistant Commissioner for Archives & Records  
and New York State Archivist

**Part I – Executive Summary**

*(Separate Document)*

**Part II – Supporting Documentation**

*(Separate Document)*

**Part III - Results of Request for Public Comments**

*Table of Contents*

**Part III-A - REQUEST FOR PUBLIC COMMENTS and SUMMARY OF COMMENTS ..... Page 1**

**Part III-B - PUBLIC COMMENTS FROM INDIVIDUALS.....Page 29**

**Part III-C - PUBLIC COMMENTS FROM GOVERNMENT..... Page 152**

**Part III-D - PUBLIC COMMENTS FROM NON-PROFITS ..... Page 235**

**Part III-E - PUBLIC COMMENTS FROM COMMERCIAL ENTITIES..... Page 296**

## Part III-B

### Public Comments Received: Individual Responses (59 comments received)

*With the exception of formatting and correction of obvious and minor spelling errors, the substantive comments are printed verbatim with any other errors or omissions intact. Salutations and individually identifying information have been REDACTED. Dates and times of comment receipt refer to when the comment arrived in the CIO/OFT e-mailbox created for this study. For any RFPC comments received slightly late, none were so late that they could not be considered for this report.*

#### B. INDIVIDUAL RESPONSES

##### **INDIVIDUAL # 1:** [INDIVIDUAL'S NAME REDACTED]: Tuesday 12/18/2007 8:02 PM

If you keep records in OOXML, better referred to as MSXML, I will be unable to read them. Many experts have commented on the incomplete and contradictory nature of OOXML and how this will prevent any vendor but Microsoft from implementing editors and readers that work. Only people with the newest version of Microsoft OS and readers are able to use OOXML today and this is not going to change.

If your goal is to have a vendor neutral implementation that provides document permanency, you will use ODF or some other ISO standard. ODF was created by a consortium of top notch tech companies and has already has multiple implementations. It is likely that this standard will continue to work for decades.

OOXML, on the other hand, will probably do no better than previous Microsoft formats. If those were worth anything, you would not be having this debate.

I hope that the state of New York adopts sane document format as agencies like the FAA have. OOXML will cost your citizens much and fail to meet any of your purported goals.

##### **INDIVIDUAL # 2:** [INDIVIDUAL'S NAME REDACTED]: Tuesday 12/18/2007 9:04 PM

Please do not trust Microsoft with their OOXML format. While they try to appear open, it is a mean for Microsoft to lock in their Office format. Please listen to IBM and their support of ODF. IBM has been a staunch supporter of open standards in which ANY office applications can use the format.

Let the ODF be the standard because it is the best choice for an unrestricted format that other competing programs will utilize.

##### **INDIVIDUAL # 3:** [INDIVIDUAL'S NAME REDACTED]: Tuesday 12/18/2007 9:14 PM

Open standards need to be endorsed by more than one corporation. ODF is the only reasonable choice.

**INDIVIDUAL # 4:** [INDIVIDUAL'S NAME REDACTED]: *Tuesday 12/18/2007 9:16 PM*

**Question 2.** As far as reading documents, PDF is clearly the best way to distribute electronics records. If you want to guarantee that your data/documents are displayed exactly as you want them, put them in a standard format like PDF. Please do not succumb to the Microsoft "standard" OOXML. Sure, plenty of people use Word documents, but Microsoft's claim that OOXML is the new standard couldn't be further from the truth. OOXML does not follow standard and "open" practices and formats. You need not look any further than the fact that even Microsoft's own products on different platforms not only have difficulty opening these documents, but even when they do open them, they do not always appear the same way.

**Question 3.** Standards, standards, standards. PDF is a standard. Use that when the general public needs only to read your documents. When it comes time to let the general public complete forms, ODF would be a much better choice than OOXML. ODF is a standard used across several platforms and with several applications. The truly open nature of the format allows much better use across platforms and applications.

**Question 4.** Same as before, PDF and ODF are the best document formats to get the job done reliably.

**Question 5.** Same as before, PDF and ODF are the best document formats to get the job done reliably.

**Question 6-9.** Same as above.

**Question 10.** Stick with standard formats rather than proprietary formats.

**Question 11-14.** Same as above, read up on ODF vs OOXML (but go with ODF in the end!)

**INDIVIDUAL # 5:** [INDIVIDUAL'S NAME REDACTED]: *Tuesday 12/18/2007 9:25 PM*

I interact with and do business with New York State government. I would like to contribute the following comment as a business person concerned with government costs and interoperability.

I feel very strongly that the ODT/Oasis/Openoffice office document standards be given special consideration as the standard document format for government interaction. I would like to posit the following reasons:

- 1) There is no vendor preference given when choosing this standard. Other office-suite players have every opportunity to support this standard and, indeed, with existing plugins, they already do.
- 2) It lowers the economic barriers to community interaction with government. There is no "technology tax" for citizens to pay to simply read and write the documentation of their government.
- 3) The "free" toolset available in Openoffice is feature-rich and more than satisfies the current needs of government file interchange. This would lower government costs right off the bat.

4) Long-term storage and compatibility is not as difficult with ODT as it would be with Microsoft Office. The Microsoft Office standard is highly dependent on the version of Microsoft office and, indeed, on the platform for which that program is written. There is a long history of difficulty in opening Microsoft Office documents with different versions of the office suite even on the same platform. OpenOffice/ODT is an open standard and not subject to the vagaries of platform and specific versions of the software.

5) Openoffice/ODT is platform independent, allowing documents to be shared on nearly every commercial and non-commercial computer operating system. No other standard boasts this.

6) There is no other truly open standard for document access. OOXML, a competing standard, from Microsoft purports to be open, but is actually quite closed in its internals (indeed, part of its specification refers to Microsoft Office as the definition of the ruleset for opening and manipulating the documents).

I certainly hope that you choose to use openoffice/odt/oasis document formats for your governmental standard. I can only see upside to this for government budgets, community involvement, and a resulting positive economic impact.

**INDIVIDUAL # 6:** [INDIVIDUAL'S NAME REDACTED]: *Tuesday 12/18/2007 9:26 PM*

I have no vested interest in the issue of which document format the state of new york adopts, but I still will put forth my opinion.

In my short life, on multiple instances, I have run into problems when data (a document, a spreadsheet, a database, or just data records) were stored in a proprietary format and the proprietor stopped supporting this format. In these cases, data were sometimes irretrievable, and at other times retrievable only with great difficulty. For long-term records, such as those a whole state may want to keep, I imagine this would be a major concern.

The Open Document Format seems less likely to produce these headaches, because, in the worst case scenario (if suddenly previous software used to interface with the documents becomes defunct), it will be easier to recover the data if it is in ODF format:

1. the ODF standard is regulated by an independent body, and the format is clearly specified.
2. source code for reading the ODF format is freely available. Even if everyone writing software that uses ODF stops writing this software, it will be easy for someone else to look at the relevant bits of source-code and figure out how to decypher the files.

I don't know what other factors are under consideration, but from my limited experience, it seems that ease and independence when trying to access valuable state data is of utmost importance.

**INDIVIDUAL # 7:** [INDIVIDUAL'S NAME REDACTED]: *Tuesday 12/18/2007 9:46 PM*

**Question 2)** Relevant records and documents should be stored in a publicly-accessible and searchable database. This will facilitate open distribution of information.

**Question 3)** To facilitate interoperability, documents should be stored in open formats that cannot contribute to "vendor lock-in". A vendor-specific format such as Microsoft's .doc, while certainly ubiquitous, is not open and totally readable by any given document editor. Furthermore, the company holding the rights to the closed format has the ability to change the format's specification at any point, perhaps leaving older versions incompatible or otherwise unusable. Open formats such as ODF are not controlled by a for-profit organization and will always remain free and open. This will lead to future stability and guarantees documents stored in such a format will always be available to the public.

**Question 4)** No comment.

**Question 5)** I feel that, for reasons discussed in #3, ODF should be the format of choice for document storage. Not only does it put format control back in the hands of the state organization running the project, it means that users are not required to purchase additional software in order to view said files. Documents stored thusly are truly free and public, as the law declares they should be.

**Question 6)** No comment.

**Question 7)** The state should keep its records as long as is reasonable. Archived content should be available free to the public, and in ODF format for openness.

**Question 8)** No comment.

**Question 9)** The state should be able to save money by using free word processing utilities (such as OpenOffice) to maintain its collection of records.

Question 10) For very highly specialized formats, there might not be an adequate open alternative. In these cases, the document in question should be available in several of the most popular formats, including open formats if possible.

**Question 11)** As discussed in #9, it is very likely that the state will be able to save money on the cost of software by using free word processing software to maintain its database of information.

**Question 12)** No comment.

**Question 13)** No comment.

**Question 14)** This is New York's opportunity to leap ahead and set a shining example to other states. Open document formats can make archival projects such as this one less expensive, and more free and open to the public. Please also bear in mind that Microsoft Word (if it's necessary to use it) can still view and modify open formats through the use of add-ons.

Thank you kindly for your time.

**INDIVIDUAL # 8:** [INDIVIDUAL'S NAME REDACTED]: *Tuesday 12/18/2007 10:03 PM*

Any owner of data that is stored for a long period of time faces problems retrieving the data later unless the data is stored in a simple, open format.

Storing them in a format supported by a single vendor will ultimately tie the owner of that data to the vendor. If the vendor goes out of business or no longer supports the closed format, then the owner of the data is faced with the expensive process paying to have the data brought up to a new format.

If the data is stored in an open format that is clearly defined, there will be an easier migration path in the event that the format is no longer supported, to extract the data. As long as it is supported by at least one provider, there will be a way to extract it.

Choose a completely open format that is not owned or backed by a single entity.

**INDIVIDUAL # 9:** [INDIVIDUAL'S NAME REDACTED]: *Tuesday 12/18/2007 10:17 PM*

I wish to reply to an open invitation to public comments regarding the State of New York's decision of electronic data storage, and which format should be chosen. I highly recommend that OpenDocument (commonly known as ODF) is chosen rather than Microsoft's Office Open XML (OOXML) format. One strong argument for the ODF format is already outlined in your RFPC invitation:

*"The director shall study how electronic documents and the mechanisms and processes for obtaining access to and reading electronic data can be created, maintained, exchanged, and preserved by the state in a manner that encourages appropriate government control, access, choice, interoperability, and vendor neutrality."*

ODF is an approved ISO standard. It is very stable and very well documented, and contains no risk for any vendor of Office suite software to implement. This vastly provides an advantage for interoperability and vendor neutrality. On the other hand, OOXML is a very poor standard. The documentation for this format is far too incomplete for any software company other than Microsoft to implement, thus severely harming the advantage of government control, access, choice, interoperability and vendor neutrality. I refer to this statement in the OOXML specification:

*"To faithfully replicate this behavior, applications must imitate the behavior of that application, which involves many possible behaviors and cannot be faithfully placed into narrative for this Office Open XML Standard. If applications wish to match this behavior, they must utilize and duplicate the output of those applications."*

In other words, much of the behavior of this standard is kept secretive by only a few software vendors. Vendor neutrality is non-existent with the OOXML format, which is why ODF is such an excellent choice for storage of electronic data, records or documents. OOXML is not a stable format as it can be changed at any time by Microsoft, who yields full control over the format. In addition, OOXML is not an ISO standard whereas ODF is.

To keep electronic data, documents and records accessible to everyone, and not just users of Microsoft's technologies, the State of New York must strongly consider adopting ODF as their format of choice, as this will ensure full accessibility to everybody not just today, but in the future, since ODF is fully documented and can be implemented by all Office software vendors at any time without risk of incomplete implementation, or patent infringement lawsuits by Microsoft.

**INDIVIDUAL # 10:** [INDIVIDUAL'S NAME REDACTED]: *Tuesday 12/18/2007 10:29 PM*

I would like to state my support for the ODF format it is open and people shouldn't be forced to buy expensive software to see gov. documents. I am a new (as of august) NYC resident. Working as a software engineer.

**INDIVIDUAL # 11:** [INDIVIDUAL'S NAME REDACTED]: *Tuesday 12/18/2007 10:31 PM*

I am very much opposed to the use of OOXML, since it is officially not an open standard. And while ODF is not great, it is a lot better and more organized than OOXML, and it is a fully open standard. And ODF is already being used by many many people as a standard, and there are already good plugins (<http://odf-converter.sourceforge.net/> ) for ODF in Microsoft Word if people want to continue to use that. Also, ODF is available for use in many other word processing programs, including KWord, which I often use, and Abiword, which is another open source word processor that I have used in the past. And of course Sun's own Open Office and Star Office. Meanwhile the only thing compatible with OOXML is Microsoft Word, and even for old editions you already need a plugin.

**INDIVIDUAL # 12:** [INDIVIDUAL'S NAME REDACTED]: *Tuesday 12/18/2007 10:58 PM*

Adopting OOXML as the open document standard within NY would send the wrong message to the nation and world and would reinforce Microsoft's stranglehold on the public and private sectors. To gain "support" for OOXML, Microsoft has engaged in many unethical and biased activities including monetary bribes and practices that edged out the room for competition.

Please take a stand and choose a truly open standard for New York. Take a stand against the fishiness of Microsoft's magically huge interest in open-source and openness all of a sudden, and their pressure to fast-track the world onto OOXML. Keep Microsoft out of the things that are truly free and open right now.

ODF is an accepted and widely-used standard already, gaining support rapidly and is supported by the largest competitor to Microsoft's office suite (Openoffice.org). Microsoft is trying to strangle the open-source community by embracing it tightly and scaring everyone away.

Please do the right thing and support ODF, don't let Microsoft continue to push the public and private sectors around. Openness is the key to a government that can be trusted, not by bringing in Microsoft and letting them run everything. Not by using black-box voting machines which nobody can trust (Diebold).

Do the right thing, support transparency, openness, and open-source. Don't let Microsoft continue to push the world around.

**INDIVIDUAL # 13:** [INDIVIDUAL'S NAME REDACTED]: *Tuesday 12/18/2007 10:59 PM*

Though I do not live in New York, I felt that I had to respond to this request for opinions. First a few facts.

ODF was conceived and developed to be an open and vendor-neutral standard for documents. Its stated purpose is to avoid vendor lock-on, and provide a document standard that can be freely used by any software company. This was done to make documents, especially archived

documents, and documents transferred across the Internet (and corporate networks) readable to anyone, using their choice of software. In other words ODF is about giving all users more choices.

ODF is also an ISO standard. It has been implemented in a number of free and open word processors, and the OpenOffice.org Office Suite.

OOXML on the other hand was created by Microsoft to try to maintain its vendor lock-in, and to protect its sales of MS Office products. OOXML is NOT an ISO standard at this time, although much evidence has come to light that Microsoft had tried to corrupt the ISO standards process to make OOXML an ISO standard. To my knowledge, OOXML has not been fully implemented in any software. On the technical side, the 6000 pages of the OOXML standard are so complex, and contain so many references to proprietary code, that no one could implement it but Microsoft...which is Microsoft's intent. They want the image that OOXML is an "open" file format, but in reality, it is just a ploy to try to prevent ODF from becoming the preferred document standard. Those who have studied the 6000 page OOXML proposal have pointed out many flaws.

I believe that the best choice for archiving and public viewing of important documents is ODF. The goal of any organization should be to make sure that its documents are readable by as many people as need to read them, and for the longest possible time. Only ODF can make this happen. In my opinion OOXML is a dead end. It can only lead to more lock-in to Microsoft Office, which is not a good thing, as Microsoft only cares about their own profits, and thus their file formats have changed with every new version of MS Office.

Thank You for taking the time to read and consider this comment.

**INDIVIDUAL # 14:** [INDIVIDUAL'S NAME REDACTED]: *Tuesday 12/18/2007 11:05 PM*

I read in Slashdot that *"In August of 2007, the State of New York passed legislation requiring its CIO, Melodie Mayberry-Stewart, to gather information on the advantages and disadvantages of adopting either ODF or OOXML as a document standard, and to report her findings by 15 January 2007."*

I don't work with ODF or OOXML professionally, so I can't comment on the specific technical merits of either. However, I do work with Open Office, which shares roots with ODF, and I work with products from Microsoft (or try hard to). OOXML's roots are in Microsoft.

I also have observed the conduct of people who produce Open Standards...and of the people who run Microsoft. Open Standards make the lives of IT professionals easier. (I work with XML, HL7, and EDI.) Microsoft products are designed to lock users (around the world) into eternal use of Microsoft products.

The State's CIO should consider whether the State embraces those who work to improve communication, or those who have been convicted of numerous types of unlawful business conduct, and have not wavered from it.

[IDENTIFYING INFORMATION REDACTED]

I noticed that there are also wikipedia articles about ODF and OOXML. The two articles can be used for a side-by-side comparison.

**INDIVIDUAL # 15:** [INDIVIDUAL'S NAME REDACTED]: *Tuesday 12/18/2007 11:06 PM*

**Question 1.** [IDENTIFYING INFORMATION REDACTED] My research falls in the discipline of Information Systems and I believe I have a good understanding of several of the issues involved in this decision. As requested, in what follows I will identify the questions to which I am replying.

**Question 2.** Concerning "access for being used for the day-to-day purpose" and other "access during their active business use for ancillary purposes": the State of New York (NYS) should first and foremost ensure that both individuals as well as organizations will not be forced to buy or license software products/services from specific vendors to be able to conduct business with (cf question 3), or gain access for ancillary purposes to records kept by, NYS. More specifically, NYS should adopt mechanisms and processes which require the use of open, fully-documented, vendor-neutral and patent-unencumbered formats. Only in this way can NYS avoid the creation of unnecessary barriers to public access (in both senses) to its electronic records.

Concerning "accessibility for their historical and research value after having been preserved": digital computing history, in spite of only covering a few decades, is littered with sad stories of data being unrecoverable due to file formats becoming unreadable as, e.g., particular software companies exit the market and their products become unsupported and then turn obsolete. As a public institution, it is of vital importance that the records of NYS resist such events. Therefore, mechanisms should be in place that require the use and preservation of records in open, fully-documented, vendor-neutral and patent-unencumbered formats.

**Question 3.** Here I write based on first-hand personal experience and provide my case as anecdotal evidence. As someone who is not a user of Microsoft products, I often have difficulty in sharing files with colleagues and institutions just because Microsoft Office file formats are so common. NYS, in order to encourage interoperability and data sharing with both citizens as well as other institutions, should definitely ensure that its electronic records are created, stored and shared in an open, fully-documented, vendor-neutral and patent-unencumbered format. Only in this way can interoperability be preserved. A special request: please be aware of claims of interoperability by particular vendors who are promoting their own standard. The litmus test for interoperability boils down to four adjectives: "open", "fully-documented", "vendor-neutral" and "patent-unencumbered". In terms of "office productivity" file formats, the Open Document Format (ODF) is the only format which truly satisfies these four criteria.

**Question 5.** Please see my answer to question 3. Both "encouraging choice" as well as "vendor neutrality" rely on precisely the same factors as interoperability (the topic of the third question): records must be created, kept and transmitted in a format which all applications can easily read and write. By standardizing on formats which a particular vendor has an advantage in implementing, NYS would be effectively introducing an (additional) important distortion in the software market and conditioning the choices available to both individuals as well as organizations. For that to be reason, and in order to encourage choice and vendor neutrality, it is important that NYS puts in place mechanisms that require the use of open, fully-documented, vendor-neutral and patent-unencumbered formats for its electronic records.

**Question 7.** (I will restrict my answer to the issue of the formats in which electronic records are to be kept.) As mentioned in the second paragraph of my answer to the second question, it is for historical and research reasons very important that electronic records be kept in formats which will be readable in the future. The best chance we currently have at increasing the chances of future

access to our electronic records involves (in addition to adequate hardware and software backup and archival solutions) requiring that all relevant electronic records be kept in open, fully-documented, vendor-neutral and patent-unencumbered formats. By doing so, NYS can ensure that the readability and access (in the third sense specified in your RFPC) will not be dependent on the future availability of a particular software product.

## Part II

**Question C. 4.** As mentioned in my answer to the third question in Part I, I have extensive first-hand experience in trying to interact and collaborate with both individuals as well as organizations which require files to be read, written and exchanged in the formats supported by a specific vendor. That vendor is, in the vast majority of cases, Microsoft and the format in question one of those supported by its Microsoft Office productivity suite.

In matters of government oversight and general constitutional importance such as FOIL, it is inconceivable that any barrier to truly universal and immediate access be put in place once a document has been made public through a FOIL request. Adoption of a format such as Office Open XML (OOXML), which does not fulfill the four key criteria specified above (those of being open, fully-documented, vendor-neutral and patent-unencumbered), would constitute such an inadmissible barrier. In an age when both the traditional as well as the newer online citizen media reacts to event with unprecedented speed, and in which the importance of a news story is beginning to be gauged by the intensity of the reaction of the blogosphere to the documents which underlie it (which are often released after FOIL requests), it would not be fair that the access -- and the timeliness of that access -- of individuals and organizations to FOIL documents be conditional on such apparently trivial matters as the office productivity suite one has installed on a computer. Therefore, this constitutes strong grounds for the adoption of the Open Document Format (ODF) by the State of New York.

**Question E. 6** Although at first this seems to be an acceptable definition of "interoperability", it is important that a clear reference be made to how stringent the State of New York must be when determining what it means to "be used together". One often sees claims of interoperability between two software systems based on the ability of system X to support basic, or "core", features of documents produced by system Y and vice-versa. Such is the case with support for earlier (pre-OOXML) Microsoft Office formats: multiple vendors "supported" importing and exporting documents in those formats, but (due to deficient documentation of those formats) were only able to support the core features, failing to preserve apparently unimportant features such as "formatting". The apparent unimportance of "minor" features is terribly misleading. E.g., I often hear colleagues say "I would gladly use [an alternative to MS Office] if only it wouldn't mess up the formatting of my slides". Those "minor" interoperability issues (between products which, according to most reading of your present definition, would be considered to be "us[able] together without modification or development of custom interfaces") are one of the primary reasons why Microsoft Office has got such a huge hold on the market for office productivity suites in spite of there having been numerous, some of them highly innovative, competitors.

In summary, I ask that you clearly specify that for two products/systems to be "usable together" and therefore considered interoperable, there must be zero tolerance for "minor" interoperability problems. E.g., to "mess up the formatting" in any way is a clear indication (in a market where software is used to generate documents to be presented to actual or potential business partners or other business-relevant third parties — a situation in which most would agree that concern for careful presentation is of great importance) that two systems are not interoperable.

Furthermore, I maintain that true interoperability can only be achieved if a complete implementation of a standard is publicly available \*in source code format\*. Only when the "nitty-gritty" of the internals of a complex system are exposed through making the source code of a reference implementation public do other vendors have a fair chance at implementing a fully interoperable (as defined above) product or system.

**Question E. 7** This is a very good definition, but it needs an important addition. At first sight, it covers all three criteria I have previously emphasized in my answer to Part I: to be open, a format/standard must be fully-documented, vendor-neutral and patent-unencumbered.

However, and in line with my answer to question E. 6 in this part, it is important that your definition specify that a standard will only be considered "fully documented" if a complete, widely-used reference implementation is made available in source code form. This is the case with multiple, much simpler technologies (see, e.g., the JPEG and Ogg Vorbis file formats, both of them widely used on the Internet by systems using a variety of implementations). A full specification in text form will fail to ensure true openness (just as it fails to ensure full compatibility/interoperability — see previous answer) unless the standard is "documented in practice" by having a leading vendor provide its source code. If such source code is not made available, it will always be possible (and relatively easy) for the leading vendor associated with that format to start adding/modifying features which will deviate from those specified in the standard and thus gain an edge over competing implementations.

Second, I submit to your consideration just how "open to all interested parties" the decision-making process of a standard so intimately tied to a particular vendor (such as is the case of Microsoft and OOXML) will be. Certainly, a non-commercial organization (both ECMA as well as ISO, in this case) will officially maintain the standard, but one cannot expect the decision-making process to be veritably open to all stake-holders in the sense that their input will be considered purely on its technical merits. This raises huge questions over the evolution of a standard and the very meaning of "openness": is a standard "open" if the public cannot be confident that it will remain open into the foreseeable future? If the decision-making process can be expected to be so biased towards a specific vendor who is in the business of producing proprietary solutions, can we truly call that standard open?

**Question E. 10** I find it disturbing that this question should be asked, since its presence in the survey suggests a lack of concern for much more fundamental issues (please refer to my answers to questions 2,3,5 and 7 in Part I). The State of New York simply should not be placing constraints on the choices made by the public by favoring a vendor-specific (or "vendor-favorable") standard when a truly interoperable, open standard exists.

I am confident that "failing to provide the NYS workforce with the capability of using ODF" would not cause "NYS interoperability problems": what would happen would be that all parties interested in interacting and conducting business with the NYS would effectively have their choices of office productivity suites constrained due to a NYS policy. Given the vital business importance of many such transactions for a number of individuals and organizations in this state, as well as the purely mandatory nature of many such interactions (e.g., the submission of tax documents), what would happen would be that non-NYS parties would have to use a Microsoft product simply to conduct business with the state. This scenario, in the presence of a truly open standard such as Open Document Format, is not acceptable.

**INDIVIDUAL # 16:** [INDIVIDUAL'S NAME REDACTED]: *Tuesday 12/18/2007 11:08 PM*

I understand that New York is considering file format standards such as ODF and OOXML. I believe that it is very important that New York chooses a standard that is truly open and free and that standard is the ODF.

I work with software, and I must say that it is a joy to be able to work with and have access to ODF files. First of all, the ODF standard allows anyone to create, view, or modify ODF files. Using standard and open source tools I can take an ODF file and uncompress it and then go into the directory that is created. I can look at the files using a simple text editor. If I want to make changes I can do so with a simple text editor, or perhaps I want to write a little program to modify the contents of a file. Everything is open and simple with ODF, so anything is possible.

Take a look at an actual ODF file, and also an OOXML file. Look at how much larger the OOXML file will be.

If New York chooses the free and open ODF format then New York will own and control its data, both now and in the future. I expect that by selecting ODF New York would save a very considerable amount of money and see increases in productivity.

**INDIVIDUAL # 17:** [INDIVIDUAL'S NAME REDACTED]: *Tuesday 12/18/2007 11:09 PM*

I am writing as a professor at [REDACTED] to offer some insight on the OOXML/ODF debate. During my doctoral work, I produced scores of educational materials in various digital formats. Some of these materials were in-house, only meant for use at one institution, while others were federally or privately funded by organizations including the U.S. Department of [REDACTED].

For each project, we constantly re-evaluated the technologies we used. Over the course of five years, we began to notice that the more a project employed open technologies, the longer it remained accessible and usable.

For example, one project was a CD-ROM meant to teach basic French to volunteers at for 2002 Winter Olympic Games. Funded by L'Alliance Francaise and L'Agence Internationale de la Francophonie, this project was built on Microsoft's Internet Explorer 5, which was the most ubiquitous content delivery platform at the time.

It was less than eighteen months before changes to Internet Explorer rendered that material unusable.

Contrast that result with the Arabic instructional materials we created for the National Middle East Language Resource Center in 2003. Those activities were built to the Sharable Content Object reference Model (SCORM) and referenced other fully open XML specifications, such as QTI. Further, their functionality was defined in standard ECMAScript, with no browser-specific extensions.

This year, many of those materials were seamlessly incorporated into the Arabic Without Walls program for the University of California Consortium for Language Learning and Teaching.

The difference in longevity between these two projects can be traced to the openness of their underlying technologies.

Although both OOXML and ODF consider themselves "standards," one must evaluate their "openness" independent of standardization claims. This issue may be distilled to one question: Is there a single for-profit corporation that owns/controls the "standard"? If the answer is yes, then the potential for the standard to slowly shift with the business interests of that corporation is much higher. While it is possible for a truly open standard to be co-opted by a powerful stakeholder, it is far less likely than if a single interest already owns the standard.

I would strongly encourage you to make considerations of this type of openness paramount as you evaluate these competing standards. If you would like more information, I can best be reached via email until January 10th, when I will return to my campus office.

**INDIVIDUAL # 18:** [INDIVIDUAL'S NAME REDACTED]: *Tuesday 12/18/2007 11:14 PM*

I am a resident of New York.

I favor the ODF standard and not the OOXML one.

- a. The ODF standard is open-source and developed by a broad community effort. The OOXML standard is essentially proprietary and developed by a closed corporation. For open, public documents, an open-source solution is more congruent.
- b. Except for those beholden or networked with Microsoft, all commentators I've read state that the OOXML standard has serious flaws. If that is the case, correction of the flaws is dependent on the acquiescence of the closed corporation.

**INDIVIDUAL # 19:** [INDIVIDUAL'S NAME REDACTED]: *Tuesday 12/18/2007 11:41 PM*

To guarantee the broadest access to data as well as the most archival document format, I am, as a New York resident, writing to request, *\*emphatically\**, that you support Open Document Format.

Furthermore, as an NY state taxpayer, I hope the state will adopt and support free software and the standards that support free software whenever possible.

Private interests, specifically Micro\$oft, are free to code their products to read and write ODF data. On the other hand, if the OOXML standard is accepted many users, such as myself, will be excluded from accessing to the standard format documents and data.

I trust you will do the right thing. Thank you for your public service.

**INDIVIDUAL # 20:** [INDIVIDUAL'S NAME REDACTED]: *Tuesday 12/18/2007 11:54 PM*

**Question 1.** Over the past fifteen years I have been involved in the development of many Internet standards including HTTP, HTML, XKMS, SAML, WS-Security and currently co-chair a [TECHNOLOGY ENTITY - REDACTED].

**Question 2.** What mechanisms and processes should the State of New York establish for accessing and reading its electronic records in order to encourage public access to those records?

Government agencies should let agencies choose the document formats that best fit their needs. Government attempts at top-down dictation of 'standards' has been counterproductive in the past and is likely to be so in this particular case. During the 1990s the government of Germany received much praise for its 'visionary' policy of promoting provision of ISDN services to the home. The US government made a considerable investment in OSI networking infrastructure which was similarly futile. Even today participation in the US government Federal Bridge PKI officially requires provision of an X.500 directory.

In this particular case an attempt to make a choice of document format is futile as is the request for public input. The choice of desktop software should not be made by central IT departments, nor should it be made by random members of the public. The choice of software should be made by the people who are going to use it at their desks every day. They will choose the tools that help them do their job best.

The state does have certain interests in ensuring interoperability and access but these can be met by any mainstream software package for reasons I will develop further in the answers below. The state should not therefore mandate the use of software that supports one particular standard, whether OOXML, ODF or anything else. Instead the state should set out the essential criteria that are necessary for interoperability.

Question 3. What mechanisms and processes should the State of New York establish for accessing and reading its electronic records to encourage interoperability and data sharing with citizens, business partners and other jurisdictions?

There are two types of access that are critical, the first is human accessibility, the second is machine accessibility. Both are important if the full potential of the Web is to be realized.

Human accessibility has been a major concern in the design of HTML since 1994 when Uri Rabbinski and others from the accessibility community met with Dave Ragget, myself and others at the first World Wide Web Conference. Despite these capabilities however, HTML is better regarded as a format that allows accessibility by blind, partially sighted and other disabled persons to be supported rather than achieved and HTML cannot currently be considered an all purpose document language as implementations lack features necessary for use in authoring books, reports, etc.

Despite the fact that Word is a proprietary format, the market presence of Word has made it in effect mandatory for all other document processing software to interoperate. Even before the Word formats were made public their content was widely known through reverse engineering. As a practical matter there is no human accessibility problem with the Word document format.

Machine access is a different issue. In particular current document presentation technology geared towards the needs of human readers and editors is poorly suited to many machine processing needs. At one end of the scale are presentation-only formats such as PDF which allow a document to be printed on the screen but unless special preparation steps are taken require considerable reverse engineering to be applied if the document produced is the going to be edited. Document formats such as Word, HTML and such occupy a middle ground. The underlying data used to produce documents can usually be reliably recovered using 'screen scraping' techniques but each set of documents requires separate treatment. At the far end of the scale are formats such as XML datasets and semantic Web technologies such as RDF which are

designed to allow automated interchange of data between machines without first rendering it to human readable format and back.

**Question 4. What mechanisms and processes should the State of New York implement to encourage appropriate government control of its electronic records?**

Rather more important than the choice of document format is the decision to apply security to a document or not. If parties are to act on government provided information it must be trustworthy and not merely trusted.

The mere fact that data was obtained from a government Website does not provide any assurance of authenticity. The Internet infrastructure does not provide security as a default condition and even if it did reliance on transport level security such as SSL is a poor example for government to set. The ability to authenticate a document using public key cryptography should be considered a prime concern. Government should set up processes to ensure that only official documents that are properly authorized receive the state signature.

**Question 5. What mechanisms and processes should the State of New York consider for encouraging choice and vendor neutrality when creating, maintaining, exchanging and preserving its electronic records?**

None.

The sole concern of the state should be to obtain the software that best meets its needs. Vendor neutrality is not a reachable goal in the current market for document preparation software. To insist on either ODF or OOXML is to effectively mandate choice of one vendor over the other.

**Question 6. Are there mechanisms and processes the State of New York should establish that are specific to the management of its electronic records in its various life cycle stages (creation, maintenance, exchange, preservation and disposal)?**

None that are relevant to the choice of ODF vs OOXML.

The human mind has proven far more able to decipher long forgotten scripts than to devise scripts that defeat comprehension. The fact of the Ultra decrypts of German ENIGMA intelligence during World War II, the deciphering of Egyptian hieroglyphs, the Myan syllabary and Linear B all demonstrate that if there is value in deciphering a long forgotten means of communication it will be found.

Archivists concerns over long term support for obsolete document formats are generally futile. While there can be no guarantees that any aspect of human society will endure in perpetuity, we can safely ignore the risk that human kind will loose the ability to comprehend any variety of HTML, Word, or any other widely deployed document format.

The only constraint that need be required is that whatever document formats are used have a sufficiently significant market position to guarantee support by the major document lifecycle management products.

**Question 7. How should the State address the long term preservation of its electronic records? What should the State consider regarding public access to such archived content?**

Long term archiving is a complex consideration that requires separate concern for preservation of the data 'bits', preservation of the ability to interpret them and preservation of the document provenance. Some state data may also be subject to confidentiality considerations and in some cases classification as secrets.

Having previously disposed of the question of interpretation, let us consider the preservation of bits: How do books survive? The answer is different depending on whether we are considering centuries or millenia. Books survive over centuries or more through redundancy, If a book is published in an edition of a thousand copies or more the probability of survival for the individual book need not be very large for the probability of survival of at least one book to be almost certain. A 1% chance of survival for the individual copy means that it is 99.995% certain that at least one copy will survive. Books survive over longer periods through copying. Paper degrades over time, our knowledge of classical times comes through us through multi-generational copies.

We need to apply the same principles to storage of data: replication of the storage media to ensure redundancy and intermittent verification and copying to new media to ensure that data is not lost due to systematic failure of the storage media.

Document provenance and confidentiality protection require the application of standard cryptographic techniques.

**Question 8. What changes, if any, should be made to the government records management provisions in New York Statutes?**

None, the technology in this area is developing faster than the statutes will keep track of. Let the executive determine policy in this area and do not try to manage change. Any attempt to legislate in this area is likely to be hijacked by special interests whose concern is making the next sales quarter rather than finding the best technology for the state.

**Question 9. What constraints and benefits should the State of New York consider regarding the costs of implementing a comprehensive plan for managing its electronic records?**

A: Small is beautiful.

B: The number of consultants required to implement a design is proportional to the square of the number engaged to design it.

I do not fully understand why large government IT projects have such a dismal track record but I have witnessed some of causes of failure. Excessive reliance on consultants is a frequent cause, I refuse to do consulting for state governments after my last experience six years ago when every single person at the 20 strong kick-off meeting turned out to be a consultant. The contract leads then decided to adopt the ever-profitable waterfall development model that produces many billable hours but little else.

Every enterprise and corporation is going to need to deploy a comprehensive document archive and retention system over the next decade or so. If the state is unable to find a Commercial Off The Shelf infrastructure that meets its needs with little or no customization today it should wait until one becomes available.

I have a hypothesis that I would like some government contracting office to test: Is it more efficient to break down large IT projects into smaller units that are specified and bid independently than the traditional approach of giving a huge blank check to one large consulting firm?

The argument over whether OOXML or ODF should be adopted completely misses the point that the real value of Open Source is not to be found in the code itself but the development and deployment approach. For example the UK National Health Service is currently experiencing a world class mega-IT project fiasco. Under the contracts agreed by the NHS the doctors at hospitals that have purchased software for use in the scheme are not allowed to report on the performance of the software, or their experiences with it to other health authorities. While this contract was probably sold to the ministers as a prudent means of avoiding political embarrassment the real motive for the vendors was more likely to suppress reports of how poorly their products performed. The ministers were not saved embarrassment as the stories all leak to Private Eye regardless.

The cost of office software is such a small part of the total cost of ownership that the choice to buy proprietary software or adopt open source is unlikely to yield savings.

The much more significant question is the status of software developed by or for the State of New York. Will New York follow traditional models of IT contracting that lead to the familiar problems of contractor-lock in, undocumented source code, etc. that lead to many billable hours for little real return?

Far more important than the status of code the state consumes is the status of the code the state produces. Can New York State develop a model for IT contracting that allows it to take advantage of the network effects that open source has provided in other areas?

**Question 10. What should the State of New York consider regarding the management of highly specialized data formats such as CAD, digital imaging, Geographic Information Systems and multimedia?**

Take a small percentage of the amount of money you spend in such areas and apply it to pure research into ways in which the state might leverage the emerging technologies of the Semantic Web to advantage in the future. In the near term all such data formats tend to be proprietary and the opportunities to exchange data with other programs tends to be limited.

Rendering data sets in XML provides a small step forward in enabling data interchange but the mere fact that a data set has angle brackets does not make it possible to exchange the data with other programs.

**Question 11. What constraints and benefits should the State of New York consider regarding potential savings or additional costs associated with the management of defined electronic record formats?**

Expect all current data formats to become obsolete over the next fifteen years or so.

The ODF vs OOXML debate is particularly futile as the whole concept of the office suite is itself obsolete, a hangover from the days when PCs were so slow and limited to run more than one function at once so the Word Processor, Database and Number crunching features were all

independent. The basic approach of the spreadsheet has changed little since the days of 8 bit processors and Visicalc.

Mathematica shows how the same functions could be integrated to produce a package that is both simpler and more powerful than current office suites.

**Question 12. 13.** Omitted.

**Question 14. What else should the State of New York consider about this subject?**

Consider security explicitly, not as an afterthought. do not think of document security in terms of confidentiality alone, consider the need for authenticity and demonstration of provenance as the primary concern.

At the present time the financial infrastructure of the USA and every other industrialized countries falls far short of the state of the art. The standard means of access to online banking sites is a username and static password. Although Chip and PIN smartcard security has been added to card payment systems in Europe, deployment has not yet occurred in the US. Yet another concern is so called 'identity theft', better known as obtaining loans under fraudulent pretenses.

Much of the data currently held by the state poses a potential risk in that it may be employed by Internet criminals to further crime. In the short term the state must take great care to ensure that its actions in making documentary data available do not further undermine the security of the financial services infrastructure. In particular the state should avoid collecting personal data that might compromise such systems and avoid transmission via insecure transmission media. The recent scandal in the UK where a senior civil servant was forced to resign after a data disclosure involving 20 million odd taxpayers (a third of the population) is an example of an all too regular occurrence.

The state must however make it clear that in the medium term it is the responsibility of the financial services industry to set their house in order and develop a financial services infrastructure that does not rely on static passwords, or the secrecy of public data such as mother's maiden name. I develop a comprehensive plan to do this in my book [NAME REDACTED] to be published by [REDACTED].

**INDIVIDUAL # 21:** [INDIVIDUAL'S NAME REDACTED]: [Wednesday 12/19/2007 12:18 AM](#)

I do not live in the state of New York but I figure I'd give my feedback, as I have extensive experience with both ODF and OOXML file formats.

The bottom line: ODF is far superior to OOXML.

ODF is a true open/free document format. Whenever a document is saved as ODF, the user has the security and knowledge to know that the information contained in this document will ALWAYS be readily available and easily readable by any software that chooses to support the format. The format is completely open and free for anyone to use, so including support for ODF into any piece of software is easily done. The key thing to remember here is that ODF is truly means what it stands for: Open Document Format.

OOXML is Microsoft's feeble attempt to create an open format for documents, but they have failed in this area. The document writes itself off as being free and open, but it is not. Most of the OOXML format appears to be free and open, just like ODF, BUT there is a catch. Certain parts of the OOXML document are PROPRIETARY and NOT OPEN. These parts require PROPRIETARY plugins/software available ONLY from Microsoft, in order to be readable and writable. This means that the data and information you store in an OOXML format is not necessarily going to be available to you in the future. This is absolutely no guarantee. The information you write into an OOXML may very well require an official (and costly) Microsoft product to read in the future. You must understand something about Microsoft: they are a company that thrives on power and control. If they can convince large governments (like the State of New York) to use OOXML, this simply means that the State of New York, will someday be required to purchase massive quantities of Microsoft software to read the information they saved in an OOXML document. This is what it's all about for Microsoft. Money. They do not care about providing a free and open document format. They simply want to make it appear that the document is free and open in order to lock people/companies/governments into requiring Microsoft products to read/write them in the future.

This past year, many international companies and other countries governments around the world have chosen ODF instead of OOXML because of the specific reasons above. It is a ridiculous mistake to use OOXML instead of ODF.

Once again, the bottom line: ODF is far superior to OOXML. ODF is free, open, and will ALWAYS remain free into the far future for all future generations. The Microsoft OOXML format is not truly open. It is not truly free. Part of the OOXML format requires PROPRIETARY software to read/write. This means it will require costly Microsoft software. Microsoft invented OOXML in order to lock people into paying for their software. ODF is completely free and completely compatible across the board.

**INDIVIDUAL # 22:** [INDIVIDUAL'S NAME REDACTED]: [Wednesday 12/19/2007 1:26 AM](#)

**Question 2:** Unrestricted public access on a standard website

**Question 3:** Use published, royalty-free, patent-free standards such as OpenDocument (ODF)

**Question 5:** For standard office documents, mandate use of the OpenDocument spec, which can be read fully by multiple independent software programs from different vendors.

**Question 10:** Whichever format is used, the State should require that the application provider make available full specifications so that other programs will be able to read and write the data should the need arise.

**INDIVIDUAL # 23:** [INDIVIDUAL'S NAME REDACTED]: [Wednesday 12/19/2007 4:15 AM](#)

I'm using only Linux computers at work and at home. The only way for me to freely access documents issued by the government is if they are delivered in a truly open format. I have no doubt that the Open Source community will some day provide me with a OOXML import filter, but it won't ever be perfect. On the other hand there are already freely available implementation of ODF in every operating system I know (including of course Windows). There is even a plugin for ODF for Microsoft Office users.

Please don't alienate your citizens by forcing them to change their operating system and software by using a patent and binary encumbered format.

**INDIVIDUAL # 24:** [INDIVIDUAL'S NAME REDACTED]: [Wednesday 12/19/2007 7:55 AM](#)

As an attorney licensed in NY [REDACTED] I would like to comment that I believe using free open source standards, specifically ODF, is essential for a free and functional society. OOXML is not a free and open standard, since it is tied to proprietary products such as Microsoft Word, while ODF is an ISO standard and can be used by any word processing program. As a solo and contract attorney, I use Sun's free and opensource software OpenOffice.org to word process in my practice in ODF format, yet I must save my documents in .doc to communicate with government agencies or other lawyers.

Why should small businesses and private individuals wishing to view government documents have to pay \$500 to view public documents from the comfort of their office or home? Why do I have to pay Adobe \$500 to fill out simple .pdf forms? Reading and filling out paperwork should be free and open. ODF was written by committee and has the support of multiple companies, organizations, and individuals. OOXML was created by a corporate monopoly and has support of no one except Microsoft and their paid skills. Choosing ODF is clearly the right choice.

**INDIVIDUAL # 25:** [INDIVIDUAL'S NAME REDACTED]: [Wednesday 12/19/2007 8:37 AM](#)

Thank you for soliciting input on this very important issue. While I can't address every one of your questions, I'd like to respond to a few of them that I have direct experience with or an opinion on.

Responses to General Questions

**Question 1. Contact information [REDACTED]**

I am a computer scientist by profession, a physicist by training, and a computer user that has had dealing with my State government in the area of electronic records. I have been doing computer software requirements analysis, software design, programming, testing, and field support for over 30 years. For a portion of this time I have also been involved in standards development in the area of communications protocols used for [REDACTED].

**Question 2. What mechanisms and processes should the State of New York establish for accessing and reading its electronic records in order to encourage public access to those records?**

Access to ALL State information should be available via the Internet, free of charge, to anyone, anywhere. There should be a process whereby a user can request a data set be put onto physical media for non-electronic transport at a nominal cost. (I'm thinking here of large, multi-gigabyte GIS data sets).

**Question 3. What mechanisms and processes should the State of New York establish for accessing and reading its electronic records to encourage interoperability and data sharing with citizens, business partners and other jurisdictions?**

You should not assume that people or businesses will use any particular hardware, operating system, or application software. That's what vendor neutrality should mean. From this starting

point, it's a lot easier to determine what software and/or e-data format is truly open, i.e., is it implemented on all generally available hardware platforms, (PC-x86/Mac/??), operating system (Windows/Linux/Mac), and applications. In general, if an open-source implementation does not exist you have to ask yourself why because there must be a very important reason for it not to exist.

**Question 7. How should the State address the long term preservation of its electronic records? What should the State consider regarding public access to such archived content?**

Long term preservation of electronic records is a requirement, archiving is not since it implies that the information is somehow off-line. All electronic records, once created, need to remain on-line forever. Storage technology is steadily improving and costs are falling so the State should plan accordingly.

**Question 10. What should the State of New York consider regarding the management of highly specialized data formats such as CAD, digital imaging, GIS and multimedia?**

First of all, I don't believe that digital imaging or GIS data is "highly specialized". These are very useful data sets that have been generally unavailable to the average person due to three factors: 1) Lack of availability either due to no data source, 2) lack of bandwidth to download the data and/or 3) proprietary GIS data formats.

Factor 1 requires the State to make data available, factor 2 is being addressed by technology, factor 3 is still a problem.

**Responses to Part II - Detailed Questions**

**Question 5. In terms of appropriate "government control" of electronic records, what factors or concerns should the State be addressing?**

As far as "government control" over public records, I don't see how a government agency can claim that they have "control" over electronic records if they don't have "control" over the software that creates, manipulates and stores those electronic records. In the case of ODF it is certainly possible to have such control over the software, with OOXML it is impossible. This is because there are currently in existence several independent, open-source implementations of ODF while there are no certified implementations of OOXML (by this I mean that it has never been proven that the Microsoft Office suite implements OOXML correctly nor in its entirety.)

**Question 6. Is this the correct definition of interoperability which the study should be using? If not, please provide a better, alternative definition.**

This definition only captures one part of interoperability. One important idea not covered though is that of "exchange of data" between different systems. System would have to include: hardware, operating system, and application software. A better definition that apparently comes from the IEEE is: The ability of two or more systems or components to exchange information and use the information that has been exchanged.

**Question 7. Is this the correct definition of "openness" and "open standards" which the study should be using? If not, please provide a better, alternative definition.**

The phrase "maintained by a non-commercial entity" is not strong enough. Since there are no "official" standard setting bodies in the sense that they can mandate use of such standards, you have to look to the history of the organization, its purpose and methods, and the rate of acceptance and adoption of its standards. ISO standards have some legitimacy, ECMA standards on the other hand, do not. This should be obvious after the recent OOXML rubber stamp by ECMA without any technical evaluation of the standard. The International Organization for Standardization unfortunately appears to be having similar problems with OOXML due to recent politicization of a supposedly technical process.

**Question 12. Other than in the office suite context, in what other ways does the State need to be concerned about electronic records interoperability?**

Electronic records interoperability is very important in the GIS domain. A similar problem exists with GIS data sets as with office suite data, namely the use of proprietary standards that cannot be implemented by open-source GIS systems because of the unavailability of documentation on the standard, patent encumbrances, etc. As a result, today it is very difficult to access and use government GIS data for personal use on any systems running an operating system other than one from Microsoft and using commercial GIS software.

**Question 13. Given the existence of tens of thousands of e-data formats, the increasingly dynamic nature of electronic documents, and a preference toward more open formats in other realms besides office suite formats, what type of an approach or mechanism should be used within the State to further the existence of openness in all relevant formats? Please describe with specificity.**

There is not easy answer to this question unfortunately. Each problem domain will have one or probably more existing electronic data formats. I think one of the main ways in which the State could further the existence of openness is to provide funds to participate in the actual development of standards that are of importance to the State. Librarians have gotten together to develop open standards related to their field. The aviation community is doing the same for air traffic services. Why shouldn't the State be involved in developing standards for office suite data, GIS data, digitized images, etc.

More specifically, choosing a particular e-data format should include the following criteria:

1. Open specification available to everyone for reading and implementation
2. Creates a fair and competitive marketplace. No vendor lock-in.
3. No royalties or other fees are required to implement the standard.
4. Standard setting body must be open to all, rules for establishing a standard must be published and followed, must not be subject to abuse by any single vendor.
5. There must be several independent implementations of the e-data format at least one of which must be an open-source implementation. This would at least indicate that the standard is really open and of some use.

I think one of the most successful standards setting bodies that can be used as an example to follow would be the Internet Engineering Task Force (IETF). While it has no real power over

anyone, its existence has built the Internet using consensus-based engineering. It is not affiliated with any government, country, or vendor. What we need is a Office Suite Engineering Task Force to create a truly open process to develop a world-wide, consensus based, set of office suite standards. Perhaps OASIS is already that organization and they are developing ODF.

**INDIVIDUAL # 26:** [INDIVIDUAL'S NAME REDACTED]: [Wednesday 12/19/2007 9:32 AM](#)

**Question 3. Does the use of particular office suite formats such as the Open Document Format (ODF) or Office Open XML (OOXML) raise any security or privacy implications and, if so, what are they?**

Yes. Microsoft has a long history of keeping deleted text within its documents. This has caused numerous security issues through the years.

**Question 7. Is this the correct definition of "openness" and "open standards" which the study should be using? If not, please provide a better, alternative definition.**

Definitively yes. Open standards must be managed through an open process, and must not be encumbered by licensing fees.

**Question 10. Will the usage of ODF among those individuals and entities with whom the State interacts be so great that failing to provide the NYS workforce with the capability of using ODF will cause NYS interoperability problems? If so, if the State did not adopt the ODF format, what would be the best method to ensure interoperability with ODF documents received by the State from others?**

ODF is the only office format supported by low cost office software (e.g. OpenOffice, StarOffice) whose longevity is assured by an open standards process. As a result, individuals and other states who care about the cost of maintaining their documents are adopting it, and NYS will be at a disadvantage when trying to interact with both its citizens and other governments. If NYS does not adopt ODF, it must invest in developing a converter between OOXML and ODF that preserves all formatting as well as meta-data. This converter should be open-sourced to assure its existence going forwards.

**Question 15. What is the "problem" that this study should be addressing? Please define with specificity exactly what the State should be trying to solve.**

The state needs to identify how to minimize the cost and maximize the ease with which it can both perform its work and maintain its records in perpetuity. To that end, it needs to both identify the entry barriers for the formats (e.g. cost for OOXML, new software installation for ODF) as well as investigate the longevity of different formats through the last thirty years.

**Question 17. Assuming this observation is correct, please provide a numbered list, with the greatest specificity and in the simplest terms possible without marketing verbiage or usage of ambiguous phrases, of exactly which customer requirements are best met by OOXML.**

1. Compatible with the existing office suite used by most individuals and businesses

**Question 18. Assuming this observation is correct, please provide a numbered list, with the greatest specificity and in the simplest terms possible without marketing verbiage or usage of ambiguous phrases, exactly which customer requirements are best met by ODF.**

Open format and existing open-source implementations assure long-term ability to access documents. Open management of the standard assures its evolution is guided by all interested parties Existence of free implementations of ODF office suites means that NYS citizens can access documents without paying a "tax" to a software vendor.

**Question 22. How valid is this concern? Is re-writing of custom in-house software also needed (and has it been needed in the past) for migration between different versions of office suite software?**

The concern is valid, but as the question points out it has been common in my experience to have to re-write applications whenever Microsoft made significant changes to its office suite or software compilers. I have had many programs I personally wrote for Windows broken when Microsoft issued service packs.

**Question 24. What weight, if any, should the State give to the fact that a particular format has been accepted by a standards body? In affording that weight, what elements should the State consider?**

This is a tricky question. Having been burnt by trying to program C++ prior to the introduction of the C++ standard, I am a big believer in only coding applications in ANSI standard languages. HOWEVER, I started making an exception several years ago when I began to use PERL. What I have learned is that the most important element in deciding on a format or programming language is the ability to inspect the code of its reference implementation. In other words, if the source code implementing a format or compiler is available for use and modification, then it tends to be true that the evolution of that format or compiler occurs in a way that minimizes future costs.

**INDIVIDUAL # 27:** [INDIVIDUAL'S NAME REDACTED]: [Wednesday 12/19/2007 10:10 AM](#)

Thank you in advance for taking the time to read this email. I am writing in response to the evaluation of a document standard, specifically in regards to ODF vs. OOXML.

In the simplest of terms, regarding "Laws of 2007, Chapter 477 (codified at New York State Technology Law § 305(4)), which requires "...interoperability, and vendor neutrality...", ODF meets this requirement whereas Microsoft's OOXML does not. As you know, ODF was written by committee and has the support of multiple companies, organizations, and individuals. OOXML exists solely as a Microsoft standard and locks users (of OOXML) in to that standard, which is not even fully specified.

If NY state desires an "open standard" then the following key points should be taken into consideration.

- The standard is adopted and will be maintained by a not-for-profit organization, and its ongoing development occurs on the basis of an open decision-making procedure available to all interested parties (consensus or majority decision etc.).

- The standard has been published and the standard specification document is available either freely or at a nominal charge. It must be permissible to all to copy, distribute and use it for no fee or at a nominal fee.
- The intellectual property - i.e. patents possibly present - of (parts of) the standard is made irrevocably available on a royalty-free basis.
- There are no constraints on the re-use of the standard.

These commonly accepted criteria render the OOXML vs. ODF discussions moot, as the conditions of OOXML patent licensing only fake compliance.

In summary, I respectfully urge you to choose ODF as it is a true open standard and will not lock NY state now, or ever in the future, to a specific company or software platform to maintain its important documents.

**INDIVIDUAL # 28:** [INDIVIDUAL'S NAME REDACTED]: [Wednesday 12/19/2007 10:50 AM](#)

Standards should be designed and maintained in an open cooperative effort for the greater good. OOXML is the exact opposite on both points! The ODF spec is managed by OASIS and approved by ISO, control was turned over to a standards body a long time ago, before it was even a standard.

OOXML is patent encumbered, ODF is not. The OOXML is an intentionally unwieldy specification and only Microsoft has the binary blobs which make it work. Other entities can achieve partial compatibility, at best. On the other hand, ODF actually is an open format which is properly documented and which does evolve in the open.

The fact is there is absolutely no reason (other than corruption) for a government body to go with MS's lock-in format considering the technical merits of both, and most especially the past behavior of MS. OOXML is a pseudo-standard, purposefully obfuscated to keep the MS monopoly gravy-train running smoothly at the expense of New York's tax payers.

**INDIVIDUAL # 29:** [INDIVIDUAL'S NAME REDACTED]: [Wednesday 12/19/2007 11:01 AM](#)

As a NY resident I hope the state implements ODF over OOXML. We do not need a corporation to devise a document standard, the open source community is more dynamic and responsive to the digital needs of our citizens than Microsoft, whose concern is more oriented around profits. Please avoid making a mistake which governments around the world are avoiding.

**INDIVIDUAL # 30:** [INDIVIDUAL'S NAME REDACTED]: [Wednesday 12/19/2007 11:01 AM](#)

I am writing to encourage NY to choose an open document format instead of a commercial format. In 100 years, the ODF will be still viable. Commercial formats make money for the company that owns them when they change. Save the money for other needs.

I was a publications manager for [COMPANY NAME REDACTED]. We used many different products in producing documents for [NAMES REDACTED], and other state and Federal customers. Often, the department was forced to spend money on new software to provide documents in customer formats. We also had to buy old software to open documents that were no more than 5

years old. The worst case was when we had to re-buy software to open our own documents. It was a huge waste of money.

The OOXML format is not open. It will force you to buy Microsoft software and their upgrades. Choose ODF instead.

**INDIVIDUAL # 31:** [INDIVIDUAL'S NAME REDACTED]: *Wednesday 12/19/2007 11:13 AM*

In response to "Part I - General Questions," under "I. Information Requested," pertaining to "Terminology - Access," in addition to the very reasonable points listed there, I define a format's "accessibility" to include openness -- namely, the format must be based on open standards, and be guaranteed to stay that way in the future. This means that those standards are completely documented and specified, and available to anyone, and will remain so. The Microsoft OOXML standard does not meet this criteria. In fact, Microsoft has failed to keep its public promises regarding control of the standard (please see <http://www.groklaw.net/article.php?story=20071206131310362> for more information on this from people involved in the ISO standardization process).

Essentially, once the format is approved as an ISO standard, Microsoft wants to keep the standard under its own control; they will be able to accomplish this because, rather than turning the standard over to ISO, the standards body they plan to turn over maintenance of the OOXML standard to, ECMA, has an OOXML group chaired by not one, but two Microsoft employees (<http://www.ecma-international.org/memento/TC45.htm>). Once the standard is in the hands of ECMA, Microsoft will then be free to add or change features at their whim, leaving any who attempt to implement their standard unable to take advantage of the now \*undocumented\* features. Therefore, they will fail to be in full compliance with the standard.

This will have the effect of locking businesses and government departments into the use of their software, just as if they were to continue to use MS' current, proprietary ".DOC" format. It will also have the effect that, in order for taxpayers to access documents whose creation they've paid for, they must also pay a private company an additional sum in order to access that information. That is plainly wrong.

As a lifelong New York State resident, I am deeply opposed to this standard, for the simple reason that it encroaches on fundamental liberty. There is no justification for creating a de facto requirement that individuals or organizations will, now or in the future, purchase software from a \*private company\* in order to access public documents. Please consider following the good example of the Dutch government in adopting a completely open standard, such as the Open Document Format (see <http://www.odfalliance.org/> for more information), and keep private companies from hi-jacking my public documents.

**INDIVIDUAL # 32:** [INDIVIDUAL'S NAME REDACTED]: *Wednesday 12/19/2007 1:06 PM*

The Open Document Standard is a very weighty title. Microsoft has many good products and titles but all of these are for the sake of money instead of innovation itself. Choosing 'ODF' (Open Document Format) will ensure the continued innovation in the field of open standards, simply because it is, and always has been created by and maintained by those who do not hide their objectives of say- a Capitalist Corporate Agenda.

I openly do not trust Microsoft to be 'open' with anything and their track record in handling software reaffirms this. Just as an example there were many Word Processors in competition at the turn of the Century, Microsoft, like wolves under sheep's skin kindly bundles Microsoft Office with every new PC from Major Retailers, for FREE- Driving all other competition out of business, Then turning around and charging \$150 for a standard suite and up to \$600 for a Professional Suite Version, How could they? An Agenda.

Microsoft will only bow down and welcome open source as it suits their Agenda - to make more money in the long run.

**INDIVIDUAL # 33:** [INDIVIDUAL'S NAME REDACTED]: *Wednesday 12/19/2007 1:27 PM*

**Question 2.)** Use completely open standards which can be fully implemented by any organization. That implies fully open specifications, which enable a full open-source organization to implement a particular implementation. This would mean ODF in this particular case.

**Question 3.)** Use completely open standards which can be fully implemented by any organization. That implies fully open specifications, which enable a full open-source organization to implement a particular implementation. This would mean ODF in this particular case. Without a fully open specification, businesses, citizens and others would be limited in how they can access and use data by \*who\* makes the full reading ability available (Microsoft in the case of OOXML).

**Question 4.)** Not sure what is considered appropriate government control. By and large, things that are available to the public should be publicly available on the internet. Probably a free and open website. Probably no registration should be required except where it is required for individual registrations or payments or the like. But for retrieving data, a simple public website with no access restrictions seems most appropriate.

**Question 5.)** Use completely open standards which can be fully implemented by any organization. That implies fully open specifications, which enable a full open-source organization to implement a particular implementation. This would mean ODF in this particular case. Any records kept by the state should be in a open, well documented, fully specified data format. That would imply ODF or PDF or something like that. Both have many open-source implementations. The state need not limit itself to \*only\* open-source software. But the resultant data from that software should be accessible by any open-source project which can be implemented from freely available public specifications. The freely available software should also be able to \*create\* said documents.

**INDIVIDUAL # 34:** [INDIVIDUAL'S NAME REDACTED]: *Wednesday 12/19/2007 1:54 PM*

I strongly encourage the State of New York to take the lead in adopting open standards for all state government electronic data.

Vendors such as Microsoft have only one goal for their software and document formats and that is more profit. This is more often than not at odds with a democratic government's principles of openness, transparency, and free access for all citizens.

Vendors have in the past purposefully broken formats, made them difficult to implement, and generally done everything in their power to lock in and limit the choices of their customers.

In addition, I have concerns about the future availability of data stored in proprietary formats. What if the vendor goes out of business? What if they decide to stop developing tools that can read and write the format?

I am a professional software engineer and I deal with both open and proprietary formats every day. Open formats empower software developers like me, and in turn empower users. Proprietary formats empower no one but the vendor who owns them.

**INDIVIDUAL # 35:** [INDIVIDUAL'S NAME REDACTED]: *Wednesday 12/19/2007 2:17 PM*

I strongly urge you to find that the public interest is best served by standardization on Open Document formats for all electronic archive documents, and all electronic documents exchanged between departments. It is especially important that documents which are intended for public use be in Open Document formats, except where the proprietary but ubiquitous and well documented Portable Document Format (PDF) would be appropriate.

The reasons for this are threefold:

1. **Guaranteed public access.** Using closed proprietary formats tied to a single vendor for public documents, is a de facto requirement that all citizens purchase commercial software from the same vendor to observe and participate in the actions of their government. Until or unless Federal law prohibits this form of corporate welfare, it is up to the States to assure that the public has access to public documents without having to pay a toll. (This is the primary reason why Israeli law mandates that all public offices use only Open Document formats.)
2. **Archival stability.** Compared to commercial competitors, Open Document formats provide two decisive benefits: Much smaller file sizes (hence smaller storage and retrieval infrastructure costs), and guaranteed forward accessibility (the documents are "future proof" and can never be made inaccessible by software obsolescence).
3. **Cost savings.** The Open Office suite and other free software which uses Open Document Formats, is available free of charge for all lawful purposes public and private. The transition from high priced commercial word processing and spreadsheet software to high quality professional freeware requires no license fees, and there are no hidden costs. Typical MS Office users adapt fully to Open Office in one or two workdays without formal retraining.

Departments which standardize on MS Office or other proprietary commercial office suites incur high costs per new workstation, and are subject to the "forced upgrade cycle" which requires that all installed copies of this software be paid for again every few years, for no reason other than format changes introduced in new versions for the sole purpose of obsoleting the installed versions.

Please be advised that the ISO Open Document standard which is now in place does not include the falsely named "Office Open XML" format by Microsoft, which is a closed source proprietary document format with patent encumbrances and "trade secret" components. Any law which seeks to mandate open document formats must explicitly prohibit the use of closed proprietary document formats which are "Open" in name only. Office Open XML was created for the

purposes of diluting the Open Document trade name, confusing regulatory agencies and purchasing agents, and perpetuating the single-vendor lock in strategy which the legislation now under consideration seeks to correct.

**INDIVIDUAL # 36:** [INDIVIDUAL'S NAME REDACTED]: *Wednesday 12/19/2007 2:22 PM*

I recently became aware of RFPC # 122807 through the website slashdot.org. I am a New York State resident, currently living in [NYS LOCATION REDACTED], and working in [NYS LOCATION REDACTED] as a Computer Systems Administrator. I would like to express my immense support of the ODF file format standard over OOXML.

As a taxpayer, I would like to emphasize that I, as well as many other citizens, care about access to public information. In this regard, it is good to requests for public comments like this readily available.

I must also state that I care very much about how government agencies spend public funds. To decree a technology specification for a file format to be used for most agency business when that specification has only partially been implemented by one single company will not allow for proper safeguarding of data. This one single company will have a virtual monopoly on state information, and the state will be forced to spend taxpayers' money to upgrade every three to four years for no other reason than this one single company decided to change the file format.

Any individual or organization that wants to correspond with state agencies will also have to upgrade their software. In effect citizens will be forced to buy software from this one single company in order to have a say in government, apply for employment in some state agencies, etc. Non-profit organizations will have to spend more money on software, with less funds going toward their intended purpose. State contractors will be forced to spend more money on software or be excluded from some contracts, thus raising their operating costs and the cost to taxpayers.

This one single company is Microsoft, and their partially implemented file format is OOXML. This file format is unreadable from versions of Microsoft Office 2003 and earlier. The OOXML specifications are so poorly written that no other vendor of office suites has released software that can read it or write to it. The OOXML specification contains such references as a "autospaceLikeWord95" routine, yet the only way for other vendors to figure out what "autospaceLikeWord95" does is reverse engineer Microsoft Word 95, which is clearly against any Microsoft terms of license and perhaps against the law.

ODF, on the other hand, is completely open, and has been implemented by numerous companies, such as IBM, Sun, Novell, and many other vendors. Some of these software products are free, or cost very little. In effect no state agency, private citizen, non-profit, or contractor will ever be forced to buy office productivity software from a single company if it uses ODF exclusively.

**INDIVIDUAL # 37:** [INDIVIDUAL'S NAME REDACTED]: *Wednesday 12/19/2007 2:25 PM*

OOXML is NOT an open standard!

**INDIVIDUAL # 38:** [INDIVIDUAL'S NAME REDACTED]: *Wednesday 12/19/2007 2:28 PM*

I am a computer programmer. I have worked in field of computer records at [REDACTED] University and [REDACTED] Corporation of America. Currently, I am working in astronomy research.

Reply to question #2:

The government needs to take a leading role in adopting non-proprietary open-standards like the ISO approved ODF standard. The government should also be more involved in creating standards for computer records for public information, and mandating compliance to such standards by software vendors. Otherwise, the software vendors will continually break existing standards, under the guise of "innovation", in order to maintain a position that prevents competition from other software producers. (vendor lock-in) Microsoft is the most notable for this anti-competitive tactic.

**INDIVIDUAL # 39:** [INDIVIDUAL'S NAME REDACTED]: *Wednesday 12/19/2007 3:04 PM*

I am a private citizen who used to be employed by [REDACTED] Incorporated at the [REDACTED] branch. During part of that time (Feb. 2000 through Sept. 2004) the NYS Office of Children and Family Services (OCFS) contracted my services as a software analyst and developer from [REDACTED] to assist in the ongoing design and development of the Connections project. If you are unfamiliar with the Connections project, it is the system of record for a variety of tasks related to Child Welfare cases for the State of New York. As such, I have first-hand experience with one New York State Office's handling of data, documents, records, and official [government] records (using the definitions specified in RFPC #122807[1]).

**Question 2. and 3.** New York State should standardize on open platform-neutral mechanisms and processes for establishing, accessing, and reading its electronic records. When I joined the Connections project, it was 100% dependent upon the FOUNDATION framework, a proprietary application framework provided solely by Accenture Consulting (which used to be called Andersen Consulting); license fees for this framework for this one project cost the State over \$500,000 annually. During my tenure, a number of new modules were added to the Connections project that used a different application framework; in this case it was MicroSoft Visual Basic and their COM remote object server. MicroSoft was the sole vendor of both the application framework and COM server, and site license fees for these technologies also ran into the tens and hundreds of thousands of dollars. Both frameworks served to lock the project into their respective vendors. It was a laborious process to implement mechanisms to get the two frameworks to communicate with each other. Both frameworks put severe constraints on the technology selection process of not only the OCFS users of the system, but of the hundreds (if not thousands) of vendors and providers who served the needs of children of New York. For one simple example, none of the vendors (who may have included schools, psychologists, and halfway houses) would have been able to use Apple products or a Linux system to perform their required duties in the Connections system. If the OCFS had selected open, platform-neutral mechanisms (web browser technology that comply with international standards, for example) instead of closed, proprietary mechanisms, millions more dollars would have been able to be applied to serving these families in crisis, instead of supplying 50% net profit margins to convicted monopolists.

**Question 4.** My sole comment on this issue is recognition of the superior work done by the OFT in conjunction with [REDACTED], Inc. in 2003 and 2004 to define the mechanisms and processes that should be adopted to encourage appropriate government control of its electronic records. I realize my perspective is influenced by my position as an ex-[REDACTED] employee, but I believe

the "Project Management Handbook" provides a framework that is independent of any particular vendor's products.

**Question 5.** One of the tasks I performed for the Connections project was to support the generation of documents in support of child welfare cases. The data for the case would be accessed from the RDB records, and the components I helped to write would manipulate Microsoft Word 2000 and a third-party ActiveX component (that had licensed Microsoft Word technology) to first produce a Word95 document that could be edited by a caseworker. Once the caseworker has completed their task, a PDF document would be created to provide an official [government] record of that particular action. Considering the data retention requirements of child welfare official [government] records (which I believe to be "time to child's majority + 5 years"), I expect that this repository of documents in the proprietary, undocumented Word95 format represents a looming crisis that will require a nightmarish conversion effort once Microsoft decides to stop licensing the technologies required to read Word95 documents. While the Word95 format has been reverse-engineered by many Open Source Software projects (such as Open Office), trying to recover the contents of these documents in the event of a court case ten years from now may be... problematic. If New York State were to adopt open, internationally-recognized standard for documents that were unencumbered with patents and were supported by several independent document editors (such as ODF and PDF; please note, OOXML meets none of these requirements), it could limit its liability to existing data, instead of adding to it every year. If the OCFS were to work to migrate existing data to these open document standard proactively, it could forestall this looming crisis.

**Question 9. and 11.** When a NYS organization evaluates the cost of implementing a plan to manage its electronic records, I believe that it must identify all software components involved in accessing and reading those records, and specify a course of action it can take if that vendor should eliminate support of that component at any point during the expected retention period of those records. I do not mean to imply that the plan must be comprehensive, nor do I mean to imply that the organization be forced to enact the plan when the that inevitable event occurs, but it is a business reality that all software will one day be retired. The plan must be feasible, however. I believe, however, that failure to have any plan at all for an event that is guaranteed to occur is a shocking omission on the part of those entrusted with spending the taxpayers' money.

**Question 13.** I can not speak to New York State's current standards, regulations, and guidelines regarding records management, as I left my OCFS assignment (and my position with [REDACTED], Inc.) in September 2004, my professional opinion as a software analyst and developer is that they were inadequate in 2004. Despite the release of Word 2000 and Word 2003, and in spite of the availability of other document formats, Word95 format continued to be used for official [government] documents in the Connections project. In spite of recent, negative experience with a proprietary application framework supported by a single vendor, COM objects and Visual Basic were selected for new development efforts. In spite of the data retention requirements of decades, no contingencies were identified to address a failure of a vendor. Change management, risk assessment, and data retention practices all point to the superiority of open, internationally-recognized standards for document formats that are supported by multiple vendors such as ODF and PDF. I heartedly recommend them to your consideration.

**Question 6., 7., 8., 10., 12. and 14.** I have no comment on these issues.

**INDIVIDUAL # 40:** [INDIVIDUAL'S NAME REDACTED]: Wednesday 12/19/2007 9:49 PM

**Question 3)** The State of New York should adopt the approach taken by the government of the Netherlands in adopting Open Document Format, an existing ISO standard, as the standard format for all newly created electronic records.

**Question 5)** The State of New York should adopt the approach taken by the government of the Netherlands in adopting Open Document Format, an existing ISO standard, as the standard format for all newly created electronic records.

**Question 7)** The State of New York should examine the approach adopted by the National Archives of Australia in converting existing documents in proprietary formats to open formats (ODF, PNG, FLAC) for the long-term preservation of digital information.

**Question 10)** The State of New York should examine the approach adopted by the National Archives of Australia in converting specialized data formats, including multimedia, to open formats for the long-term preservation of digital information.

**INDIVIDUAL # 41:** [INDIVIDUAL'S NAME REDACTED]: *Friday 12/21/2007 9:54 AM*

I am but a humble social worker. Still, after 30 years of public service, I've worked with enough systems to have an idea of what should work.

First, the problems . . . it's been a hellish nightmare going from Microsoft version to Microsoft version. It was a crap shoot trying to figure out how to reformat everything I'd written earlier every time the agency upgraded. Please don't leave me at the mercy of Microsoft again! They promise but don't deliver.

Second, I don't need something with a lot of bells and whistles. Most word processors do fine, but make sure it's a format that will be supported until I retire in another 30 years. (I don't want to work past age 90.)

Thanks for the opportunity to vent. It's nice to think I may make it out of word processor hell.

**INDIVIDUAL # 42:** [INDIVIDUAL'S NAME REDACTED]: *Friday 12/21/2007 1:46 PM*

ODF. period.

**INDIVIDUAL # 43:** [INDIVIDUAL'S NAME REDACTED]: *Friday 12/21/2007 3:17 PM*

[EXPERIENCE REDACTED]

### **Disclaimer**

Though an employee of [COMPANY NAME REDACTED], I am filing this comment as a resident of New York State concerned with public policy in the area of electronic document standards. It is my understanding that [COMPANY NAME REDACTED] is filing an official response to the RFPC; that response is entirely separate from this personal contribution. I would be glad to expand upon what follows and to consult further with the Office of the CIO if my expertise in this area is considered helpful.

### **Analysis**

In the interests of brevity and clarity, I will in what follows divide the electronic document problem space into *page-oriented documents* and *transactional documents*.

By *page-oriented documents* I mean the output of the word processing (WP) component of the typical office productivity suite. Such documents make up the majority of the reports and legal instruments generated by State agencies. Slides used in presentations have some of the same characteristics, and portions of what follows will be found to apply to them as well. Due to their tight binding to specific software packages, spreadsheets constitute a separate problem outside my sphere of expertise and will therefore not be treated here.

By *transactional documents* I mean EDI-like XML documents such as purchase orders, invoices, and shipping notices. While both page-oriented and transactional document formats are now being based on XML encodings, they have very different requirements.

### **Page-oriented documents**

Until recently, the primary deliverables of office productivity systems have been paper documents produced by means of WP software. Now we are considering a refactoring of this workflow in which primary deliverables may be electronic documents. In this connection, it becomes essential to recognize the distinction between a final format and an editable format. A final format is one that specifies pages to a level of detail that “locks down” line breaks and page breaks; an editable format is one that allows arbitrary changes to be made to the electronic document. These categories are not logically mutually exclusive, but they are mutually exclusive in practice because of the way that WP formats and applications are licensed and constructed.

This distinction between editable and final formats is important because many official documents include line, page, and footnote references. No conversion between existing editable WP formats can preserve line and page references with complete accuracy, because the different editable formats assume different formatting models and are tied to specific applications that implement different formatting algorithms. Differences in the algorithms that place footnotes, for example, can result in quite different page layouts and can push text from one page to another or result in a different number being applied to the same footnote. Differences in hyphenation and justification (H&J) algorithms and the dictionaries on which they are based can very easily cause changes to line and page breaks, creating errors in line and page references.

Meaningful standardization in this area sufficient to allow multiple applications to produce identical line and page breaks would require all WP vendors to use a common core of running code. This is only possible under two conditions: a virtual monopoly by a single commercial vendor or universal adoption of the same opensource product by all the vendors. Monopoly control has unacceptable cost implications for government agencies, and universal adoption of a common code base is at this point far from reality. Absent this level of standardization, there can be no guarantee that a page, line, or footnote reference embedded in a text encoded in any current editable electronic format will still be correct when that same text is rendered by a different application or even by a different version of the same application. Such differences can create obvious problems in informational documents and critical problems in legal documents.

For this reason, an electronic deliverable that can replace all or some paper documents in the traditional government workflow must today use a final format. The only standard final format worthy of discussion in this connection is ISO 19005-1, aka PDF/A. It follows that decisions about

WP formats and applications are (or should be) decisions about a work environment whose eventual output will be PDF/A files and/or paper realizations of those files. In particular, it should be understood that the publicly archived and published deliverables of State agencies should never be presented in an editable format, because the public has no business editing published documents, and the State even less so. Decisions about WP formats and applications should therefore focus on their role within the agency as part of document preparation workflow and not on some presumed role as a delivery medium.

#### ***Word processing formats: one or several?***

Within the context just described, the question arises whether to allow multiple WP formats going forward or to require the adoption of a single format. In my opinion, it will be much better to adopt a single format across all State agencies, for the following reasons.

- While it is impossible, as explained above, to achieve complete page and line fidelity even across different versions of the same application, in practice the presentational fidelity of documents in progress will survive exchanges between different authors and across different agencies far better if all authors are required to use the same editable format.
- The character-based XML encoding used by current WP formats opens up documents to user-written scripts and programs in way that was impossible with older binary formats, enabling the development of a rich assortment of document processing tools created by State workers. Use and management of such tools will be greatly enhanced by the adoption of a single format within a given agency and across multiple agencies. This applies to commercial indexing and search tools as well. Everything about document management becomes easier with a reduction in the number of document formats.

#### ***Word processing applications: one or several?***

A similar question arises with regard to WP programs; assuming the adoption of a single WP format, should everyone in an agency be required to use the same application to create documents in that format? While adoption of a final form such as PDF/A for electronic deliverables can solve the problem of line and page references, other factors relating to management make it advisable to standardize on a single WP application.

- Author training, stylesheet design, and the maintenance of a consistent “look and feel” are vastly simplified by the adoption of a single WP application.
- Software management across an agency is similarly benefited by the adoption of a single WP application.

#### ***Choice of a word processing format***

At this writing, the chief contenders for a standard WP format are ISO 26300 (aka ODF) and Ecma-376 (aka OOXML). For public sector use, it is my opinion that the best choice is ODF, for the following reasons:

- ODF is an already existing International Standard created by multiple vendors and maintained in an accountable standards process open to all interested parties. OOXML,

on the other hand, is a proprietary specification designed to maintain market domination by a single vendor — Microsoft. That current attempts to promote OOXML to International Standard will probably succeed should not be allowed to obscure the fact that its development is intended to further Microsoft's historic monopoly control of the office productivity software market — control that has resulted in huge unnecessary costs to public agencies and to members of the public seeking access to documents created by public agencies when the editable WP format has been used (mistakenly, in my opinion) as the publicly available deliverable.

- Market considerations aside, ODF is a reasonable, well-designed specification that leverages other well-designed standard technologies and does not carry the technical burden of years of proprietary Microsoft engineering decisions.
- OOXML, by contrast, is an exceedingly complex specification nearly an order of magnitude larger than the ODF Standard and weighed down with sections that exist only to cater to Microsoft applications, including fossilized versions of software bugs in those applications. From an engineering standpoint, it is a mediocre specification that is not well designed for meaningful application interoperability.
- Complete implementation of OOXML cannot be achieved without access to Microsoft code. ODF, on the other hand, can be implemented by anyone and is unencumbered by any vendor's intellectual property claims.

Two arguments commonly offered for the adoption of OOXML turn out upon closer examination to be specious.

First, it is argued that adoption of OOXML will promote interoperability between multiple implementations. Leaving aside the fact that the ODF standard already provides such interoperability, OOXML, despite its enormous size, is too poorly specified in places to allow separately developed implementations to seamlessly interoperate without access to Microsoft's proprietary code. In fact, the authors of the OOXML specification explicitly disclaim any expectation that independently developed implementations will produce the same results when given the same OOXML document, saying that "application conformance is purely syntactic" (Ecma-376 Part 1, Section 2.3). In other words, according to the OOXML specification itself, an OOXML application is conformant if it just accepts an OOXML document as input; it need not produce results that resemble those produced by any other OOXML application. Consequently, the only way that an agency can be reasonably confident that OOXML documents will produce expected results is to continue to use Microsoft applications — which is, of course, the point of the exercise.

The second specious argument offered for the adoption of OOXML is that such adoption is necessary to provide interoperable access to billions of legacy documents trapped in older proprietary Microsoft formats. This claim is simply not true. An owner of legacy Microsoft Office documents who wants a completely faithful rendering of those documents will still be required either to continue maintenance of the Microsoft products that created them or to purchase new Microsoft products that recognize the old formats using Microsoft's proprietary knowledge of how those formats were intended to be rendered. No one other than Microsoft can have the necessary knowledge for how to do this, because the OOXML specification does not provide a normative mapping between the OOXML format and the older binary Office formats, and access to documentation on the older formats does nothing in itself to correct this. On the other hand, an

owner for whom a rough conversion of legacy documents to a standard XML format is sufficient can already accomplish this using existing free products such as OpenOffice. In either case, OOXML adds nothing to the ability of legacy document owners to access their documents, because OOXML does not put competing products on an equal footing with Microsoft products when it comes to correctly rendering the existing document base. Once again, the actual objective of OOXML is to keep users locked into continuing upgrades of expensive proprietary software, directly contrary to the economic interests of the taxpayers of the State of New York.

### ***Choice of a word processing application***

I recommend OpenOffice as the WP application to be adopted by governments. The OpenOffice suite provides all the functionality needed by public agencies (including an excellent free PDF export facility) with none of the costs imposed by reliance on proprietary software. Furthermore, the fact that OpenOffice code is completely open to inspection offers public agencies the only absolute guarantee that the software is free from security holes and hidden dependencies on proprietary technology. And the availability of OpenOffice on multiple platforms enables a later transition to open-source operating systems such as Linux and Solaris. For these reasons, OpenOffice is achieving growing acceptance by governments all over the world. It would be very much to the financial advantage of the State of New York to follow their lead.

### **Transactional documents**

Implementation of a standard XML-based electronic format for documents such as purchase orders and invoices can also result in very significant savings for state governments.

While public attention has been focused on the debate over XML office productivity formats, XML has been quietly expanding the scope of, and in some cases replacing, older EDI encodings for government procurement. XML transactional schemas define a standard set of tags for business data, just as XML WP schemas define a set of tags for printed representation.

For example, Universal Business Language (UBL), an OASIS Standard, provides a set of 31 transactional XML schemas specifically designed for government procurement. UBL has been endorsed by the United Nations Centre for Trade Facilitation and Electronic Business (UN/CEFACT), and adoption of UBL in the public sector is underway in a number of European countries.

When mandated by governments, standard XML transactional formats such as UBL can incrementally replace paper forms in existing workflows and achieve a savings of several dollars in the cost of each transaction. In Denmark, where UBL Invoice has been mandated for all public sector business, some 1.25 million electronic invoices are exchanged every month, with savings to the Danish government estimated at 100 million euros annually. This is in a country with a population of 5.5 million, less than one-third the size of the population of New York State.

[COMPANY NAME REDACTED] has no particular financial interest in UBL and has sponsored the development of UBL purely for the public good, just as it sponsored the development of XML itself. As a NYS taxpayer who also happens to chair the [TECHNICAL ORGANIZATION NAME

REDACTED], I would be gratified to see this technology used to increase the efficiency of NYS government procurement. [IDENTIFYING INFORMATION REDACTED].

**INDIVIDUAL # 44:** [INDIVIDUAL'S NAME REDACTED]: *Saturday 12/22/2007 5:26 AM*

**Question 1. Contact info:**

I am currently retired, but continue to consult for several national and state governments on IT technology.

-----

**Part II**

**Question A1. Are the distinctions described in Part I of this RFPC between the definitions of electronic data, documents, and records useful? Are there any specific elements or distinctions in those terms which CIO/OFT should be taking into account?**

These distinctions are not drawn carefully enough to be useful in the roles they're intended to play. For example, e-data is defined as information evidencing an activity. But consider the electronically-stored latitude and longitude of a point on the boundary of a legal description of a plot of land. It is true that (in some sense) this data is evidence that somebody measured or calculated it. But that does not capture the essential or significant meaning.

It is probably better to say something like: electronic data is information stored or transmitted by electronic means in such a way as to permit consistently reliable interpretation, when it is subjected to predefined interpretation processes properly performed.

Regarding the definition of e-documents: firstly "physical realization" is completely out of place here. It is no more or less germane to electronic documents than it is to anything else electronic -- data, records, whatever.

It is essential to recognize that an electronic document is an abstraction. This abstraction is meant to support a mental model that humans use when they think about paper books, ledgers, photographs, etc. The mental model itself is not particularly well defined. We can mentally envision two identical copies of an imaginary book. But if we try to envision a million copies next to a billion copies, our imagination fails. It is best not to rely very much on these imaginary mental models, and to focus instead on the document abstraction per se.

We are left with something like this: a particular e-document is an instance of an abstraction. All instances of e-document abstractions are imperfectly realized, nonetheless, all valid realizations have two necessary components: a particular set of e-data, together with software that interprets and renders the e-data into useful forms. Importantly, the e-data in the realization must conform to a set of rules that specify how the e-data are organized and assigned values (otherwise the e-data cannot be usefully rendered). Those rules constitute the format to which that e-document adheres. It is essential to recognize that a given ruleset or format may be renderable by software created by a multiplicity of sources. The more diverse the multiplicity of sources is, the more open the format can be said to be. It is also important to recognize that format definitions can be descriptive (defined in some external, independent form), or proscriptive

(defined implicitly by the programming of the software that created it). The more descriptively defined a format is, the more open it generally is.

This is a lot of philosophy for a comment written for government consumption. The key takeaways are these:

- If you are going to try to make a distinction between e-data, documents, and/or records, you need more precise definitions before those distinctions are useful in practice. It is probably just as useful to lump everything under the heading of electronic information and let it go at that.
- The notion of an e-document is particularly pernicious, because we all think we know what we mean when we use the term. I hope the previous paragraphs have illustrated that our suppositions don't survive close inspection.
- When we do take careful pains to define e-documents, we see that the concepts of format and interpretive software are intrinsic. Without them, the term e-document is almost vacuous.

-----

**Question B2. Is the description in Part I of this RFPC of three types of access needed for electronic records – day- to-day utility access; ancillary active record access; and historical access – a realistic and useful conceptualization of the main uses of electronic records? If not, please describe with specificity recommendations for alternative methods for conceptualizing the study's issues.**

Although these 3 distinctions are meaningful in certain contexts, the over-riding point is this:

Due to the increasingly tight inter-relationships between items of electronic information (think of a URL in a email pointing to a website page), and due to the rapid time-variance of that information (we have all had the experience of following a pointer to a web page only to discover that the information we want has been replaced or discarded), it is almost impossible to know, during day-to-day utility access, which parts of which documents will need to be accessible in ancillary access or historical modes. *Even if we have a good idea of a document's historical value during day-to-day operations, the more effort it takes to convert between "day-to-day" format and "historical" format, the less likely it is be that historically significant documents will be preserved. This ceases to be a problem if there are no substantive format differences between the three types.*

-----

#### **G. Functionality.**

**It has been observed that "it is very important that customers have the freedom to choose from a range of technologies to meet their diverse needs. [OOXML] and ODF were designed to meet very different customer requirements."**

I believe this argument is specious. When it comes to interoperability standards, choice between essentially equivalent standards is NOT desirable.

I am sensitive to the fact that New York is locked in to Microsoft Office formats, just as most businesses are.

If Microsoft were truly committed to Open interoperability standards, it would endeavor to fold OOXML into ODF in the spirit and letter of International open standards processes --- instead of continuing its attempts to stuff ballot boxes and game the system.

For New York, a transition plan to Open document formats (per the South African definition) will surely be required. But that effort should proceed using EXISTING Office 2003 binary formats, and avoiding OOXML as a needless, wasteful, and attractive-only-to-the-gullible detour.

**INDIVIDUAL # 45:** [INDIVIDUAL'S NAME REDACTED]: *Sunday 12/23/2007 2:55 AM*

**Question 1. [IDENTIFYING INFORMATION REDACTED]**

**Question 2. To encourage public access to electronic records:**

It would be useful to present the information in HTML pages for quick reference. Which of course could be saved by any modern graphic browser. Ideally there would be links to save the information for later reference in ODF format, or alternately PDF format if the exact presentation is important. The advantage of using ODF is that the user can readily edit the information for their own usage to, for example, keep only the first 20% of the lines, if that is all they find of interest.

**Question 3. To encourage interoperability and data sharing:**

Minimise the formats used, selecting open formats accessible on the widest range of operating systems and a large number of independent vendors. This helps ensure that the largest numbers of users have access to the information. Note that to enhance the sharing of information, they must be a modifiable formats (such as ODF text, spreadsheet, presentation and drawing formats). It is important that the format NOT be restricting to a sole operating system, and NOT be restricted to a sole vendor. As well, for long-term utility, the format must be open and well-defined, so that other vendors can add support for these formats, as inevitably some vendors will disappear over time, and there has to be an avenue to alternate support for the formats.

Note that it is not necessary that the software use the chosen formats as their native internal format (although that would be better), as long as the chosen format can be reliably imported and exported without loss of function. Again in practice this requires that the chosen formats be well-defined, so that any vendor can reliably implement them.

**Question 4. Appropriate control of electronic records:**

Internally the records can be stored in any convenient format, it is the publishing of the information that should be in a universally accessible format. The best control is to require that all internal office-suite type communication be kept in the universal format, except in special cases where there is no software available for specific needs. (I can't think of any example that can't be converted to ODF and function with Openoffice.org, but some may exist.) Evidently large amounts statistical information would be best stored in some sort of relational database, a spreadsheet not being a practical option. A strict policy requiring specific permission for any new information stored in non-universal format would be useful. For existing information in non-universal format, a conversion committee to monitor and enforce the conversion of such records would be useful.

**Question 5.** Insist that all new office suite software acquired be capable of reliably importing, processing, and exporting the universal formats chosen. Of course, allowing for special cases where there is no software available for specific needs, if they exist.

Although such a policy may seem a bit harsh for a certain large vendor that has (to date) chosen not to implement a truly open format, that is their choice.

**Question 6. Life cycle management:**

- a) all new documents should be in the new universal formats.
- b) maintenance: after a period of transition, all modified documents should be in the new universal formats.
- c) exchange: all exchanges in the new universal formats
- d) preservation: all old documents should be eventually converted, unless they are not considered to be worth keeping.

Again, this should be monitored and enforced by a conversion committee.

**Question 7. Long term preservation of electronic records:**

Archived content should be eventually converted, as long as it is considered worth keeping. This ensures long-term access, as software capable of reading older formats is likely to disappear in the long term. Note that data stored in open format databases will likely remain accessible.

**Question 8. (Not familiar with statutes.)**

**Question 9. Constraints/ benefits.**

Should think long term. To provide better access to government information has its costs, just as does the existing program, and the maintenance of electronic records. By using open formats, the cost of software will be substantially reduced. (Due to both the availability of free software and the increased competition around common file formats.) The rationalization to fewer open formats which will be available in the long term should also reduce costs. The current situation with significant changes in proprietary format every few years is expensive. In the short term, there are higher costs of conversion, in the long term these will disappear. In short, probable higher costs short-term with the likelihood of considerable savings in the long term. And a better quality of service for the public. Conversion is a good long-term investment.

**Question 10. Highly specialized data formats:**

These should be lowest priority for conversion. Convert when and if open standards exist. The impact on external communication is considerably less. Especially in multimedia, open formats are developing. By the time office suite formats are converted, many more open specialized formats are likely to exist.

**Question 11. See also 9.**

The standardization of electronic record formats should result in long-term savings, due partly to reduced software costs, and reduced long-term conversion costs.

**Question 12. (No examples)****Question 13. (Not familiar)****\*Part II\*****Question 1. Distinctions (electronic data / documents / records):**

I would say that (information / documents / databases) is probably a more useful model.

**Question 2. Types of access: model is appropriate.**

**Question 3.** The use of ODF is quite secure. It can be encrypted so as to be totally impossible to decypher without a password. (The XML code content is totally unintelligible, even though other elements - like embedded graphics - can be extracted.)

OOXML presents a security risk - it must be converted by an external program, if available - to be read on most systems. To be converted, it must be NOT encrypted - thus insecure. Considering that many systems will not have a reliable OOXML conversion utility available - it could be secure, but unreadable.

Note that OOXML is available only on the latest version of Ms-office on Ms-Vista. So to be able to import OOXML directly, one has to be part of the small minority that is running Ms-office-vista on ms-vista. Not available with openoffice.org, not available on Linux or Mac or Unix, not available with Ms-office-2003 or previous.

And if by chance one doesn't have an office suite capable of reading ODF, a citizen can download Openoffice for free for Ms-windows / Linux / Mac / BSD / Solaris. (And one can compile the source code for any Unix.)

**Question 4. Accessibility / Rapidity of response for FOIL.**

Adoption of OOXML would considerably reduce the accessibility for those not having (the rather expensive) ms-office-vista on ms-vista. Adoption of ODF would provide universal access, due to the availability of Openoffice and other software at zero or minimal cost. The rapidity of response could be problematic at first for either of these options government side due to the conversion process. It is conceivable that temporary relaxation of required response time for FOIL would be necessary. It would be useful for the transition committee to closely monitor this process.

**Question 5. Appropriate "government control":**

This should not change much, assuming that current practices are adequate. Note that it is always risky to put confidential information on line. Password-encrypted ODF files with authenticated logon should be a minimum for accessing this type of information.

**Question 6. Definition of interoperability: very well said.**

Note that the definition includes ODF (available with numerous products on numerous systems) and \*not\* OOXML (available on a single product, single system).

**Question 7. Definition of openness, open standards:**

The best definition of the many I've seen. (By the way, this definition is promoted by the Ubuntu distribution of Linux, which is produced by a South African company.)

**Question 8. (Not with a state agency.)**

**Question 9. Gartner's prediction re ODF:**

If you count the population represented by the governments, I would say probably ... ODF is gaining momentum in France, Germany, the rest of Europe, India, Brazil, South Africa, a version of ODF in China ...

Many predict that Microsoft will wake up and finally accept ODF ... and that OOXML will die ... with or without Ms-office-vista. If New York does not accept OOXML, that will accelerate the process. (How useful is it for New York to select OOXML if it dies in a few years?)

**Question 10. If New York does not adopt ODF formats:**

Problems with ODF documents from other governments ... very good point ! Many other governments are moving toward mandating ODF in all their communications, where possible. Note: If New York adopts ODF, that will accelerate the adoption of ODF by citizens and businesses, who can download an ODF-compatible suite for free.

If New York adopts OOXML, that will have little effect on the adoption of OOXML, since Ms-office-vista is expensive and only available on Ms-vista. The only way to ensure interoperability for ODF documents is to install Openoffice.org or some other ODF-capable software. That is, to in practice go part way to adopting ODF.

By the way - Openoffice.org (or Staroffice, the variant with support services from Sun Microsystems) has built-in support for legacy Microsoft office formats. (Not perfect, but reasonable for reading documents, and available on all supported platforms.)

**Question 11. Governments adoption of ODF and/or OOXML and other formats:**

**ODF exclusively:**

**-countries:**

Belgium; Brazil; Croatia; Finland(ministry of justice); France; Hong Kong; India(electoral commission); Italy; Japan; Korea (South, recommended); Malaysia; Netherlands; Norway; Poland; Russia; South Africa

**-provinces/states:**

Misiones(Argentina); Kerala(India);

**-cities:**

Freiburg, Munich (Germany); Bristol(UK)

**OOXML exclusively:**

(to my knowledge, none)

**Both formats:**

Denmark; Massachusetts

**Other formats supported:**

PDF, HTML

**Question 12. Interoperability outside the office suite context:**

Formats for CAD, multimedia

**Question 13. Complex question.**

There is an important difference between formats used internally, formats published, and formats accepted from the exterior. Where the usage differs between these usages, there is also a problem of conversion.

Divide the formats into 4 groups:

- "preferred" formats (open formats),
- "tolerated" formats (largely proprietary formats in wide usage, accessible by alternative products),
- "depreciated" formats (proprietary formats in common usage for which equivalent preferred or tolerated alternatives exist), and
- "redundant" formats, proprietary formats for which viable alternatives are readily available.

Open should be defined according to the excellent definition by South Africa. (Note my selection of names for the groups is arbitrary.) For this purpose, open formats under development could be put in the tolerated group. Certain formats, such as PDF, although originally proprietary, could be treated as preferred, due to wide usage and availability of multiple products on multiple platforms.

Internally, the usage is largely arbitrary. The usage of "preferred" formats where available should be mandated.

Formats accepted from the exterior fall in 2 groups: Documents specifically requested (like Requests for Proposals), where the usage of "preferred" formats could be mandated. Exceptions may have to be made for certain types of information, in which case "tolerated" or "depreciated" formats could be accepted.

General invitations for feedback on various issues could be done by email (as in the present case) or web pages, to ensure the reception of the responses in "preferred" formats. Non-solicited

documents could be accepted in a wider range of formats. It would be reasonable to refuse documents in the "redundant" group.

For publication, the usage of "preferred" formats should be mandated, where available for the information released. Note that PDF is classified as preferred, so those not have installed an ODF agent (such as the free Openoffice.org) would still have ready access to office-type documents. The state's web pages could provide a link to download an ODF agent, much as many web pages now provide a link to download Acrobat Reader for PDF documents.

A similar approach could be made for all other open formats. In this way, the state would be actively promoting the use of open formats, which in the long term facilitates the task of the state in the distribution of information. By the way, as some sites do present alternative agents for PDF files (e.g. Acrobat\_Reader and Foxit\_Reader), the state could present alternative open format agents, when available.

There will be a period of transition required, such changes cannot be made overnight. A good approach would be to give warnings of the new policy on web pages and other publications, with reference to a specific web page giving the policy in detail. This page would list the specific formats according to their acceptability for publication / reception / internal use.

Note that according to my classification, ODF formats are "preferred", and OOXML formats (being non-open) are "redundant" (the lowest category). However, ms-word and ms-excel formats would initially be "tolerated" (the second category), due to wide usage and existence of alternative products for the format.

**Question 14. The proposed focus is very good.**

Focusing on office suite formats is highly appropriate, as this is where the most immediate progress can be made.

**Question 15. The most important problems to solve are**

- a) Long-term security of data, through the use of open formats.
- b) Interoperability, again through the use of open formats.

Open formats facilitate multiple implementations of agents to process the formats. Formats based on XML offer an additional advantage, facilitating the exchange of information. The resulting cost savings should be considered a bonus, to be realized in the longer term. The goal should be better service and stability.

**Question 16. Current formats (independent computer consultant)**

<0,6% .doc .xls .ppt (ms-office) (mostly documents received)  
<0,2% .rtf  
4,3% .pdf  
70,6% .html  
17,7% .txt  
4,6% .zip  
xml formats:

1,2% .sxc .sxw .sxi (early version of ODF, readable by Openoffice and others)  
0,9% .ods .odt .odg .odb .odi (ODF formats)  
(no support for - or intention to support - other xml formats for documents, although used in various programs)

New documents generated exclusively in .txt or ODF formats, with conversions to .doc or .pdf format when required by an external party.

(The ODF document remains the document of reference.) .txt and ODF formats have the advantage of being identically generated and viewable on multiple operating systems. (Use a dual Microsoft / Linux environment.)

Would prefer the availability of more documents in ODF formats, replacing all ms-office formats, and many documents now available in .pdf and .html formats. This would facilitate quickly removing parts of documents of no interest.

Note that ms-office .doc format is approximately twice as large as the equivalent document in ODF .odt format. (Due to .zip-compatible compression of ODT, counts conversion in BOTH directions.)

**Question 17. Observation is apt.**

OOXML is designed to serve:

- 1) Users of Ms-office-vista running on Ms-vista who prefer not to use other data formats.
- 2) Users with conversion utilities available who transfer documents to/from users in category (1)
- 3) Users who appreciate the advantages of XML-based documents, and who prefer to remain in a Microsoft-only environment, and do not value exchanging information with those outside such an environment.
- 4) Microsoft

**Question 18. ODF is designed to serve:**

- 1) Users who wish to preserve the long-term access to their data, without being subject to the whims of a single vendor. This will include many current users of Ms-office and Microsoft operating systems.
  - 2) Users who prefer the flexibility and interoperability of XML-based documents.
  - 3) Users who use non-Microsoft operating systems. Such as Linux, BSD, Macintosh, Unix.
  - 4) Users who do not use the Ms-vista operating system.
  - 5) Users who do not use Ms-office
  - 6) Users who appreciate using free or low-cost software
  - 7) Users who favor the open source model of software.
- (end question 18)

**Question 19. Office suite format best for state's day-to-day operational functions:**

I would say ODF, for the following reasons.

- For legal applications, Wordperfect has long had a preferred position. Wordperfect supports ODF, and due to its completely open specification, they will be able to fully support it. (Assuming the state uses Wordperfect for such applications.)
- For any applications using embedded graphics, the support in ODF-based Openoffice.org (or Staroffice) is excellent, and more versatile than the support in ms-word .doc format. (This shows up in conversion. One must avoid certain means of anchoring graphics in order to display properly in ms-word. Only special effects cause a problem in conversion from .doc to .odt) It is possible that Microsoft will correct this problem with OOXML, but I wouldn't depend on it.
- Since ODF was designed from ground up to be an integrated text / spreadsheet / presentation / drawing format, it integrates very well. Import a spreadsheet table into a text? The appearance is identical. It is stored in exactly the same format. This can be verified, as an ODF document is a zip-compressed file tree. Decompress the file, and all components can be readily compared. (If the document is password-protected, the non-graphic components are encrypted so as to be unreadable.)
- All elements of ODF are well-defined. This is a major advantage of ODF over OOXML. There are no definitions which depend on the functioning of a particular program, which may produce logically incorrect results. ODF does depend on formats defined elsewhere, such as .gif, .png, .jpg. But these formats are not defined in terms of the functioning of another application, which is one major weakness of OOXML.
- ODF is already available from multiple vendors, on many platforms. Since there is no fee for using the format, it promises to become an almost universal office suite format, at least for import/export.

To date, the only major hold-out is Microsoft. If important governments do not accept their format, they won't likely hold out for long. In any case, ms-office users will still have access to the older ms-office formats, which will permit them to transfer information to/from others for a considerable time. (end question 19)

**Question 20. For long-term preservation / production of electronic records:**

ODF is the ideal solution.

- It is well-defined - explicitly defined, with reference to well-defined external standards.
- It is already accepted as an international standard, and mandated by a growing number of governments and organizations around the world.
- It is well integrated - for example table in a text is defined identically to the same table in a spreadsheet. Similarly for a drawing. Since it is a zip-compressed file tree organized around an XML file with a manifest, this is easy to verify.
- The availability from multiple vendors further ensures the formats long-term viability

**Question 21. Any electronic format could be useful for purposes of litigation, as long as it is readily readable and searchable.**

The fact that ODF is an openly defined and freely usable international standard already mandated by a number of governments and already available from a large number of vendors on many operating systems assures its viability for the foreseeable future. Thus using ODF would

facilitate the utility of electronic records for this purpose. Microsoft's OOXML offers no such assurance. Other formats such as plain text, pdf, and Mozilla/Netscape's email format could also be useful for this purpose.

**Question 22.** Such customization must be rewritten from time to time to accommodate changes in the user agents. This is necessary even if one stays with Microsoft products - there are significant changes on a regular basis, and some aspects of any application are likely to change.

Under ODF, macros (embedded scripts) are defined as application-specific, but are signed when saved so as to signal changes by other applications. ODF-based Openoffice/Staroffice have java-based macro languages, which have the advantage of being executable on all the operating systems for which the Java Runtime Environment (JRE) is available, which currently includes Microsoft, Linux, Solaris and Mac.

(Java is a semi-compiled language which requires the JRE to execute.) One of Openoffice's macro languages is a version of Basic similar to the VBA Basic used in macros in Microsoft office. It is a fairly simple process to import an Ms-office document with its macro, modify the macro to function under Openoffice, and save it in the ODF format.

(Note 1: To use macros in Openoffice, a Java-enabled version of Openoffice must be installed, along with the Java Runtime Environment. To prevent the use of macros, which could be a security requirement, simply install a version of Openoffice/Staroffice which is not Java-enabled.)

(Note 2: This refers to Ms-office formats excluding the new OOXML, which cannot be converted by Openoffice.)

Microsoft Access is a simple relational database system, which can be replaced by more advanced alternatives available on many operating systems, some of which are open source. Note that ODF-based Openoffice now has a database access function (which formerly existed only with the related Staroffice suite). It allows creating/connecting to a wide variety of databases, including such variants as a plain text file, a spreadsheet, Oracle, the open source MySQL, several generalized database access protocols, and Microsoft Access.

The general advice in the computer industry is to avoid ActiveX due to potential security risks, since ActiveX can freely write to disk. Java is recommended as inherently more secure, since writing to disk is normally blocked. However there are reasonable circumstances for using ActiveX-type functions. Note that if ActiveX is used to link between different ms-office applications, that is totally unnecessary with ODF: ODF defines a totally integrated text / spreadsheet / presentation / drawing / database environment. The format of a table is identical in a spreadsheet or embedded in text or a presentation. Similarly for a drawing. A table in a text document can be embedded (physically in the text document) or external. It could be even accessed on Internet.

It could be reasonable to decide, for existing custom software, to keep using proprietary technologies such as ActiveX and Microsoft Access, especially in the shorter term. However, especially for long-term preservation of data, it would be advisable to use open technologies for any new custom software, and in the longer term convert existing applications.

In short, if one uses Openoffice/Staroffice for ODF, there should be little problem with rewriting software for migration, even with certain Microsoft-centric technologies. Probably not much more than a typical upgrade to a new version with Microsoft.

**Question 23.** (Not with a state agency)

**Question 24. Weight to fact that format accepted by a standards body:**

The primary consideration should be the probable persistence and reliability of the format. Open formats have the considerable advantage of being easier to verify as reliable, and tend to attract more vendor support, as is the case for ODF.

For example, Microsoft's OOXML has already been approved by a standards body (but not the overall ISO), but it is not well-defined, is not currently supported by other vendors, is not likely to be in the future, and certainly not on non-Microsoft platforms. For those reasons, it is a crippled format that has little probability of persistence in its current form. (And consequently little probability to be accepted by the ISO.) ODF has the advantage of being relatively well-defined, by an open process, is already supported by a large number of vendors on many platforms. And it has been approved by the ISO. The approval by the ISO is largely a consequence of the probable persistence and reliability of the format. Of course, ISO approval helps the probable persistence.

The PDF format has only recently been approved by the ISO as a formal standard, but in practice it has long been a "de jure" standard, widely used, supported by multiple vendors on multiple platforms. Thus already a high probability of persistence, with a reliable format. In sum, formal acceptance by a standards body is a plus, but probable persistence and reliability should be the key question.

**Question 25. Standardization on a single office suite format:**

would certainly promote competition in the IT marketplace. Vendors would have to compete on program features, not on locked-in proprietary formats. WordPerfect has long had innovative word processing software, with a number of interesting features that its competitors do not have. They are a major player in the legal office suite niche, despite domination by Microsoft in other sectors. Although ODF was developed by Sun Microsystems with strong support from IBM, both competitors to Wordperfect, Wordperfect was quick to support ODF in its software. With ODF, Wordperfect can compete on features, which can only improve consumer choice. It is also good for vendors. They don't have to keep reinventing the wheel. A single office suite format focuses competition on features.

Imagine a world without standards. Your car needs repair. Every manufacturer uses its own size nuts and bolts. The mechanic has a different set of wrenches for each brand of car. You buy replacement headlights according to your brand of car. You buy gas, every gas station uses a different size pail to distribute the gas. (Can't use a pump, no standard way to measure the gas!)

Microsoft would give you 2 standards. OK, so your mechanic only needs 2 sets of wrenches. Only 2 ways to measure gas. Perhaps a little better. But still more expensive. And what if you do some of your own repairs, and you have a set of wrenches for car type A. You see an interesting car, are thinking of buying it, but it is type B. You'd have to buy a whole new set of wrenches.

A single standard simplifies life for everyone.

**Question 26. Who should enforce standards regarding creation of electronic records?**

For the conversion, there should be a conversion committee with enforcement powers, which could become a conformance agency for enforcement in the long term. It could be appropriate to give the NYS Archives this authority, I'm not familiar enough with the state's organization to say. In any case, the conformance agency should also have veto powers over procurement not consistent with the state's standards, as that is an area of potential deviation from the standard. E.g., policy mandates ODF format, procurement of 1000 copies of ms-office-vista would be vetoed by the conformance agency (unless a policy-approved exception applied).

**Question 27. State mandates a single document format for state agency use:**

For state: If an open format, reduced costs through standardization and increased competition. If a closed standard like OOXML, any reduced cost through standardization would be more than offset by increased costs due to expensive software.

For citizens: Reduced costs if open standard, especially if ODF since a number of free ODF agents exist, on many platforms. If closed standard such as OOXML, increased costs as it is expensive, or lack of interoperability if the citizen does not have (or can't afford) a recent computer with ms-vista running ms-office-vista.

For vendors: somewhat reduced costs if open standard such as ODF, since vendors would no longer have to develop/maintain their proprietary formats. Inability for most vendors to compete if closed standard such as OOXML, since it would be almost impossible to implement without a clear and unambiguous definition.

**Question 28. State mandates multiple document formats for state agency use:**

For state: If only open standards (such as ODF and PDF), could be reduced costs. If includes closed standards such as OOXML, increased costs due to expensive software and hardware, and greater difficulties of conversion.

For citizens: Reduced costs and greater convenience if open alternatives such as ODF and PDF. Could still be reduced costs if ODF and OOXML accepted, as ODF would allow reduced costs.

For vendors: If OOXML is accepted, this would reduce the demand for ODF, if only due to the inertia of citizens in software choices. This would reduce profitability for vendors supporting ODF, but other clients in the market would still favor the open format, so vendors would continue to support ODF in place of proprietary formats. (Using ODF would still simplify their development process.) Of course, Microsoft would benefit from this choice, as the state would be a profitable client. If ODF and PDF are used, the situation for vendors would be much as if only ODF were chosen.

**Question 29. Which option would be most cost-effective?**

In terms of benefits, choosing both ODF and PDF would be the most cost-effective. For the state, choosing PDF as well as ODF would be not much more expensive than ODF alone. (No need to add expensive software, easy generation of PDF from ODF with at least Openoffice/Staroffice.)

For citizens: ODF and PDF would be preferable, as many end users prefer to receive PDF documents, to those in an editable format. But an editable format is necessary for correspondence and forms to be submitted, for example. This choice would not require acquiring expensive software or hardware.

So both ODF and PDF formats would be preferable for most citizens.

For vendors: Avoiding OOXML would be preferable almost all vendors (with one big exception). This would enhance the trend of conversion to ODF. The PDF format would not detract from this trend.

Thus ODF only or ODF + PDF would be equally acceptable for most vendors.

**Question 30. Difficulty of adopting multiple standards:**

Yes, if the standards are not openly and unambiguously defined. Even if open and well-defined, certain elements of the standards may not lend themselves to exact translation.

An illustrative example from another domain: Before the rationalization of British-based units of measure in 1960, the inch in Britain, Canada, and the US differed, by less than 1%. The Canadian inch was exactly 25,4 millimetres, the British and US measures slightly more and slightly less, all defined in terms of millimetres. The Canadian measure for the inch was accepted as standard. (Other compromises were made for other measures, reducing discrepancies between the 3 countries systems.) Before the standardization, with normal rounding, for a length of a few inches, one could not translate exactly a length of a few inches from one country's measure to another. After standardization, there was no need to translate.

OOXML has particular problems. It attempts to incorporate legacy proprietary microsoft formats, making reference to an application as arbiter of the correct format. It incorporates errors in calculations, because these errors exist in current microsoft applications. Since the proprietary formats referred to in the definition have not been openly defined previously, the format has the unenviable task of publicly defining these formats for the first time, without succeeding to completely define them.

**Question 31. (n/a)**

**Question 32. (n/a)**

**Question 33. Why a vendor cannot/ will not directly support OOXML:**

- ODF is a well-defined open format for the same type of documents, already approved by ISO. OOXML is thus redundant.
- OOXML defines logical errors in the spreadsheet format to maintain compatibility with Ms-excel
- OOXML format is overly complex, with much of it is defining proprietary formats previously publicly undefined, with references to applications to complete the definition.

- For OOXML to become widely accepted by vendors, it must become well-defined without reference to specific applications, correct internal logic errors, and ODF must cease to exist.

**Question 34. Why a vendor cannot / will not directly support ODF:**

- There is no reason why a vendor CAN not directly support ODF, as it is a well-defined open format. Compatibility with older formats can be achieved by conversion in the vendor's applications. The vendor doesn't even have to reveal any details of the older formats.

- There is a short-term economic advantage for a certain large vendor to NOT directly support ODF: It will be more difficult for clients currently using the vendors' applications (and formats) to switch to other vendors' software, as they will have considerable difficulty converting OOXML documents.

- In the longer term, any vendor that refuses to support ODF natively will progressively lose clients.

**Question 35.** Unfamiliar with any specific studies on the conversion from the Wordperfect suite to Microsoft office.

However it is worth noting that in the spreadsheets, the syntax Wordperfect suite's Quatro pro (which matches that of Lotus 1-2-3), differs moderately from that of Microsoft office's Excel. Openoffice/Staroffice has chosen to use the syntax of Excel, which will ease the conversion of spreadsheets. In any case, Excel spreadsheets seem to display very well when imported into Openoffice.

Microsoft office text documents without special effects import readily into Openoffice, and always display identically when converted in either direction, if they do not contain graphics. Microsoft office presentation documents seem to display very well when imported into Openoffice. Of course, if Openoffice's drawings module is used, one will not be able export it to Microsoft office, as Microsoft office has no equivalent module. Microsoft Access has no exact equivalent under Openoffice, but Openoffice's database module can link to Microsoft Access. (Via a specific option in the menu.) Via the menu can also easily create plain text or spreadsheet databases, or link to MySQL or Oracle databases, as well as configure several general-purpose database link protocols.

**Question 36.** Having managed many software migrations, I would suggest starting with one or two agencies of a less critical nature, migrating to a supported ODF agent such as Staroffice. (A somewhat enhanced version of Openoffice, with support from Sun Microsystems.)

This will give internal support personnel experience with migration. Be sure to include some internal support personnel with some experience with macros, or at least some version of Basic. (Don't worry, Openoffice/Staroffice is well documented by context-sensitive help buttons on almost every page.)

After this initial migration, you should be aware of potential pitfalls, and probably be willing to continue without external support. Most significant potential problems would be due to your particular needs, as Openoffice/Staroffice are relatively mature products with few problems.

One very nice feature is the ability to export any (readable) document in PDF format. It could be useful to hire temporary staff to assist during the conversion, especially if it is decided to convert immediately all existing documents to ODF. This temporary staff would better be allocated to non-conversion tasks, to let the regular staff verify the correct appearance of converted documents. The appearance is the most likely problem to be encountered in conversion.

Alternately existing documents could be converted only when used. Immediate conversion has the advantage of reducing the period when staff must deal with 2 different applications for equivalent tasks.

I find it very hard to believe that any study could legitimately find that ancillary costs increase after conversion. It is worth noting that Word files generated by Ms-office sometimes do not display correctly in Microsoft's Wordviewer, which is supposed to accurately display Word documents for those who do not have Word. Especially those containing graphics, which sometimes overlap text. The Openoffice writer module has many ways to anchor graphics, to overcome display problems (and facilitate positioning), and I have never seen an Openoffice-generated writer document have problems with text overlapping graphics. (Although the best ways of anchoring graphics tend to display poorly when converted to .doc format.) From my experiences, I would be inclined to think that efficiency would increase with Openoffice, reducing costs in addition to saving licensing fees. Also, when changing from one ODF application to another, which will inevitably happen sooner or later, there will be no need to "convert" documents.

**Question 37. Validity of studies showing saving after converting to ODF:**

In a word, Yes. The savings in licensing costs is obvious. The long-term savings when one changes the agent used, through lack of need to convert, is obvious.

Another advantage is in collaboration, which should result in cost savings. User-1 can generate an ODF document on Linux. And send it to user-2 on a Ms-windows machine, who modifies it, and returns it to user-1. User-1 user opens the document (running on Linux), and with all the formatting and appearance preserved, sees only the specific changes made by user-2. I have this experience almost every day. (Sometimes I am both user-1 and user-2, as I work in both environments.) Interestingly, while the applications display the document identically, the application used can display with its own theme. In my case, using Openoffice, it displays with one theme on Ms-windows, and a Gnome-compatible theme on Linux. I could just as easily use the Koffice suite on linux, with the same results. Or the IBM-Lotus or Google suites on Ms-windows.

To put this in the context of the state, an ODF-based form could be put on the web, downloaded by a citizen, who fills it in and returns it by email. The original form could have been generated using the IBM Lotus on Ms-windows. The citizen might be using Openoffice on a Mac. The form when received could be processed using Koffice on a Linux workstation. And everything would just work !

Tell me. Where do you see increased costs?

**Question 38.** I strongly doubt that any independent study will show reduced costs by migrating to OOXML from ODF.

Microsoft may succeed in finding niche cases showing savings. But the State should also consider the costs for its citizens, and the accessibility of services provided. The increased collective costs to license OOXML software is obvious. For citizens not interested in using Ms-office-vista, or unable to pay the price, the cost is inaccessibility to state services. For the state, the cost of licensing is obviously higher. And of course, Microsoft, being the only OOXML player, will be free to change the standard at will, thus adding another long-term potential cost. When Microsoft ceases to support OOXML (which is inevitable, as no company or product lasts forever), there will be the cost of conversion. (Probably to ODF.) (I think the most likely case is that Microsoft will finally support ODF.)

**Question 39. Assistive technologies:**

I have no special knowledge in this area, but it seems reasonable to procure a means of access for users with adaptive mechanisms, such as braille interfaces. Openoffice already has some configuration options for those with minor disabilities, such as high contrast display, as well as being able to use any such options offered by the operating environment.

**Question 40. Which format will better facilitate access through use of assistive technologies?**

Currently:

I don't know about other formats, but ODF has a non-normative guide for accessibility, including assistive technologies. According to the guide, most assistive technology functions are provided by the operating environment, rather than then the ODF agent. These functions are generally available for Ms-windows, Mac, Linux, and other UNIX variants. ODF has the capability of passing any information needed to activate adaptive technologies. The guide gives numerous suggestions for the development of ODF agents enabled for adaptive technologies. In practice, some assistive elements are directly configurable in Openoffice, such as high-contrast or special color schemes. Considering that the operating system provides most adaptive functions, I would be surprised if the choice of format has much effect on the adaptive functions available.

In the future:

Since ODF is capable of passing any information needed to activate adaptive technologies, ODF will be capable, without changes to the format, of adjusting to any new adaptive technologies. Other formats may require modification.

**Question 41. Adoption of ODF acceptable if conversion to other formats available to allow usage of assistive technologies?**

It would be perfectly acceptable, since it is important to facilitate assistive technologies. However any assistive technologies not available would only require modification of the ODF agent, since ODF itself is capable of passing any information necessary. Since a number of ODF agents are open source, I would expect that any such shortcomings would be quickly rectified.

**Question 42. Should state engage in initiative similar to Massachusetts MOU, "to design, procure, certify and develop training for software that is accessible to people with disabilities" and "establish unit devoted to accessible technology"?**

Excellent idea. Even better, the state should try to cooperate with Massachusetts so as to avoid duplicating efforts, avoid reinventing the wheel.

One very good way to assist in the development of such software is to donate (money or computer personnel) to open source projects for that purpose. The advantage of contributing to open source projects is that the source code is readily available, normally at no cost. Thus once the software is developed, the same functions can be readily replicated in other applications, at little further cost. Thus the investment has much a greater impact.

**Question 43. Stakeholders most conversant with issues related to document formats and assistive technologies.**

I don't know who is MOST conversant, but a very good reference is available at [http://docs.oasis-open.org/office/office-accessibility/v1.0/ODF\\_Accessibility\\_Guidelines-v1.0.odt](http://docs.oasis-open.org/office/office-accessibility/v1.0/ODF_Accessibility_Guidelines-v1.0.odt) A search on Internet will give many other references. Wikipedia < <http://en.wikipedia.org> > is another good source.

**Question 44. Setting office suite software format standard premature?**

Not at all. The idea is to define the present and an avenue for future development. It is more than a little naive to think that office suite software will move exclusively on line. Sure, certain types of users will find the on line model compelling, and it does provide an interesting revenue stream for vendors.

But on line processing invites all sorts of security problems, which are best avoided by keeping processing as local as possible. Local processing also means, for the same resources, faster processing. So there are compelling reasons for most software suite users to avoid on line office suites. With the arrival of open source and free or low-cost office suites, cost is no longer an impediment to owning a sophisticated locally processed office suite. It only takes a significant portion of users to make a norm such as ODF relevant. We are already there.

**Question 45. Setting office suite format standard premature because:**

**a) ISO standardization in process for OOXML?**

Not at all. With ODF, we have a well-defined and viable standard. Why do we need 2 standards for the same thing? Especially if the second standard is NOT well-defined.

**b) ODF format undergoing revisions?**

All standards undergo revisions. It is the sign of a viable standard. ODF is relatively new, so enhancements are important. For instance, the assistive technology guidelines was a non-normative part of ODF 1.1 (See reference in answer 43.) Note that HTML is a much older standard, and is still undergoing important revisions and enhancements. ODF is a layered standard, where one can be compliant to a specific basic subset, a specific larger subset, or to the entire standard. Similar to the "Transitional" and "Strict" variations of HTML.

**Question 46. Appropriate timeframe for CIO/OFT recommendations?**

I'd like to see them before February 2008 (and ISO's OOXML ballot). We don't need 2 standards for the same thing. Microsoft should simply wake up to the fact that there is no point combining new XML formats with legacy binary formats, when it suffices to simply import the legacy document and save it the new XML format - which should be ODF.

The nature of the questions gives the impression that one is close to decision.

I hope my input helps.

**INDIVIDUAL # 46:** [INDIVIDUAL'S NAME REDACTED]: *Tuesday 12/25/2007 11:56 AM*

I have just read your request for information on: <http://www.oft.state.ny.us/oftnews/erecords-study.htm>

My country, the Netherlands, has just decided on a new set of policies concerning document standards and software procurement for the public (i.e. government) and semi-public sector (i.e. hospitals, school, etc).

[IDENTIFYING INFORMATION REDACTED] On december 12th the Dutch parliament unanimously approved the policy and implementation will start next month.

The goals of the Dutch policy are improving interoperability in the public sector, lowering dependence on vendors and improving the functioning of IT-markets and thereby supporting Dutch knowledge- economy.

This will be achieved by a series of measures of which universal and mandatory adoption of ODF by the entire public sector over the next 18 month is the most important one. Another major measure is to give preferential treatment to opensource software in cases of comparable functionality in all tax-funded procurements.

Because of great international demand the Ministry for Economic Affairs has decided to release a first English translation for download here: <http://appz.ez.nl/publicaties/pdfs/07ET15.pdf>

The Dutch government is looking into the possibility of releasing the entire document as a Creative Commons licensed work so other can re- use it for non-commercial purposes.

Norway has formulated comparable policies:

<http://virtuelvis.com/archives/2007/12/norway-mandates-html-pdf-odf>

As have Belgium and Denmark. Various other European countries (Spain, France, germany, Italy) are studying comparable measures.

ps: for background information on the development of the policy please see [REDACTED] presentation at the CCC IT conference in Berlin on December 29th:

<http://events.ccc.de/congress/2007/Fahrplan/events/2387.en.html>

([REDACTED] full presentation and a video registration of the talk will be online on or after the 29th)

**INDIVIDUAL # 47:** [INDIVIDUAL'S NAME REDACTED]: *Thursday 12/27/2007 5:45 AM*

## ELECTRONIC RECORD POLICY FOR NEW YORK STATE - DRAFT

All New York State Employees, except legislators and their staff and full-time professors, shall print two hard-copies of every electronic record created using their state-assigned password. All hard copies shall be archived for a minimum of three years. The first copy shall be retained in a personal chronological file. The second copy shall be retained in a topical file.

Hard copies shall be created immediately after sending the electronic record. In the case of portable computers used in the field, hard copies shall be created and archived in the employee's office, normally within one week.

Hard copies of electronic records regarding criminal investigations, medical records, and personnel matters shall be marked "CONFIDENTIAL" and kept in separate chronological and topical files within a locked file cabinet. The New York State Inspector General or his designee, sworn to confidentiality, will randomly check confidential files without prior announcement to insure ONLY that the contents correspond to the definition given in the first sentence of this paragraph.

Meetings and telephone calls, other than informational calls, shall be summarized with soft- and hard-copy memoranda.

Any New York State Legislator, the New York State Governor, or their designees may demand, without prior announcement, hard or soft copies of any New York State employee's chronological file. Transmittal shall be made within twenty-four clock hours if the employee is not traveling. Quality of that file may be considered with regard to promotion or retention to the extent that the employee's job description indicates responsibility for electronic records. Furthermore, timeliness of responses and thoroughness of responses to incoming electronic records must be considered. Initiative in proactive creation of queries must also be considered. The employee's supervisor shall not be responsible for subordinate employees' chronological files. The employee must protest in writing to a Legislative Oversight Board if the supervisor does not guarantee adequate time for record-keeping within working hours, operating printers with ink and paper, and desk space in a quiet and private room conducive to literary production.

The employee shall not engage in paid or unpaid work for other individuals or institutions while in New York State offices or while on officially-approved field visits.

**INDIVIDUAL # 48:** [INDIVIDUAL'S NAME REDACTED]: *Friday 12/28/2007 11:39 AM*

Thank you for requesting public comment on this very important public policy decision. I am a professional software engineer for an Albany-area media server and control system company, living and breathing how technology impacts our daily lives. Those choices, and the laws behind them, influence our artistic expression, products and business opportunities. The ramifications are more diverse and profound when considering the needs of the public in creating, securing, storing, sharing, studying, and analyzing the product of our government. One must also consider how state-sanctioned technology and standards will impact their uptake within the state and the world at large, as no boundary exists for the sharing of ideas.

I must state that I am a long-time user of Linux and other open source software, and approach this inquiry from that perspective. My wife and I use Linux on all of our home computers, and all of the products that my company builds are built on Linux and many other open source tools. I rely on all of this software and its interoperability every day, for every critical task. My professional

and personal life intersect when I spend hours reading up on analysis of technologies, devices, and software design.

I have not read either the OOXML or ODF specification, though I have read many critiques of both formats. What I understand of their relative strengths comes from a variety of analysis of OOXML which point to serious flaws in design, notably through a lack of specifications. Serious ethical questions have also been raised regarding the influence of voting on OOXML specifications in standards committees throughout the world. These concerns give me pause, but I urge you to consider a more fundamental principle on which to base your decision.

The choice before you is not one of software or technology, but one of language. The specifications we use define what we can express, how we can express it, and how it can be understood. No different than our familiar written languages, a good specification allows one to open up new avenues in communication, sharing their ideas in new ways. Our world is filled with simple yet powerful standards that have redefined how we interact: TCP, HTML, email.

In making this decision, I urge you to consider that by the very nature of the data itself, we cannot predict the needs and opportunities which will arise as the product of government is recorded, analyzed and expressed through the decades. The needs of the blind and the deaf, lawyer and politician, accountant and surveyor, can only be met when there is every opportunity to create products and services which empower them to use this data. Public, open technologies have a proven history in meeting these needs now and in the future, and choosing the ODF specification will ensure that all New York State residents will have access to these documents in perpetuity.

**INDIVIDUAL # 49:** [INDIVIDUAL'S NAME REDACTED]: *Friday 12/28/2007 6:05 PM*

I realize that a comprehensive assessment of New York State's electronic records policy must confront a daunting variety of requirements, constraints, and considerations. Fortunately, one of those considerations -- the format of data to be preserved and manipulated -- is straightforward, if only it is allowed to be so.

Vendors with proprietary solutions are highly adept at pretending to abide by open standards. The result is invariably a technological bait-and-switch maneuver. Not all such attempts succeed, but many do, and some vendors are breathtakingly good at it. Each successful case represents an operational disadvantage to the user. That means you.

You face a basic choice between Open Document Format (ODF) and Office Open XML (OOXML). I would argue that the relative merits of your choice are clear:

**ODF:** ODF is entirely usable today for all purposes claimed. Rough edges will doubtless be exposed for new uses in the future, but the open, transparent nature of this standard ensures that any mismatches will be brief and easily resolved. Vendor neutrality and interoperability will, of necessity, be preserved.

**OOXML:** OOXML cannot genuinely be stabilized and made unambiguous, even by its creators. It is a proprietary moving target masquerading as an open standard. A future based on OOXML will be a future of translation utilities, workarounds, and reformatters that, at best, sort of work most of the time. Fragile relationships between future documents and future applications will be the norm. It will be a de facto requirement to stick with a de facto monopoly vendor. You will be forced to live with failures of interoperability, even between successive

versions of the monopoly vendor's own applications. Novel security vulnerabilities will find their way into OOXML documents on a regular basis.

We have all lived for too many years with flawed technologies in a flawed market, when simpler and fundamentally more desirable alternatives are available at little or no cost. Your study can tilt real-world usage toward open and robust solutions.

Anyone, Microsoft included, can implement ODF, and if Microsoft had to support ODF by the book, they would do so. Microsoft can, of course, implement OOXML better than anyone else, but no one, Microsoft included, can fully implement OOXML.

Standards matter, and New York State is now in a good position to establish sensible, tractable standards for itself and others. Adoption of OOXML would, at heart, be a minor variation of the status quo and would preserve the technological disadvantages and difficulties that led to your study in the first place.

I urge you to opt for ODF.

**INDIVIDUAL # 50:** [INDIVIDUAL'S NAME REDACTED]: *Friday 12/28/2007 7:48 PM*

Re: RFPC # 122807, A study concerning electronic record policy for New York State.

Given the detailed questions provided at [http://www.oft.state.ny.us/oftnews/FINAL\\_e-record\\_study\\_RFPC.Part%20II.pdf](http://www.oft.state.ny.us/oftnews/FINAL_e-record_study_RFPC.Part%20II.pdf)

**Question 1.** The distinction may not be useful in that all of this is "data" at the machine level and must be stored in an accessible fashion. A more effective separation would be to say that the issue is storing data, some of which describes documents or records. This leaves the focus on maintaining an effective data storage technology.

**Question 2.** Daily, ancillary, and historical access to records are largely a matter of physical storage format (e.g., disk vs. CDROM jukebox). The overriding issue in all forms of access is that the data is decipherable by the currently running software.

An excellent example of this is NASA, which has terabytes of data which can neither be replaced nor read: the unmanned spaceships are no longer working but the readers for data formats they transmitted have been lost.

The States issue in saving this data will be finding a vendor-neutral, standards-based format for storing the data so that it can be used. Using a data format that supports records, documentation, and historical 'data' (e.g., census numbers) will be the priority.

The issue of vendor neutrality is central here: all formats managed by specific vendors are subject to change at their whim. Microsoft has been especially egregious on this, but all vendors have left some number of their customers stranded at one point or another. The legal necessity for recovery of data makes this especially important for the State's records; the mass of data will make conversion to arbitrary formats over time difficult if not impossible. As a result, finding a format which is likely to be stable is a necessity; and standards-based, stable formats are supported by all of the vendors.

XML and ASCII are two examples of these formats: ASCII generated in the 1970's is still perfectly readable today (given the low-level formats are known). Unicode is likely to be around for a long time.

The ultimate goal of standards organizations is to ensure backwards compatibility and long-term usability; the goal of software companies is to sell software that reads the current format. This leaves the former in a better position to describe data formats the State will be using long term.

**Question 3.** On the issue of ODF or OOXML, the main issue will be whether the State has the option to choose the encryption method used to handle its data. The advantage to ODF is that being open and well-described, it can be handled by any collection of software available.

OOXML is a proprietary Microsoft ("MS") product. It does not support any standard and its use is restricted to Microsoft software. The history of MS security shortcomings is well known, and I won't re-hash it here. The point is that an open standard offers better portability of data, which then offers the State a better choice of enciphering tools, which eventually gives better security.

**Question 4.** The main advantage of ODF to FOIL would be a greater variety of available software to read the data. As needs and applications change, ODF has a vested interest in keeping their format up to date and reliable for handling the new data; OOXML has a vested interest in seeing that software reading it is written by or licensed from MS.

This will be especially true for long-term storage, where MS has an especially poor record of backwards compatibility and there are legal issues in not being able to access the data.

**Question 5.** The single most important issue in maintaining government control of data will be ensuring that the State has control of the mechanism for storing, retrieving, and using the data. Standards-based formats give the government more control over the data by offering choice between packaged or in-house applications to read the data and stable definitions of the data being managed. Without published, stable standards for the data the government will be at the mercy of whomever writes the core software, and changes to that format can be difficult to deal with.

Although certainly difficult to employ, MARK has persisted in part because it is stable and well described: anyone wishing to wade through the documentation has access to the complete definition. Access to the data can be maintained by the State using in-house or contract programmers with access to the format has allowed a debate on its merits and helped lead to better formats -- along with conversion utilities.

Net result: An open format offers the State its best chance to control any data by giving us the choice in how to process it.

**Question 6.** This definition is a good start: the point is to ensure that no one vendor can hold the format's users, effectively, hostage to using it in the first place.

Central to the utility is open access to the doc's at no or nominal fee and open access to the irrevocable standard. These are what guarantee that the user can select their own applications for managing and using the data. Without these a vendor can restrict access to the knowledge or mechanism for managing the raw data.

**Question 7.** N/A

**Question 8.** N/A

**Question 9.** N/A

**Question 10.** Failing to support the standards will cost the State, if not now then later.

**Question 11.** Long-term storage is one obvious issue: if the state does not adopt a published, well-supported standard for this data there is no guarantee that it will be accessible in the future.

**Question 12.** The State needs to concern itself with electronic records interoperability at all levels, from the low-level binary format on disk drives to the printers and screens. This gets back to maintaining control: if the State lacks a choice in how the data is managed then it lacks control over how to manage the data. Using proprietary formats reduces the choices to what any one vendor may or may not choose to support.

Other than office suite documents, the State should concern itself with medical records, census data, minutes of the legislature, and the body of laws enacted by the state: All of which need to be accessible long term for the State to function.

**Question 13.** The approach should be to first adopt formats that are open, stable standards. XML will evolve over time, and -- given the history of computing -- be replaced with something else. The advantage to using XML for lower-level formats is that the changes will be gradual, thought out, and have specified migration paths for existing data. This last point cannot be stressed too much: the State must be able to utilize new standards as they evolve and adapt to new standard as they are adopted in other places. An open, standards-based format will give us the ability to migrate existing documents and applications to the new formats as they arise.

The most specific mechanism I can describe is for the state to adopt only Open Formats (see item E) and purchase only software that fully supports the chosen standards. This doesn't mean that the office suite software cannot read any other formats, just that it support storing and retrieving the data in those formats. As with MA, this will force the software vendors to support the open standards.

Any inter-state agreement between agencies on open formats would also help. These could include exchanging inter-state data and documents (e.g., tax records) using those formats.

**Question 14.** It is reasonable: starting with a specific, common example and moving outwards from there is a workable approach. The process of evaluating formats for "office" data should provide information that helps jump-start any further investigation.

**Question 15.** The study should be looking at ways to maintain effective access to State documents and records; which I think it is.

**Question 16.** My own company stores nothing in MS '.doc' format: the risk of being unable to read it going forward is too high. All legal documents are kept as PDF (vendor specific but intended as interoperable); at some point we'll probably migrate to ODF.

Aside: This example is important. Because both PDF and ODF are well-described we CAN convert between them easily.

Our images are stored as JPEG and PNG depending on use (JPEG is an industry-wide standard and unlikely to change on a whim), along with source data in vendor format (largely Nikon's '.nef' format).

Most data on disk is some form of flat ASCII or XML (used with bioinformatics processing).

Anything sent out of house goes as PDF, Postscript, standard image format, or flat ASCII (e.g., this message).

**Question 17.** None. There simply is no data that can be stored more effectively or flexibly in this format.

**Question 18.** (numbered list)

1. Exchanging data between agencies within the state and inter-state or with Federal agencies. Use of open standards, from TCP/IP to SMTP to ODF enhances range of places where data can be sent and used.
2. Long term storage and recovery is enhanced with open standards. The standards agency has a vested interest in seeing that their format accommodates current usage and grows in a stable way. This includes ensuring migration paths for existing data and maintaining backwards compatibility of the current standard with previous versions.
3. Access to the widest range of users outside of the State. Users on applications other than MS products will be able to read ODF data; users who use MS products will be able to read ODF data. Given the support for ODF in a number of places (e.g., MA) the likelihood of vendors continuing to support it is good. This not only includes the ODF format in general but older versions of the standard (i.e., older data on State disks).
4. There are also licensing issues with any MS product. With ODF itself there are none, and many of the tools or libraries used to process ODF are open source. This obviates any licensing issues for the State, with their attendant licensing fees and costs of monitoring the licenses.

**Question 19.** I don't know enough about the State's needs, but my own include the ability to operate with documents produced on a variety of platforms, including MS, MAC, Linux, and Unix platforms. Using ODF to exchange and archive the data I can work with documents from all of these sources.

I can also pre- or post-process data in ODF format using my own tools. This includes filtering the content or spell-checking with external checkers (necessary for some of the more specific language I use in bioinformatics). A simple Perl program can easily extract and rebuild content for ODF files. This is because the format is designed to support simple access with a variety of tools.

**Question 20.** Long term preservation requires long-term standards. There are .doc format files that cannot be read between single versions of MS Office or MS Windows. Please see the previous answers for why I think the standards-based format is necessary.

**Question 21.** The State can get any number of canned or open-source tools that easily filter ODF data (see also answer #19, above). With ODF you can also easily hire people to design any filters for searching, summarizing, or data mining the State's electronic content.

OOXML -- and MS in general -- do not support open standards. It is against their business interests to do so. Using MS' format the State will be restricted to using only MS tools for the projects; which assumes that MS will continually support all formats that have been made available over time. You might wish to believe that MS can do this in the future, in the past they have not (witness issues with single-version upgrades of Office and Windows).

XML is not perfect, but at least it is a well-known and -described product. It also one that the State can use without liability for having used patented material by mistake. ODF falls in the same category: not perfect but certainly the best choice from what is available today.

**Question 22.** Migration to ODF will require some rewrites of existing code. So will conversion to OOXML. The question going forward is which one of them is less likely to introduce major issues later on. The ODF software can largely recycle existing open-source projects' content without licensing issues for the State. In all likelihood the switch to a standards-based solution for the data will entail fewer, less costly re-writes in the future.

**Question 23.** N/A

**Question 24.** This would depend heavily on the standards body and how well accepted the standard is. The IEEE and ANSI standards are a good starting place (e.g., ASCII), with an additional review of how well-used the suggested standard are. The problem with de facto standards is that they usually are not standards at all (e.g., 'CSV' format files). Lacking a published standard with a body assigned to maintain it there the possibility of mis-information is greater and the cost of maintaining software to support the de facto standard is usually higher.

The single largest element is whether the standard is used (e.g., IEEE networking standards) and whether the standard is appropriate for State use (i.e., describes the data being formatted with it).

**Question 25.** Open standards always increase competition in the marketplace. Closed standards allow a monopoly to the company who owns the standard, lack of access prevents anyone else from creating truly competitive offerings.

Look at the internet: ethernet, TCP, HTTP, HTML standards have provided a relatively level playing field for anyone who wants to try and write a product for serving or viewing web content.

This openness of standards only helps the State since any of the HTTP or HTML products available can support any server that delivers HTML via HTTP (which excludes MS web servers since they deliver "MSHTML" that doesn't quite match the standard).

Being able to use any variety of products even with the State is an advantage: in most cases users can switch web clients without the State having to convert anything since the HTML and HTTP content is common to all of the products. This fosters competition since pilot studies or switchovers do not require re-tooling: anyone who can build a better mousetrap is welcome to catch the mice.

**Question 26.** Enforcement would be largely a matter of pre-filtering the content. The main issue would be allowing NYS Archives to reject non-compliant content -- and giving it the option to run standardized conversion programs where appropriate. As an example, there a standard PDF -> ODF conversion utility might be run, but if it fails to convert the document it would be sent back to the originator.

The important thing is that some part of the State government be given authority to reject products or documents that do not pass a standard, published test for "acceptable" documents.

**Question 27.** Converting to a single, standards-based format would obviously incur some cost. The benefit would be lower costs going forward in the form of better access to archived data, simpler data exchange within the state and inter-state agencies, and simpler migration of data between versions when the need arises.

The important thing here is not that the format is a single standard, but that it is an open standard which is subscribed to by more than just this state. This gives the standard enough weight that vendors will tend to support it as a matter of fact (e.g., as with Postscript or PDF formats).

**Question 28.** Multiple formats add a cost of frequently having to have a single document in multiple formats, having to convert it in multiple formats over its lifetime, and perhaps having to license products for handling all of the formats at once.

Most of these arguments are similar to why we support a single currency in the US and why the Euro benefited EU members.

**Question 29.** A single open format would allow any vendor or individual to access all of the State's records. The State would spend less money converting between formats to use the content, less time and effort exporting documents to various formats for use, and fewer interoperability issues within the State agencies trying to exchange or mine existing data.

**Question 30.** Licensing and documentation limitations of OOXML prevent its use in most cases. Looking at MS today, I cannot find any published standard for OOXML. This would prevent me from using it even if the licensing issues did not arise. There are no cases to date where MS has fully released their internal standards to the public (see both cases of U.S. v. Microsoft for background).

**Question 31.** N/A

**Question 32.** N/A

**Question 33.** My software company heavily uses open source for Bioinformatics. I cannot afford the licensing required to run Windows, nor can I afford to upgrade the machinery to support current Windows versions. I also cannot afford the cost of purchasing the documentation required to work with the MS formats.

I also cannot release software that depends on MS libraries to my customers, many of whom do not use Windows on their data processing systems. This would require me to have separate Windows and \*NIX products which I cannot afford to do.

**Question 34.** Vendors who do not support ODF in any way will have to find ways of formatting their data as ODF. This is an impediment to their adopting it, but it is also an issue with switching to a more recent version of Office Word format. It might not be that difficult for most vendors to handle ODF since there are ODF to postscript and PDF converters; nearly all office-document vendors support both of these. Ripping the guts out of an existing utility and making it speak the vendor's internal format is not all that convert in and out of ODF has already been done and extensively documented.

**Question 35.** Good example: you have data in one format, need it in the other, need to run an external conversion utility on all of it to get from here to there.

One of the main expenses isn't running the conversion: it's finding all of the documents and getting them where they can be converted (e.g., getting all of the CDROMs out of all of their nooks and crannies). After this, however, the SysAdmins only have to deal with managing conversion w/in the one format.

**Question 36.** One approach is to simply mandate that all new documents would be in the single format for a few years. During that time software would have to deal with existing formats and save everything as ODF (i.e., much of the conversion would be on the fly). During that window some group w/in the State would have to bulk convert documents, archiving their original contents. This could happen by department, geographical region, whatever.

The point is to avoid shutting everything down while a mass conversion is handled.

**Question 37.** N/A

**Question 38.** N/A

**Question 39.** The same features of open standards that foster competition also foster invention. By attempting to encapsulate the meaning of data, XML actually makes it easier to a single, open format would simplify the process.

**Question 40.** N/A

**Question 41.** N/A

**Question 42.** N/A

**Question 43.** N/A

Note: I don't have any specific experience for answering the previous four questions.

**Question 44.** The standards will always be evolving to accommodate new uses. Picking one open standard now will only lessen any eventual costs of conversion. If we wait for the "final" standard we will simply never adopt any.

**Question 45.** Selecting the standard today would be reasonable since any changes after the selection is made will simply be water under the bridge: whatever is there when the standard is developed will be the issue. The State is a whole lot more likely to adopt a stable standard with ODF than OOXML given the current state of standards org's dealing with both.

**Question 46.** Given that the format is what's being chosen (vs. applications), adopting that could be done w/in the year. Switching over to software that can write the chosen format would take probably another 1-2 years statewide. There is no reason to insist that the software handle only ODF, just that it be configured to write that format by default. In this case all documents that are handled would be converted as necessary over a 1-2 year period with bulk conversions of archived data over a 5-10 year period.

**Question 47.** Given that many products would be able to write ODF by default today, not all that long. A single product cycle should be sufficient for existing products to write ODF as a default option. Given this, probably 1-2 budget cycles to get the software converted over.

**Question 48.** There is certainty when dealing with open standards: vendors cannot sue for the use of a public implementation. This is a major advantage to using open standards. The CIO should be validating that ODF is an open standard, after that the issue is not a concern.

In contrast, MS has sued vendors for using technology it advertised as a "standard", though not an open one. This would be a serious concern for the State using even public standards based on MS proprietary offerings. There might be cases where MS would consider use of what the State depends on as an open standard as their own. Even if the State could sue to rectify the issue, even having to deal with the problem is a significant risk.

In my own case, I would prefer the open standard.

**Question 49.** The State would be immune from intellectual property issues in the case of ODF since the standard is open (see the definition in E of the RFC). There might be issues for the State using a vendor's product in ways it was not licensed, but that is not an issue from dealing with the format per se.

**Question 50.** The best way for the State to protect itself is using open source standards for the documents. In much the same way that no one can sue the state for using HTTP or SMTP, there simply is no grounds for suit for the State dealing with an open standard.

**Question 51.** Bit of a broken record here, but insist on truly open standards for the record formats. In the same way that libraries cannot be sued for utilizing the Dewey Decimal or LOC catalog systems, the State cannot be sued for storing records in XML (or ASCII for that matter), or measuring in Metres. These are open, public standards intended for unrestricted public use.

**Question 52.** Using an open format for the data will enable a wider range of tools to mine, summarize, and process it.

**Question 53.** For most documents PDF/A is a usable output, though perhaps storing the documents in a richer ODF format that supports lexical processing might be helpful for discovery issues. Nearly all of the documents distributed by the State could be sent out in PDF/A.

**Question 54.** N/A

**Question 55.** PDF/A might be a good choice for storing documents that are not expected to change over their archival lifetimes (vs, say, spreadsheets). There is plenty of software that handles it already and the standard is open.

**Question 56.** The biggest shortcoming of OOXML is the history of MS litigation and their "embrace and extend" mentality for standards. This has shown up in multiple suits by the U.S. against them on trade practices.

**Question 57.** Any of the open formats will be something of a horse designed by committee. Proprietary systems will advance faster and incorporate newer features sooner. Open standards will be more stable and have hooks for extending them to meet your needs.

I apologize: I simply do not have time to finish any more of this. Thanks for the opportunity.

**INDIVIDUAL # 51:** [INDIVIDUAL'S NAME REDACTED]: *Wednesday 1/2/08 1:27 AM*

This is why we need to adopt truly open document formats, as specified in the ODF standard.

(The OOXML standard was created by Microsoft to gum up the Open Format works and keep people locked in to MSFT products).

Another lesson from this is that Office 2007 (the next version of Office for Windows) also introduces new formats, that will also eventually replace the .doc and .xls and .ppt formats in which zillions of files are currently stored. Think about all those public documents created by governments around the world --- they are hostage to Microsoft decisions like this one:

*"In Service Pack 3 for Office 2003, Microsoft disabled support for many older file formats. If you have old Word, Excel, 1-2-3, Quattro, or Corel Draw documents, watch out! They did this because the old formats are 'less secure', which actually makes some sense, but only if you got the files from some untrustworthy source. Naturally, they did this by default, and then documented a mind-bogglingly complex workaround (KB 938810) rather than providing a user interface for adjusting it, or even a set of awkward 'Do you really want to do this?' dialog boxes to click through. And of course because these are, after all, old file formats ... many users will encounter the problem only months or years after the software change, while groping around in dusty and now-inaccessible archives."*

Original article: <http://it.slashdot.org/article.pl?sid=08/01/01/137257>

**INDIVIDUAL # 52:** [INDIVIDUAL'S NAME REDACTED]: *Thursday 1/3/08 6:22 PM*

This article, recently covered by Slashdot.org highlights some of the concerns that myself and many of my fellow computer users share regarding OOXML:

<http://www.robweir.com/blog/2007/12/those-who-forget-santayana.html>

A real open standard should be chosen, not one which Microsoft is pushing purely to gain advancement in the market.

***"An Antic Disposition - Rob Weir, thinking the unthinkable, pondering the imponderable, effing the ineffable and scruting the inscrutable - Thursday, December 20, 2007***

***Those who forget Santayana...***

*It must have passed beneath my radar it when it first was filed in 2004, but it caught my eye recently when Andy Updegrove mentioned it in Chapter 3 of his book-in-progress, The War of the Words. I'm talking about Novell's November 2004 antitrust complaint against Microsoft, filed shortly after settling an different, OS-related, complaint with Microsoft for \$536 million. You can view the second complaint, which I'll call the "WordPerfect" complaint, here [PDF] on GrokLaw.*

*What is interesting to me, and why this "old news" is worth talking about, is the analysis Novell made in their complaint of Microsoft's treatment of document format standards. The concerns of 2004 (or 1995 even) are very similar to the concerns of 2007. Let's go through Novell's argument and see where it leads us.*

*91. As Microsoft knew, a truly standard file format that was open to all ISVs would have enhanced competition in the market for word processing applications, because such a standard allows the exchange of text files between different word processing applications used by different customers. A user wishing to exchange a text file with a second user running a different word processing application could simply convert his file to the standard format, and the second user could convert the file from the standard format into his own word processor's format. This, a law firm, for instance, could continue to use WordPerfect (which was the favorite word processor of the legal profession), so long as it could convert and edit client documents created in Microsoft Word, if that is what clients happened to use...*

*This is a good statement of the benefits of an open document standard. Note that Novell is not arguing that the benefit of a standard is to get information in or out of a single vendor's product, like Microsoft Office. The benefit is that a standard provides for interchange between any pair of word processors.*

*...Microsoft knew that if it controlled the convertibility of documents through its control of the RTF standard, then Microsoft would be able to exclude competing word processing applications from the market and force customers to adopt Microsoft Word, as it soon did.*

*Note also that Novell is not complaining here about Microsoft's control of the binary DOC format (and its many variations). Instead, what Novell complains about is Microsoft's control over the document exchange format RTF, or Rich Text Format, used in those days to exchange data between word processors. He who controls RTF, controls document exchange, controls vendor lock-in and has the sole means of improving the fidelity of document exchanges.*

*In fact, Microsoft claimed that RTF addressed this very concern -- document exchange in a cross-platform, cross-application fashion, as stated in the introduction to version 1.0 of their self-styled "standard":*

*The RTF standard provides a format for text and graphics interchange that can be used with different output devices, operating environments, and operating systems. RTF uses the ANSI, PC-8, Macintosh, or IBM PC character set to control the representation and formatting of a document, both on the screen and in print. With the RTF standard, you can transfer documents*

*created under different operating systems and with different software applications among those operating systems and applications*

*It should have been obvious at the time that vesting exclusive control of an interoperability interface in a single company was a bad idea. But I guess the world didn't realize what dealing with Microsoft meant. But we know better now. So why are we making the same mistakes in 2007?*

*Those who control the exchange format, can control interoperability and turn it on or off like a water faucet to meet their business objectives. I don't know how many people noticed the language in Microsoft's press release announcing their sponsored interoperability track at XML 2007 a few weeks ago:*

*In its approach, Microsoft strives to bring technologies to market in a way that balances competitive innovation with the real interoperability needs of customers and partners.*

*Let that sink in for a minute. Microsoft is saying that they need to balance interoperability and profit. (Their profit, not yours) They can't maximize for both simultaneously. They need to trade one off for the other.*

*Continuing with Novell's 2004 complaint:*

*92. The specifications for RTF were readily available to Microsoft's applications developers, because RTF was the format they themselves developed for Microsoft's office productivity applications. Microsoft withheld the RTF specifications from Novell, however, forcing Novell to engage in a perpetual, costly effort to comply with a critical "industry standard" that was, in reality, nothing more than the preference of its chief competitor, Word. Indeed, whenever Word changed its own file format, Microsoft unilaterally and identically changed the RTF standard for Windows, forcing Novell and other ISVs constantly to redevelop their applications. In this manner, Microsoft gave Word a permanent, insurmountable lead in time-to-market and made document conversions difficult for users otherwise interested in running non-Microsoft applications. Many WordPerfect users were thus forced to switch to Microsoft Word, which predictably monopolized the word processing market....*

*So, the RTF standard was just a dump of Word's features, done when and how Microsoft felt like doing it. As one wag quipped, "RTF is defined as whatever Word saves when you ask it to save as RTF."*

*This should sound familiar. OOXML is nothing more than the preferences of Microsoft Office. Whenever Word changes, OOXML will change. And if you are a user or competitor of Word, you will be the last one to hear about these changes. ISO does not own OOXML. Ecma does not own OOXML. OOXML, in practice, is controlled and determined solely by the Office product teams at Microsoft. No one else matters.*

*Consider that Microsoft has recently proposed over 1,700 changes to the OOXML specification, including fixes that presumably will be made into a future Office 2007 fixpack. Microsoft knows what these fixes will be. The Office developer teams know what these fixes will be. But if you are a competitor of Microsoft's in this space, do you know what these changes are? No. Microsoft has decided to keep them a secret, claiming that the ISO process allows them to withhold interoperability information from competitors in what they maintain is an "open standard".*

Further, the coding of Office 14 a.k.a. Office 2009 is well underway. Beta releases are expected in early 2008. But are file format changes needed to accommodate the new features being discussed in Ecma? No. Are they being discussed in ISO? No. Are they being discussed anywhere publicly? No.

*Is this how an open standard is developed?*

*My prediction is that the first time anyone hears about what is in the next version of OOXML will be when Office 14 Beta 1 is announced at Microsoft's 2008 Office System Developers Conference (ODC). Other vendors will not hear a word about the format changes until after the Beta 1 is already released. Not even Ecma will hear about the changes until after the ODC.*

*DIS 29500 is already obsolete, has already been embraced and extended. You just don't know about it yet. You weren't meant to know. In fact, pretend you don't know. Give Microsoft a big head start. They need it.*

*Further from the Novell complaint:*

*93. ...As in the case of of RTF, Microsoft forced Novell to delay its time-to-market while redeveloping its applications to an inferior standard. Because these standards were lifted directly from Microsoft's own applications, those applications were always "compatible" with the standards.*

*And that is the key, isn't it? By owning the "standard" and developing it in secret, without participation from other vendors, in an Ecma rubber-stamp process, Microsoft rigs the system so they can author an ISO standard with which they are effortlessly compatible, while at the same time ensuring that their products maintain an insurmountable head start in implementing these same standards. There is no balance of interests in OOXML. It is entirely dictated by Microsoft, and voted on, in many cases, by their handpicked committees in Ecma and ISO.*

*So much for Novell's complaint from 2004. I'm told that this is still case is suspended as of November, 2007, as the two parties pursue mediation. A status report on that mediation is due to Judge Motz by January 11th, 2008. Maybe we're hear more then.*

*Looking at this long history of standards abuse by Microsoft, in the file format arena and elsewhere, I'm drawn to take a broader view of this controversy. It is not really a battle between ODF and OOXML. It isn't even really a battle between OOXML and ISO. It is, in the end, a battle between having document standards and not having them. Microsoft is trying to dumb down the concept of standards and interoperability to a point where these concepts are meaningless and ineffective. This is not because they want to support standards more easily in their products. No, it is because they do not want standards at all.*

*Remember, standards bring interoperability, the ability to try out new tools and techniques, the ability to migrate, the ability to chose among alternatives, the ability even to run non-Microsoft products. If standards are meaningless and ineffective, then the incumbent' vendor lock-in will win every time. At that point, isn't it convenient for them to have a monopoly in operating systems and productivity applications? This, in my opinion, is the essence of Novell's 2004 complaint, Opera's present complaint, and the ongoing file format debate. Microsoft's monopoly power and the resulting network effects have lead to a relationship with standards where they win by winning, by drawing, or even by cheating so much that they discredit the system."*

**INDIVIDUAL # 53:** [INDIVIDUAL'S NAME REDACTED]: *Tuesday 1/15/08 1:40 PM*

### Freedom as a Fundamental Goal

I believe that New York State (NYS) should explicitly require the use of free and open standards for all electronic/digital records, and these standards should apply to all communications entered into by the state, including:

- \* document formats ([3]UTF-8 text, [4]OpenDocument Format, [5]W3C-[6]compliant [7]XHTML, etc)
- \* communications protocols ([8]IEEE 802 networking, [9]IP, [10]TCP, [11]HTTP, [12]TLS, etc)
- \* certification mechanisms ([13]OpenPGP, [14]X.509, etc)
- \* operating systems and software ([15]GNU/[16]Linux, [17]OpenOffice.org, [18]Firefox, etc)

When I say free, I should be clear that I'm referring to liberty, not price. The standards and tools promoted/required by the public sector must not be held under the control of a private organization. Anyone should be able to make use of the technology in question, including modification and redistribution of any tools required to access the data.

There should be no proprietary hurdles to jump in order to have access to government data or records. This should be an explicit state mandate.

Below, I detail some of the specifics about why choosing free standards is important.

### Encouraging public access

#### **Question 2. What mechanisms and processes should the State of New York establish for accessing and reading its electronic records in order to encourage public access to those records?**

Requiring free and open mechanisms for accessing public data will encourage public access because there are no proprietary requirements to be met in order to view the data.

For example, no NYS World Wide Web site should ever require the use of Microsoft Internet Explorer. This would cause problems for citizens who have not purchased a Microsoft operating system, since IE only runs on Microsoft's Windows (IE has not been supported on Mac OS for years, and has never been available on GNU/Linux or any other operating system). Instead, NYS websites should adhere to the W3C's XHTML 1.0 standard. Users of all modern computer systems have access to free tools which can render this public standard, so no one is disenfranchised.

A more subtle example: no NYS electronic records should be stored or published in Adobe's Flash format (a.k.a. swf). While Adobe makes players for this format available without charge for most modern operating systems, users are not allowed to modify or redistribute these players, and [19]the license of the format specification itself prohibits use of the specification to make

another player. This lack of liberty (despite the absence of immediate financial cost for those platforms supported by Adobe) means that the public is required to make some sort of arrangement with Adobe (a specific private company) in order to access public data. Since Adobe has no requirement to interact with the public in an evenhanded way, and citizens have no legal recourse to accessing the data themselves, they are effectively discouraged from accessing it.

### **Encouraging interoperability**

#### **Question 3. What mechanisms and processes should the State of New York establish for accessing and reading its electronic records to encourage interoperability and data sharing with citizens, business partners and other jurisdictions?**

Free formats, standards, and tools encourage interoperability and data sharing because there is no restriction on adoption for the other parties in communication.

If NYS was to choose a proprietary format for electronic records, it would need to pay the proprietor of that format a fee for its use. If a neighboring jurisdiction was to choose a proprietary format for electronic records, it would also need to pay the proprietor a fee for its use. If the two jurisdictions happened to choose different proprietary formats, then both jurisdictions will need to pay fees to both vendors if interoperability is desired.

While this would be a windfall for the vendors, the sum of the costs to all jurisdictions scales exponentially as the number of jurisdictions desiring mutual interoperability grows. Better to choose free interchange standards so that there are no additional per-jurisdiction per-vendor costs due to the proprietary nature of the records.

### **Encouraging appropriate government control**

#### **Question 4. What mechanisms and processes should the State of New York implement to encourage appropriate government control of its electronic records?**

Free formats, standards, and tools allow governments (and other entities) to retain control over their own data.

As a fictional example, consider sanitation district boundary records stored in GBG format in 1995. GBG was (fictionally) owned at the time by Dispatch Service Co. (DSCo), which specialized in sanitation logistics. Fast forward to 2008: DSCo has decided that it does not want to support the format any longer (or has been sold to a parent company uninterested in sanitation logistics, or has simply collapsed). The State is now in a weak position to have any access to tools needed for modifying garbage truck routes.

Choosing free formats and free tools from the beginning would allow the government better control over the data because it would be insulated from the fate of any particular instance of DSCo.

### **Encouraging choice and vendor neutrality**

**Question 5. What mechanisms and processes should the State of New York consider for encouraging choice and vendor neutrality when creating, maintaining, exchanging and preserving its electronic records?**

Free formats, standards, and tools avoid vendor-lock-in and promote competitive industry.

As in the above example with sanitation logistics, choosing free formats and free tools would let NYS negotiate among any vendor to meet its sanitation dispatch needs. Even if DSCo was still a thriving, healthy corporation, if the formats and tools used were all free, it would be forced to compete with any other entity willing and able to do the work needed by the city. The underlying freedom translates into a better negotiating position for the State, and encourages healthy competition and interoperability among the potential vendors.

**Electronic record life cycle**

**Question 6. Are there mechanisms and processes the State of New York should establish that are specific to the management of its electronic records in its various life cycle stages (creation, maintenance, exchange, preservation and disposal)?**

Free formats, standards, and tools preserve the integrity of electronic records, and enable verifiable disposal.

When the underlying technologies are unencumbered by any restrictive vendor-applied rules, it becomes much easier to keep data up-to-date, and to be certain that destroyed data is actually destroyed.

If you want to ensure that records can be kept up-to-date, again choosing free formats, standards, and tools helps you to make changes and negotiate between vendors.

Similarly with record disposal. Only unfettered, transparent access to your systems can verify that no trace of the record remains. While i don't expect complete destruction of records to be a high priority for most government tasks, it is extremely important in some cases. For example, court-ordered destruction of data gathered by unconstitutional methods should be verifiable. With a proprietary toolset, it is significantly more difficult to verify that a record has been completely purged simply because no one but the vendor of that tool knows exactly what it does.

**Data preservation**

**Question 7. How should the State address the long term preservation of its electronic records? What should the State consider regarding public access to such archived content?**

Free formats, standards, and tools make it easier to transfer any electronic records to new media, and ensure that the government and the public of the future will be able to interpret the data. Can you read a Lotus 123 spreadsheet (or any other obscure-today, once-popular proprietary format) on your computer today? Can you read an old ASCII text file? The free, well-documented, legally-unencumbered ASCII format is much more accessible now, and will continue to be in the future. Similarly, choosing UTF-8 and the Open Document Format today will give a better shot for the readers of tomorrow.

Preserving data for your own future consumption is very similar to sending data to a relatively-sophisticated party with whom you only have one-way communications. Remember that you don't have any way of knowing if the particular operating system, interface, or libraries are available to your future self. If you want that remote party to be able to understand your data, you are better off using well-documented, freely-available tools and formats. Your future self will have access to these because they are already widely documented and legally unencumbered.

### **Specific statute changes**

**Question 8. What changes, if any, should be made to the government records management provisions in New York Statutes? (Please reference those laws which are cited here: [20][http://www.archives.nysed.gov/a/records/mr\\_laws.shtml](http://www.archives.nysed.gov/a/records/mr_laws.shtml)).**

I wish I had a specific suggestion for modification of the Statutes, but my experience and skill lies in working with electronic data, not in working with laws or legal language. So at the moment, I'll refrain from making specific suggestions for legal language.

### **Implementation costs**

**Question 9. What constraints and benefits should the State of New York consider regarding the costs of implementing a comprehensive plan for managing its electronic records?**

There will be significant costs to State-wide adoption of any electronic record plan, whether proprietary or free. Some proprietary vendors will most likely try to offer short-term cost discounts to encourage the State to choose a format or a system that they control. Accepting such an offer would be a very bad deal for the State and for the citizenry in the long run, due to the problems with proprietary tools, formats, and standards outlined here.

Transitioning to free infrastructure should have similar real costs to transitioning to any particular proprietary infrastructure (less of course any surcharge that the proprietors levy on the latter). But choosing freedom will have many more payoffs in the future (see question 11).

### **Highly-specialized data formats**

**Question 10. What should the State of New York consider regarding the management of highly specialized data formats such as CAD, digital imaging, Geographic Information Systems and multimedia?**

Choosing free data formats wherever possible will ensure the widest range of tools to be able to work with any given dataset, even in highly-specialized domains.

Digital imaging and multimedia have significant, functional free options in the wild. For example, images can be stored and transmitted at arbitrarily high resolution in the [21]Portable Network Graphics format. Video and audio can be stored and transmitted using the unencumbered formats provided by the [22]Xiph foundation. GIS data can use a variety of free formats, including the widely-available [23]TIGER format. XML-based free formats like [24]SVG can also be used for arbitrarily-detailed vector graphics.

Where there is no free format, standard, or tool available (this would be a highly unusual case), the State should collaborate with jurisdictions with similar needs to craft a free format to be

published for public review. In this way, NYS can retain the benefits of free infrastructure for its own data, and ensure that its own needs are met in the future without being locked into a single vendor.

### Ongoing costs and savings

#### **Question 11. What constraints and benefits should the State of New York consider regarding potential savings or additional costs associated with the management of defined electronic record formats?**

Choosing free formats, standards, and tools will provide significant cost savings as time goes on and State-wide policy stabilizes.

Because of the lack of vendor lock-in, the State will be able to choose from a healthy range of groups offering services to maintain and manipulate the data. No overhead or maintenance will need to be paid to any particular vendor. And as new needs arise, the State can employ teams of their choice to craft new tools or modify old ones without needing the permission of any particular proprietor.

Additionally, by making a public commitment to free tools the State avoids the risk of becoming embroiled in litigation concerning breach of licensing. Even threats from litigious proprietary vendors for breach of license can prove costly, as [25]the city of Philadelphia found in 2001.

### Existing policies as precedent

#### **Question 12. What existing policies and procedures in the private or public sector for the management of electronic records would be appropriate for the State of New York to examine? Please cite specific examples.**

Massachusetts recently adopted the OpenDocument format as its standard document format. While their decision to retain Microsoft Office and use an ODF plugin was unfortunate, their selection of a free format allows them the ability to move to free software in the future. Unfortunately, their more-recent decision to allow the use of Microsoft's OOXML format (a [26]non-free format because of its numerous hidden "gotchas" and its control by a single vendor, despite Microsoft's claim to the contrary) is deplorable, and returns them to the very real possibility of one-vendor lock-in.

Minnesota's house bill [27]H3971 from Legislative session 84 (2005-2006) was a concrete proposal for a State mandate for free document formats, but I don't think it was passed. It would be good to see New York take a similar position to the one proposed by Rep. Thissen.

[28]Brazil ([29]english translation by google) and [30]Croatia have also in recent years made explicit government mandates for free standards and free software. While the implementations have not been as vigorous as they could be, but following their intent in NYS would be admirable.

The EU has also issued [31]the European Interoperability Framework for pan-European eGovernment Services, which states principles to be adopted by governmental services, including:

- \* Use of Open Standards
- \* Assess the benefits of Open Source Software

\* Use of Multilateral Solutions

While this framework offers little in the way of concrete guidance, their general principles seem sound.

The Extremadura region in Spain [32] adopted government-wide free software policies in 2006.

[33] Mannheim, Germany [34] began transitioning their computing infrastructure to Linux in 2006.

The City Council of [35] Munich mandated a switch to Free and Open Source software in 2003 [36] This transition project is now under way, known as LiMux.

### Adequacy of existing policy

#### **Question 13. Are New York State's existing standards, regulations and guidelines regarding records management adequate to meet the challenges of electronic records retention? How should these standards, regulations and guidelines be changed?**

As with question number 8, I'm going to decline to offer specific commentary on general legal language since it is outside my particular expertise. I expect to see a strong mandate for freedom in any newly-created modified standards, regulations, and guidelines. I hope that New York State legal and legislative representatives will keep these goals in mind while drafting new requirements.

Thanks for considering these important goals. I look forward to seeing New York State become a leader in this area.

#### Links:

1. <http://www.oft.state.ny.us/oftnews/erecords-study.htm>
2. <http://cmrg.fifthhorseman.net/wiki/NYSRFPC122807>
3. [http://unicode.org/faq/utf\\_bom.html](http://unicode.org/faq/utf_bom.html)
4. [http://www.oasis-open.org/committees/tc\\_home.php?wg\\_abbrev=office](http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=office)
5. <http://w3.org/>
6. <http://validator.w3.org/>
7. <http://www.w3.org/TR/xhtml1/>
8. <http://www.ieee802.org/>
9. <http://rfc.net/rfc791.html>
10. <http://rfc.net/rfc793.html>
11. <http://rfc.net/rfc2616.html>
12. <http://rfc.net/rfc4346.html>
13. <http://rfc.net/rfc4880.html>
14. <http://rfc.net/rfc3280.html>
15. <http://gnu.org/>
16. <http://kernel.org/>
17. <http://openoffice.org/>
18. <http://mozilla.com/firefox>
19. <http://www.adobe.com/licensing/developer/fileformat/license/>
20. [http://www.archives.nysed.gov/a/records/mr\\_laws.shtml](http://www.archives.nysed.gov/a/records/mr_laws.shtml)
21. <http://www.w3.org/Graphics/PNG/>
22. <http://xiph.org/>

23. <http://www.census.gov/geo/www/tiger/>
24. <http://www.w3.org/Graphics/SVG/>
25. [http://archive.salon.com/tech/feature/2001/07/10/microsoft\\_school/print.html](http://archive.salon.com/tech/feature/2001/07/10/microsoft_school/print.html)
26. <http://www.nooboxml.org/petition>
27. <http://www.revisor.leg.state.mn.us/bin/bldbill.php?bill=H3971.0.html&session=ls84>
28. <http://www.iti.gov.br/twiki/bin/view/Swlivre/WebHome>
29. <http://www.google.com/translate?u=http%3A%2F%2Fwww.iti.gov.br%2Ftwiki%2Fbin%2Fview%2FSwlivre%2FWebHome&langpair=pt%7Cen&hl=en&ie=UTF8>
30. <http://www.linux.com/feature/56376>
31. <http://europa.eu.int/idabc/en/document/3761>
32. [http://www.regiondigital.com/modulos/mod\\_periodico/pub/mostrar\\_noticia.php?id=43131](http://www.regiondigital.com/modulos/mod_periodico/pub/mostrar_noticia.php?id=43131)
33. <http://www.mannheim.de/>
34. <http://www.heise.de/english/newsticker/news/64856/>
35. <http://www.muenchen.de/>
36. <http://www.muenchen.de/Rathaus/dir/limux/english/147197/>

**INDIVIDUAL # 54:** [INDIVIDUAL'S NAME REDACTED]: *Tuesday 1/15/08 5:15 PM*

As an undergraduate studying history at Cornell University, with acceptance into the history Ph.D. program at SUNY Buffalo next fall, I would like to submit comments on the electronic record policy for New York State.

I would like above all to stress the importance of the preservation of free, open access to public records. As an aspiring professional historian, I understand the centrality of archival research to the protection of the public's ability to explore its past. Without ready availability of public records to groups like historians, the difficulty of preserving and extending a society's historical, cultural and political heritage becomes manifestly more difficult.

Historians already face numerous challenges accessing historical sources like public records: archives may restrict access or documents may be lost or corrupted, for instance. Electronic data, which will doubtless become increasingly central to the historians of the future as it replaces traditional storage methods, adds a further challenge, as researchers must be able to properly decode electronic records in order to read them. Electronic data created in a format that no computer can access is as useless to historians as a document written in a language that no living human can read.

I believe that the best way to ensure that records remain as readily accessible as possible is to adopt truly vendor-neutral, community-developed formats for the storage of electronic records. The OpenDoc Format (ODF), in my view, is the most effective means of electronic storage that will ensure that public records are always as accessible as possible to everyone.

I understand that ODF's chief competitor is the Office Open XML (OOXML) format, developed largely by Microsoft through a non-transparent process. While I am unqualified to address the technical differences between the file formats, I am concerned that vendor-neutrality and open access would be compromised if the OOXML format were adopted as New York State's standard, simply because of the lack of vendor neutrality inherent in the implementation of the OOXML standard.

Even given public availability of the OOXML specifications, there is a significant risk that end-users will remain in large part dependent upon Microsoft to access and create records in the OOXML format. Should the published OOXML documentation prove incomplete or unclear, it may be impossible for problems to be addressed without Microsoft's assistance.

Moreover, in light of the fact that Microsoft played such an integral role in the development and marketing of the OOXML format, it is difficult to rest comfortably with the assumption that the OOXML format is not tailored towards Microsoft's business objectives at the expense of those of its competitors, or that Microsoft will not use its expertise in the OOXML format to prejudice developers who, because they were not involved in the creation of OOXML, will find it more difficult than Microsoft employees to implement that format in their applications.

ODF, conversely, was not dominated in its development or distribution by a single organization; instead, it is the product of a highly transparent effort involving representatives from a diverse group of companies and organizations. This reality helps to safeguard against vendor prejudice and accessibility difficulties. In addition, while OOXML is a relatively new file format with minimal use outside of the most recent Microsoft products, ODF has been publicly available for a considerably longer length of time, and has been used by numerous software developers in various products, further attesting to the breadth of its neutrality and free availability.

The historians of the future should not be forced to face more challenges than their profession already entails in accessing documents.

The adoption of an electronic storage format like OOXML poses a potential for dominance of public record storage by one vendor, adding another obstacle to the work of historians and other individuals seeking to decrypt data in a format lacking a veritable guarantee of free accessibility. Since a truly vendor-neutral, community-developed alternative exists and has already been implemented in numerous situations, New York should follow the lead of other states, like Massachusetts, and adopt ODF for public record storage.

**INDIVIDUAL # 55:** [INDIVIDUAL'S NAME REDACTED]: *Friday 1/18/08 12:54 PM*

I am gratified for the opportunity to make this response to your Request for Public Comment with respect to the study concerning electronic records policy for the State of New York. [CONTACT INFORMATION REDACTED] I hope you will receive my response as that of a private individual (educated in New York State and having resided there for the 24 years prior to this one). As an Associate Member of the Free Software Foundation, a strong early contributor to the ODF community, a professional software consultant, a technology journalist and author on desktop software as well as a participant in software standards development work on document formats, I am deeply concerned and committed to general improvement in productivity, fair markets, as well as improved access to public information. (If useful, further detail on my background may be found here: [CONTACT INFORMATION REDACTED])

My answers will be confined to my area of expertise: web pages and desktop documents. This means electronic documents authored on the common desktop personal computer or mobile handset which I define to include e-mail, word-processed works, spreadsheet works & presentation works.

In the context of the discussion, such works are contained in what are called "files" (which in all cases consist of some specific "format", e.g., in usage, "file," "format" as well as "file format");

and my comments recognize a distinction between document files, their document formats and the “software applications” that are used to author them.

For the purposes of this discussion also, 'interoperability' refers to applications working together harmoniously and files are “compatible” with respective applications. Files are not “compatible” with other files and applications are not compatible with other applications unless in the latter case they are made to be so (and they are not yet); although a theoretical single universal document format would enable files to be compatible with all applications implementing that format.

My responses will focus strictly on a subset of what New York State conceptually defines as “documents”, “records” and “official [government] records” but only those document, records and official [government] records which are authored in the above-noted desktop PC, notebook PC or mobile contexts.

Wherever I discuss electronic data or discrete information elements (such as a birth dates, social security numbers or driver's license numbers, for example, which pass through and may be shared across various electronic interfaces and storage systems), I will refer specifically to 'data' or 'data elements'.

Wherever I discuss document record or data archiving, I will use appropriately specific terms.

## **Response to Part I**

### **Question 2. What mechanisms and processes should the State of New York establish for accessing and reading its electronic records in order to encourage public access to those records?**

Standardize on the following...

- a) HTML or XHTML for web pages;
- b) TXT (or .txt) for plain text editable but unformatted documents;
- c) PDF/A for fixed uneditable documents;
- d) Web pages must read, work well and conform to specifications for at least one free & open source web browser, such as Mozilla Firefox or Konqueror;
- e) wherever editable documents are published, the file format used must not be a proprietary document format (or a format that is enuring to other related proprietary software) such as .doc, .xls, .ppt, .docx, .xlsx, .pptx or other formats which are either proprietary in nature or nominally open but abetting to market abuse.
- f) in the case of editable documents, ODF has become a popular policy option; however, events in the 2005-2006 Massachusetts ITD “Pilot of ODF-Ready Applications” demonstrate the human difficulties of implementing ODF policy through migration to OpenOffice.org on Windows XP across de-centralized state government agencies, let alone full desktop migration to GNU/Linux.

(Disclosure: the author was the professional consultant to Mass ITD for Pilot implementation and remains under mutual non-disclosure agreement with the Division.)

ODF development has proceeded in 2007 on a path contrary to principals of file format interoperability with legacy document formats (such as those originating from Microsoft applications). ODF's sponsors are pursuing interoperability at the application level, which is not conducive to public policies seeking to identify a durably useful document format that can be relied upon to match the openness and life-cycle requirements of public sector agencies and the general public.

For editable documents on the desktop, some Universal Document Format is necessary; however, no software application exists today to meet the requirements of such a format. It goes without saying the Microsoft's new formats, "OOXML," (regardless of events at ISO) are unfit to meet open document policy requirements too. Government agencies cannot however avoid the need to handle both ODF and OOXML, but neither format is suitable for public policy aimed at standardizing daily desktop work or archiving.

g) In all circumstances, it will be necessary to force document editing software application vendors to add HTML 5.0, XHTML 2.0 and CSS 3.0 import and export capabilities to their applications natively. This would make a suitable first step in promoting application interoperability and file-to-application compatibility for the purposes of business processes moving across the common desktop and also for archiving.

**Question 3. What mechanisms and processes should the State of New York establish for accessing and reading its electronic records to encourage interoperability and data sharing with citizens, business partners and other jurisdictions?**

- a) Distribute information via the Web in HTML, XHTML, TXT, PDF/A and some Universal Document Format (neither ODF nor OOXML are sufficient or adequate);
- b) standard, interactive Web forms & processes, following best practices based around LAMP implementations;
- c) simplicity & consistency are encouraged;
- d) Libraries Example: see Georgia Pines based on the Evergreen open source interactive catalog system, permitting full internal catalog management while also providing public exposure of the catalog along with MyAccount interactive customer management facilities. (<http://www.georgialibraries.org/public/pines.php> )

**Question 4. What mechanisms and processes should the State of New York implement to encourage appropriate government control of its electronic records?**

- a) difficult to comment on productively in general;
- b) good, clear & enforceable e-document policy will enable and abet downstream control objectives.

c) PDF/A is a good document distribution format for controlling document content & layout because for most practical purposes it ensures that documents do not re-circulate with unauthorized or confusing changes.

**Question 5. What mechanisms and processes should the State of New York consider for encouraging choice and vendor neutrality when creating, maintaining, exchanging and preserving its electronic records?**

a) policy must identify specific non-proprietary FORMAT standards for each relevant specific use or business process context (see Massachusetts ITD ETRM 4.0:

b) Use W3C standards (i.e., “Web standards”) wherever possible; W3C standards are highest quality and possess more integrity in development process and governance which yields greater vendor-neutrality than standards from other standards bodies; W3C standards are at least equivalent and in many aspects superior to ISO standards, though this may not be a permanent condition.

c) policy must avoid identifying specific SOFTWARE APPLICATIONS for procurement because doing so introduces vendor-specificity to policy;

d) software procurement policy must not rule out any kind of software and should be designed to affect a fair procurement process environment; favoring Free or open source software application acquisition over proprietary application acquisition doesn't work; Free or open source Software is easier to acquire and implement and therefore benefits from a healthy systematic advantage if the procurement process environment is made to be and stay fair;

d[sic]) in light of monopoly software's initiatives to pass proprietary mechanisms with partially open aspects through standards bodies like ISO, government IT policy must consider not only the nominal openness of a format standard itself but whether the standard's use may effect some indirect benefit to proprietary software. Microsoft's OOXML, XPS, Silverlight, XAML and other initiatives threaten to have such effects.

e) with respect to FORMATS, policy should define to specificity what is an “open format standard” by establishing clear requirements and maintaining a list of accepted versus rejected standards;

**Question 6. Are there mechanisms and processes the State of New York should establish that are specific to the management of its electronic records in its various life cycle stages (creation, maintenance, exchange, preservation and disposal)?**

a) Life cycle stages are useful in conceptualizing and generalizing about business processes, their characteristics and human requirements. However, it is unfortunate that no

specific formats, policies or behaviors apply to any specific life cycle stage across all business processes.

b) Massachusetts ITD's ETRM 4.0 uses a concept of architectural "domains" as a motif to generalize about the different modules which occur within an idealized Services-Oriented Architecture. Their domain designations -- "Access & Delivery Domain," "Information Domain," "Application Domain," "Integration Domain," "Management Domain," and "Security Domain" -- reduce an enterprise IT architecture into modular components and in doing so identify the available open standards that can be used to knit the whole together. This is a software- or standards-based concept which contrasts to the behavioral emphasis of life cycles. Both are important and both are useful for formulating policies that with affect business process change.

**Question 7. How should the State address the long term preservation of its electronic records? What should the State consider regarding public access to such archived content?**

a) Each business process and record type will suggest its natural or ideal long-term archival format as well as define the length of the necessary term of preservation as well as the ease with which people can find it.

b) In the case of a document, for example, different business processes or end-user contexts will suggest requirements as to CONTENT and LAYOUT fidelity to the original. Sometimes the original appearance of a document is necessary (PDF/A) and other times only the content matters (TXT). Sometimes limitations of storage capacity will guide the appropriate choice of format as well as its implementation.

c) A format openness (in terms of its development, governance, software implementation & access) is a critical consideration. Open formats are essential to long-term preservation of records because they are i) developed by public consensus (everyone gets a chance to define how they work and no one gets special consideration); ii) cheaply and easily accessed by a variety of free software tools over a possibly undefined number of decades in the future; iii) fully specified in an obvious public place (therefore easily re-implemented in software or converted accurately in both content and layout); iv) absent of patent encumbrances in the specification of the format (which would make them impossible to be implemented freely & easily at any time in software applications)

**Question 8. What changes, if any, should be made to the government records management provisions in New York Statutes? (Please reference those laws which are cited here: [http://www.archives.nysed.gov/a/records/mr\\_laws.shtml](http://www.archives.nysed.gov/a/records/mr_laws.shtml) ).**

(see other responses)

**Question 9. What constraints and benefits should the State of New York consider regarding the costs of implementing a comprehensive plan for managing its electronic records?**

a) The comprehensive plan approach – what I call “the Big Bang!” -- has appeal due to the perception of transition cost savings and transition time compression as well as the perception of synergies from cutting out wasteful processes on a large & concerted scale. These benefits never, ever, pan out because business processes on the ground (like wars) are never in reality quite what they seem from 30,000 feet up. It is too hard to plan

unified and concerted transitions in environments of complexity and where many business processes were created ad hoc or improvised by workers, who have a poor conscious idea of their own processes and no ability to describe them effectively.

b) Major risk of being too early! Policy is determined by the technology available. While ODF is not it (due to its failures on format-level interoperability with Microsoft legacy documents), there is no question that a Universal Document Format – one surrounded by a rich ecosystem of application suppliers including all Office 2.0 services -- is possible, implementable and would meet widespread adoption if local, regional, state & national government agencies around the world achieve a consensus on the technology and protect the concept. Establishing wide-scope policy too early would have two problems: i) technological change can force a change in a component of the policy; and b) policy change can subsequently cascade throughout the wider plan. Frustration and cost are directly proportional here.

c) Major Large-Project risk! Small and manageable phases, accompanied by frequent celebrations of success is the route. Base hits versus home runs. A comprehensive framework will be necessary, but since technology and opportunity evolves it is important to maintain flexibility and eradicate execution bottlenecks.

e [sic] Implementing a Big Bang! with outside help (for example, from global professional services firms) can cost a mint and can impose some of the same vendor-dependencies that the Plan was intended to remove in the first place. It is also less efficient because the outsiders don't know the business or culture as well as natives. The alternative approach – building an in-house taskforce of natives based on numerous small teams to execute a 10-year plan -- is being pursued with success by the City of Munich (Germany) at significant cost savings.

**Question 10. What should the State of New York consider regarding the management of highly specialized data formats such as CAD, digital imaging, Geographic Information Systems and multimedia?**

a) At the high end of the market, where there is specialization, the market is not subject to the same commoditizing forces as at the low-end, where generic desktop is changing so dynamically that we have begun to rethink our common business processes. Expect no fundamental changes in in these categories, except such changes and new techniques which New York State learns from Planning and Implementation of its Comprehensive Plan.

b) Multimedia is less specialized and formats can be expected to be commoditized rapidly. Accordingly, there will be room for the State government agencies to manage their media most effectively through enterprise relationships with the reliable Web 2.0 streaming vendors (and not do this internally).

**Question 11. What constraints and benefits should the State of New York consider regarding potential savings or additional costs associated with the management of defined electronic record formats?**

a) Surprising total cost of ownership savings on basic software and systems, better-working and more user-friendly business processes and long-term productivity gains (for State employees and citizens) as well as procurement flexibility that will permit New York

State agencies to keep up with current technology improvements within the year they are first implemented publicly (or instantly if Web 2.0/3.0 applications are being used).

b) Risk of early-mover disadvantage. If New York State were to effect a multi-million dollar migration to ODF/OpenOffice.org, for example, on existing Windows XP machines across agencies in 2008 – 2010, and later find out that it would have been preferable as well as cheaper to run Office 2.0/3.0 applications like Adobe Buzzword or Google Apps from the newly specified Linux, Mac & Windows desktops, then tax-payers will have a valid complaint.

c) Transition costs to migrate business processes throughout and across agencies may be high or unpredictable depending on the level of talent involved and the structure of the project.

d) The de-centralized structure of State government agencies makes comprehensive change affecting basic every-day systems difficult to implement efficiently and it makes business process transitions unpredictable to manage. In mitigation however, tools & techniques exist in the operations- and incentive-management fields which can be employed in desktop migration settings to moderate the human risk-factors.

**Question 12. What existing policies and procedures in the private or public sector for the management of electronic records would be appropriate for the State of New York to examine? Please cite specific examples.**

a) Massachusetts ITD's ETRM 4.0 is an elegant policy created and iterated through the unusually close collaboration of State IT policy professionals with the very high-quality IT community in that region.  
([http://www.mass.gov/?pageID=itdsubtopic&L=4&L0=Home&L1=Policies%2c+Standards+%26+Guidance&L2=Enterprise+Architecture&L3=Enterprise+Technical+Reference+Model+-+Service-Oriented+Architecture+\(ETRM+v4.0\)&sid=Aitd](http://www.mass.gov/?pageID=itdsubtopic&L=4&L0=Home&L1=Policies%2c+Standards+%26+Guidance&L2=Enterprise+Architecture&L3=Enterprise+Technical+Reference+Model+-+Service-Oriented+Architecture+(ETRM+v4.0)&sid=Aitd) )

**Question 13. Are New York State's existing standards, regulations and guidelines regarding records management adequate to meet the challenges of electronic records retention? How should these standards, regulations and guidelines be changed?**

no comment; see other responses

**Question 14. What else should the State of New York consider about this subject?**

a) The question of implementability of certain format standards is of primary importance. It is okay to declare a policy in support of a certain standard, but if that standard is affixed to specific software which does not go well with legacy systems, then the policy endeavor cannot advance far.

b) It is important for New York State to watch what is happening to ODF policies in other "Blue" US states as well as in the countries where ODF policies have been advanced early (Belgium, Holland, Denmark, Norway, South Korea, for example). If there is a lack of US states other than Massachusetts declaring ODF policies, and there is, New York must consider carefully and seek advice from other state IT leadership as to why. (Where is California's position on ODF, for instance?). If New York notices the early-adopting ODF

countries each sliding back to ODF + OOSML policies, there is a reason for this and it is important to understand why this is happening.

## **Response to Part II**

### **Question 1. Re the definitions of “electronic data,” “documents,” and “records,” in Part I...**

a) The definitions are over-broad for purposes of both procurement and standards aspects of IT policy, which should necessarily specify clearly to which system architectural context the policies happen to apply. The Massachusetts ITD ETRM 4.0 offers a useful reference here.

### **Question 2. Re the definitional components of “access” -- “day-to-day utility,” “ancillary active records access,” and “historical access”.**

a) These distinctions may too be over-broad for the purpose of policy relevance to either existing or projected business processes.

b) In considering a relevant framework for policy, visibility of included business process 'inventories' – both existing and projected – would seem to be essential. Policy written by an overly abstracted breakdown of process activities risks irrelevancy.

### **Question 3. Re security or privacy implications of ODF v OOXML?**

a) Desktop systems are inherently insecure and there is no practical or meaningful difference between the two formats within this large context.

b) OOXML does preserve legacy binary data in files – i.e., not all data elements originating from older MS Office files gets converted to XML when opened and saved, for example, in Office 2007. This legacy binary data is not human-readable and may represent differentiated, added, risk being carried by adopters of OOXML. Additionally, OOXML's specification references certain non-open formats and certain mechanical details of function which are protected by patents and are not visible or understood by the public. In these respects, OOXML is theoretically less appealing to security experts because ODF is the more fully specified (despite the length-differentials of the respective specification documents – over 6,000 for OOXML and less than 1,000 for ODF). The argument will and perhaps should be made; but it must be received in healthy perspective as hair-splitting in light of the overwhelming security issues based on technological and human factors attending documents, business processes and desktop software systems.

### **Question 4. Re will the FOIL (Freedom of Information Law) be affected?**

a) When a reasonably open document format which evinces most-preferable characteristics of openness (vis development, governance, access and market-balancing influence), a state government should not publish any information that is designed to potentially be accessed by the public in a format other than a W3C-standard format. In such business processes where a W3C-standard format may not be available, it is of critical importance that policy favor the relatively more open format and rejects any

format that enures to the benefit – even indirectly through parallel and orthogonal designed-in system dependencies -- of any monopoly-abusing entity.

b) OOXML is a detriment to FOIL and most regulations pertaining to records and documents because OOXML creates and reinforces systematic dependencies on many different products, protocols, methods and formats throughout the Microsoft catalog of products -- in addition to Office 2007 and Windows Vista operating system. These include for enterprises Sharepoint, Exchange/Outlook, Windows Server 2003, MS SQL Server and others; and for developers the .Net Framework, Silverlight, XPS and others.

The existence of a freely downloadable application for native ODF – i.e., OpenOffice.org – makes ODF preferable to OOXML in the FOIL context.

#### **Question 5. Re “government control”?**

a) One context of control relates importantly to the desktop document authoring segment. PDF/A is an Adobe format that is not editable. While many people feel it is therefore not “open,” that is not true. Adobe has contributed sufficient specification information to the public and offers a useful and lightweight reader application which makes viewing PDF files universal for all practical purposes. PDF is a useful format BECAUSE it is non-changeable. This makes it a very good choice for governments to publish documents in a fixed format without the threat that the content of the document can be changed (easily with common desktop systems).

For documents published in editable formats (for example, .rtf, .txt, .doc or .odt), agencies should always provide good reasons for publishing this way (and not do this by default) because of the risk of loss of control or compromise of good information.

#### **General response to D. Definitions - “Interoperability”**

a) It is nice that governments require ODF usage, which forces the installation of the free OpenOffice.org office suite APPLICATION onto public administration desktops. Doing so saves public funds – principally in recurring software licensing fees to proprietary software companies. The problem is that it is very difficult to migrate offices from MS Office to OpenOffice.org. Why? Because the two applications are not interoperable enough when used in the same work group environment.

Among the serious difficulties users experience: a) document files sometimes look different when opened in the different application (the font-substitution problem, or application behavior differences yielding file format artifacts that are unreadable by one of the applications, such as the case with nested tables or page-break dynamics); b) document software feature differences between the two different applications cause users to slow down or stop work (for example, in spreadsheets, when typing long text label descriptions into cells the carriage-return commands are both different and non-intuitive; or, with documents, when users experience frustration finding formatting styles or adding footnotes); c) file corruption when documents are “roundtripped” (opened and saved in the application repeatedly such as during collaborative authoring via e-mail attachment); and d) increased risk of file-loss or version-confusion (user errors) from poor understanding of the different path-saving interface procedures designed into the different applications (such as when colleagues e-mail a spreadsheet file that contains formulas pointing to cells

in other spreadsheet files -- "relative references" -- the referenced paths change when the file is saved to a new location and the links break).

Given the imbalance of large difficulty from superficially quite minor application design differentials, an enterprise must go through an inordinate amount of pain -- or use very clever migration techniques (which exist but need to mature through practice) -- to shift its work groups from MS Office on Windows XP to OpenOffice.org on Windows XP, for example, without experiencing significant productivity constraints during and well after a migration. The problem of migration management is compounded by the complexity of geographically disbursed state government agencies with de-centralized authority in the IT organization.

b) The definition of 'interoperability' for the purpose of desktop software policy must be explicit and establish standards limitations of acceptable loss of CONTENT and LAYOUT in specifically cited contexts of document exchange.. In the case of Microsoft's competitive advantage over other Office suite providers, it is important to understand where their advantages come from. Microsoft creates advantage through owning and hiding the 'recipe' for its document formats, which creates and reinforces the need among the public to purchase and repeatedly upgrade their document APPLICATION software. This much is clearly understood in the industry.

However, creating an alternative document-creating software application and an alternative document format with some relatively open characteristics is not sufficient to shape a more competitive market-place nor to solve state government's documents & records requirements. The reason is because the demands of businesses are complex and they require the NEW and the OLD document software APPLICATIONS to function as if they are the same, which they are certainly not (MS Office and OpenOffice.org are deliberately different and their controlling vendors insist on the right to do things differently -- or "innovate"); moreover, business processes demand NEW document files to work equally as well in the OLD software application as in the NEW SOFTWARE application, and the same for OLD document files. They do not.

We say a certain type of document file (.doc, for example) is perfectly compatible with a certain Office suite application (MS Word, for example); the same is equally true of .odt with OpenOffice.org Writer. Even though OpenOffice.org Writer can open the odd .doc file and save it as .odt with success, these applications do not behave together IN COMPLEX BUSINESS PROCESSES at the level of the professional enterprise work group in a way that can be said to be sufficiently interoperable. In effect, MS Office (2000 -- 2003) and OpenOffice.org (2.3) are not interoperable according to what is a pretty good and workable general definition given by CIO/OFT -- "products and systems from multiple vendors that can be used together without modification or development of custom interfaces and tools."

Curiously enough, that definition describes e-mail: a common, agreed text-based format around which many different vendors provide a rich plethora of value-added interfaces and services.

There is no good reason why documents can't have the same thing.

c) The way to solve this is to respect that definition of interoperability and follow through with a plan which achieves for documents and document authoring tools what e-mail standards have achieved for the e-mail medium. State governments must join together around the world and, working with the W3C, create an agreed general document format – a Universal Document Format – which facilitates perfect document file compatibility and document authoring application interoperability while making policy toward a common single open standard document format feasible.

It can be done.

### **General response to E. Definitions -- “Open standards”**

a) It should be noted that this phrase is difficult to define today; and due to Microsoft's assault in 2007 upon ISO, it will get more – not less – difficult to define in future.

b) Open standards are W3C standards. They are the only open ones, for most relevant and practical purposes. Other standards bodies fail to meet the W3C integrity of vendor-neutrality, technological talent, soundness of governance and firmness of leadership direction.

c) What is the test of an open standard? The South African Department of Technology's definition is okay, but a flexible interpretation would permit ODF to meet the criteria. It is certain that ODF's development direction in 2007 compromises its openness on grounds of a) failure of OASIS to encourage an open public development environment; and b) failure of the ODF specification to address the compelling need to preserve MS Office file format internal data in a way which makes the OpenOffice.org and MS Office APPLICATIONS interoperable around a single, good document format.

### **Question 6. Re correct definition of interoperability. Is there a better alternative?**

a) In light of the broken market environment, the high level of complexity and ease of made-to-order disagreement about the definition of “interoperability,” New York State must establish a new implementable definition of interoperability for documents and document authoring tools. Interoperability for desktop documents and their authoring tools should be equivalent in effect to that which we enjoy today with e-mail (IMAP, POP, SMTP, HTTP & TXT) and web pages (HTML, XHTML, CSS)

Declaring a definition won't work without running code. However, resources do exist to implement a proof of concept that would provide a sufficient foundation on which the State of New York could partner with California and other like-minded government entities around the world to advance the software implementation and build a successful governance model around what we visualize as an implementation of W3C Compound Document Format (CDF).

### **Question 7. Re correct definition of openness?**

a) see General response to E. Definitions -- “Open standards”, above.

b) The criteria described reflect standard ideas developed by the ODF community (with contributions from me) after ODF was ratified by ISO in May 2005. Since these criteria

would pass ODF, and ODF is no longer a desirable standard for fulfilling New York State's requirements, it is fair to conclude they are too broad.

c) New York State would come closer to the objective for documents and records if overall requirements were amended to "...an Open Standard that also must meet our revised definition of Interoperability..."

**Question 8. Re state agency respondents**

no comment

**Question 9. Re is Gartner's prediction correct?**

a) Gartner's prediction quite likely to be irrelevant. It is relatively easy for individual citizens to adopt OpenOffice.org due to the lower complexity of their desktop workflows. But it is much more difficult for enterprises to do so, as noted above.

A mandate is not a choice.

Governments may require a certain format, but if the government itself cannot easily implement that format by installing the software application to which that format is native, then you have a situation where the market reverts to de facto standards immediately.

**Question 10. Re consequences of not adopting ODF?**

a) If ODF is the chosen open alternative document format to which everyone worldwide agrees, then New York State must pursue a costly rip & replace of software just like everyone else. In this context Gartner's numbers are correct enough in scale. It will be impossible to co-exist with ODF-ready supply chains & citizenry without migrating agency tools to the same.

Likewise, it will be equally difficult to co-exist without some OOXML capability, unless OOXML / Office 2007 faces a total rejection at ISO, a determined ban of some kind or an astonishing reversal in the market-place.

b) If New York State helps to establish a better, more open, more interoperable alternative document format solution – other than ODF (for example, we have proposed a solution which would enable state agencies to deploy ODF or a better alternative document format on existing Windows & MS Office installations without a costly software migration) – then a costly rip & replace can be avoided.

**Question 11. Re which governments have adopted ODF, OOXML exclusively?**

a) Massachusetts' ETRM 4.0 -- ODF + OOXML

b) Denmark – OOXML (still studying ODF)

c) Belgium – ODF (still studying; I expect their policy to shift to ODF + OOXML)

- d) Holland – ODF (may need to follow whatever Belgium does due to interconnected processes based on the multi-lingual Flemish/Walloon/French/German language requirements of government services)
- e) The situation is still fluid – Norway, S. Korea, Malaysia and S. Africa are establishing ODF policies too -- and generally looks like most policies will be forced by reality to adopt the dual format – ODF + OOXML – strategy.
- f) Other formats identified in other government policies include HTML, TXT and PDF/A.

**Question 12. Re any other relevant contexts than the office suite?**

- a) If state agencies are aggressive at pushing application services into the web browser, there is no other important area – except multimedia formats – where interoperability is as pressing a concern.

**Question 13. Re means of the State to further openness of all relevant formats?**

- a) Search for the work of The OpenDocument Foundation on W3C Compound Document Framework (“CDF”). CDF is an excellent potential unifier for different document format standards – particularly those existing at W3C.
- b) The State could work together with other state CIOs and government IT experts and the W3C. If specific processes or proprietary formats are identified, then working groups can be formed to create standards in a more ideal form.

**Question 14. Re proposed focus for the study?**

- a) yes ... that works

**Question 15. Re what is the “problem” that this study should be addressing?**

- a) Is ODF the right XML-based format for New York State or any enterprise?
- b) If ODF is inadequate, why? Is there reason to hope we can have a better open alternative?
- c) Can ODF | OpenOffice.org reasonably be implemented (installed) on all the office desktops of our agencies?
- d) How, when, at what cost, with what personnel will such a project be undertaken?
- e) Is a New York State Pilot of ODF-Ready Applications necessary, given that Massachusetts has already done one?
- f) What do we know from the Massachusetts experience? What can we know?
- g) What are our office suite format requirements?
- h) What kind of format would ideally fulfill our office suite requirements?

- i) Are document formats a public utility?
- j) Should/can state governments worldwide get together to create a satisfactory office suite format / application framework that can bring the openness of e-mail to office suite documents? Will that stifle innovation? Will that enable innovation?
- k) Do state government IT people want/need to be in the standards or applications business?
- l) Can we get the state government IT agencies in the US and around the world to come out from under their rocks on this (we are immensely powerful together)?

**Question 16. Re mix of formats?**

[no comment]

**Question 17. Re customer requirements met by OOXML?**

- a) all the basic word-processing, spreadsheet & presentation page-authoring tools (paragraphs, bolds, fonts, frames, graphics headers & footers, etc.)
- b) custom schemas for industry vertical processes (Office as platform)
- c) custom scripting functionality
- d) fluid interchange with legacy Microsoft document formats and zero migration cost.

**Question 18. Re customer requirements best met by ODF?**

- a) all the basic word-processing, spreadsheet & presentation page-authoring tools (paragraphs, bolds, fonts, frames, graphics headers & footers, etc.)
- b) ODF-to-PDF file export
- c) freely downloadable native application -- OpenOffice.org

**Question 19. Re office suite features?**

- a) For day-to-day, ODF's features are well-adequate.
- b) The important question is not about features, but can ODF be implement across a single State's disbursed agencies within a single year at reasonable dollar, productivity and opportunity costs?

**Question 20. Re long-term preservation off the desktop?**

- a) ODF is the lesser evil here because it's patent encumbrances are lighter whereas OOXML's are a significant deterrent to confidence in long-term access to CONTENT and LAYOUT of files. If Microsoft stopped supporting OOXML in its applications, public access

would be cut off (It is necessary in such circumstances to take the worst-case scenario and therefore it is necessary to assume that Microsoft WILL END-OF-LIFE OOXML AT SOME TIME IN THE FUTURE IN ORDER TO FORCE FURTHER UPGRADES UPON ALL CUSTOMERS. SUCH ACTION IS EVEN MORE LIKELY IF MICROSOFT'S BUSINESS SHOULD WEAKEN AT SOME POINT. MICROSOFT IS ALREADY WORKING ON ITS NEXT FILE FORMAT PLATFORM STRATEGY WHICH IS VISIBLE IN THE NEW .NET SDK IF ONE FOLLOWS INFORMATION CONCERNING "FIXED" AND "FLOW" DOCUMENTS.)

b) The answer will also depend on what the associated APPLICATIONS can do, and that is unknown since we are dealing with scenarios that are decades into the future.

**Question 21. Re requirements of FOIL, litigation, etc?**

a) ODF has the edge because the application is freely available by download.

**Question 22. Re proprietary scripting functionality?**

a) Quite valid. Re-programming legacy VBA scripts and macros in migration settings will increase time and cost by a significant proportion and sometimes by magnitudes.

b) Equivalent tools for scripting in the open source environment or for OpenOffice.org do not exist (though IBM's Lotus Symphony has the Eclipse Framework), so there is a big question as to how one would go about replacing Microsoft-based custom functionality; assuming the why and cost factors are justified.

c) Often, however, scripting solutions on which people are mortally dependent are simply not necessary. So, a strict approach in triage during migration takes some concentration but mitigates some pain.

d) Many applications used in workgroups are simply natively developed, over-grown scripted solutions which were built poorly or on the wrong platform in the first place. A great many of these will need to be reprogrammed to the Web, anyway.

**Question 23. Re agency estimates for application re-writes?**

[no comment]

**Question 24. Re weight attributable to standards body?**

a) Attribute zero weight. Any benefit that ISO ratification for ODF version 1.0 may accrue to ODF would be negated by subsequent ratification of OOXML. Also, if OOXML should be fully and completely rejected in the February 2008 ISO BRM, OOXML would still penetrate into the market as a de facto standard. If the February 2008 ISO BRM is inconclusive about OOXML's future or if there is a trend toward merging the two standards, OOXML would still penetrate into the market as a de facto standard. Also, ODF version 1.2 (upon which late versions of OpenOffice.org have already implemented) will need to go through a long separate ISO ratification process itself.

Consider that Microsoft Office 2007 (which defaults to OOXML) has been released now for approximately one year and best estimates of its installed base in that time indicate that it has matched the total installed base of OpenOffice.org (which defaults to ODF).

b) The principal factors for determining the suitability of a format are its intrinsic characteristics of openness and freedom from monopoly-inducing or -supporting dependencies and encumbrances.

c) Standards ratifications in this day of broken and corrupted corporate-dominated standardization processes are not meaningful. Along with software patent reform, the standards field (apart from the W3C) is in bad need of repair.

d) This is why market measures based upon plug-ins and combined with W3C standards are the only solution to the problem of document formats.

**Question 25. Re would office suite standardization promote or stifle competition?**

a) Question should be rephrased: "would standardization by the State around a single format for office documents promote or stifle competition for office suite software?"

b) It would certainly and without the slightest doubt promote competition for office suite software APPLICATIONS.

The best predictor of what would happen in the market-place is to observe the e-mail application and the Web application markets and the rich variety of tool-sets which result when de jure standards are enforced and become and remain de facto standards in the market-place.

In e-mail we have ...

- Outlook
- Mac Mail
- Eudora
- Thunderbird
- Evolution
- emacs
- mutt
- pine

... each of which APPLICATIONS function interoperably around standard e-mail (document) format which is based upon universal plain-text. Absent standardization we would have a variety of e-mail applications which would only be able to read documents originated by the same brand/type of software. E-mail would not exist without effective, durable & sufficient standardization.

On the Web, we have ...

- Internet Explorer
- Mozilla Firefox
- Safari

- Konqueror
- Opera

... each of which browser APPLICATION functions interoperably around W3C-standard (X)HTML – despite Microsoft's best and continuing efforts to extend and break this standard format for its own business benefit.

If the office document format were successfully standardized, we can expect a wide variety of SOFTWARE APPLICATIONS to co-exist in the market-place with equal access and with performance differentiated not by features but by the implementation of the specific features which are frozen and dictated by the single standard format.

It is important to recognize that we have the opportunity to freeze the basic feature-set of documents and document authoring tools (this is good because, today, we know what the important basic features are); it is argued by some that this “stifles innovation”. However, it has been necessary to “stifle innovation” in e-mail and Web formats – in order to, in effect, stop improvement on the basic universally agreed functionality (by moving the process to a managed consensus environment; only the W3C has done this successfully; and the large IT vendors avoid the W3C in the cases of “standards” they wish to control .. for which they form their own standards bodies like OASIS, ECMA and others) -- in order to gain the interoperability necessary to increase the overall value of the total market. It is important to accept and fully understand that standardization is in fact the eradication of competition on select domains of technology, and this is done by choice. This is the purpose of standards processes and the result, when it is done well, is to create enormous growth & residual value in markets (e.g., the Internet).

It is fair to say that standardizing the office document format channels competition in a deliberate fashion to a desirable place in order to increase total competition among office suite application vendors. This is only painful to market participants who's positions and business interests favor no change in the status quo. It is further important to recognize that where standardization may diminish value to established businesses, the successful standardization of certain technologies will increase the value in the total system by many magnitudes greater than the value lost by established players and this is borne out historically in railroads, electric power generation, telecommunications, and many sub-areas of the technology markets as well as others.)

**Question 26. Re state's involvement in the creation of standards?**

a) This is a very apt question and one for which the answer is clear. States (not only US but all states globally) should participate in the standards-creation and -maintenance process through Working Groups created within a global standards body modeled on the W3C, if not the W3C itself. It is imperative that the states only enforce the use of these standards through state-based IT policy, and that the enforcement, per se, of standards development should strictly stay with the dedicated global standards body where all state governments would participate by consensus.

**Question 27. Re costs and benefits to the State of single format?**

a) Single format is the only way to go for quality reasons (which include productivity as well as security and economic efficiency); so the burden of cost/benefit analysis should fall upon the one party arguing in favor of multiple formats.

**Question 28. Re costs and benefits of multiple formats?**

a) The benefit to the Microsoft Corporation of state government acceptance of OOXML is quite exceptional. This plays the company's flagship profit-center franchise, Office, as well as most of the company's new products (which function around OOXML) into the frame for an indefinite period of time.

Conversely, the rejection of OOXML by state governments could precipitate a drag on Office 2007's (and other products') acceptance in the market-place and contribute to a business and financial crisis at the company.

b) Excess costs from multiple formats are immense and create an untenable situation to be avoided at great effort. Difficult to estimate, these costs are the costs forced upon the system of having to maintain two parallel but non-interoperable systems for document and records management & storage. It is analogous to having to take land away from citizens because the two different competing railroad companies need 150 feet of bed-width (instead of merely 75 feet) to facilitate 2 parallel sets of track of different gages & widths. Clearly this is not politically acceptable to the people – so long as they understand the issue and where the unnecessary cost burden ultimately falls.

c) Such costs are much more significant in magnitude for documents than the costs, for example, of maintaining multiple the measuring standards – where the standards are embedded in work at only slight cost of translation from time to time. Whereas parallel document standards require redundant systems -- which someone must pay for. (Large IT vendors – not just Microsoft -- are happy about this, in general.)

**Question 29. Re which option is most cost-effective? Why?**

a) The option of a single standard format is by far the less expensive alternative – in terms of software expenses, system expenses, personnel productivity opportunity costs and information access opportunity costs.

Please refer to the response to question 28, above, as to why.

**Question 30.** no comment

**Question 31.** no comment

**Question 32.** no comment

**Question 33.** no comment

**Question 34.** no comment

**Question 35. Re WordPerfect-to MS Office transitions relevance to today's migration?**

- a) That experience is highly relevant to today's situation of the need to migrate from MS Office 2000, XP or 2003 (legacy binary formats) on Windows to OpenOffice.org 2.x (ODF) on Windows.
- b) The legal profession represents a good place to research the experiences of that migration. WordPerfect offered certain features which made migration highly unattractive and disruptive in the legal profession, which explains why the law offices market vertical is WordPerfect's most significant customer segment today (quite a few offices never switched).
- c) Here is a page from the website of Fayetteville State University which provides an example of the kinds of Help resource used in the migration of WordPerfect to Word 2000 (<http://www.unctsu.edu/itts/training/wp2wd.htm> ).
- d) There is documentation of the ODF pilot process of over 300 pages residing with Massachusetts ITD. This document is held under terms of non-disclosure but addresses many dimensions of this question. Contact Linda Hamel, esq., General Counsel, Information Technology Division, [linda.hamel@state.ma.us](mailto:linda.hamel@state.ma.us) to request release, perhaps under agreeable nondisclosure terms between yourselves.
- e) In general, a state agency or group of agencies approaching an office suite migration such as WordPerfect-to-Office 2000 requires i) a (total) cost analysis; ii) software selection process; iii) a documented strategic migration plan & budget; iv) migration team and resources allocation (including space, geographic location procedures, team roles & hierarchy); v) protocols: management & user communication processes (about what is going to happen to their software & systems); vi) user Help & Support resources (a user-facing Help website and manned Support mailing lists; vii) user survey system resources (to gather feedback on user issues and for migration process optimization); viii) a software roll-out plan & team (assembled with local agency desktop administrators); ix) document conversion services and support; x) a businessprocess triage system (BP inventory, analysis, redesign & implementation); xi) software feature feedback process (vendor-facing); xii) user training resources & processes; xiii) bed-side manner, psychology and sleight-of-hand techniques to help users get through disruptive desktop change QUICKLY (for example, give out new or larger flat-screens to difficult users at the same time their software changes...they will invest more effort and also notice desktop differences less readily).

**Question 36. Re specific measures for NYS agencies' migrations?**

- a) Imperative #1: Stop any and all new Microsoft Vista installations or the introduction of Vista on new OEM hardware. A zero tolerance policy toward Vista procurement must be enforced across all New York State agencies or your ODF strategy is already 50% under water. If Vista is already coming in, give up your ODF migration plan and pass the document standardization idea to the likely Governor of 2022 – who is in about the 5th grade about now.
- b) Imperative #2: Stop Microsoft Sharepoint and new introductions of MS Exchange Server, as above, with extreme prejudice. These systems are toxic to Free Software, open standards, and to non-Microsoft software throughout much of your IT architecture.

c) For ODF, install OpenOffice.org 2.x on existing (Windows 2000 & Windows XP?) machines. When Windows XP reaches end-of-life (estimated 2009), your agencies will have been producing ODF for some time and their lack of dependency on Microsoft's proprietary formats will have neutralized that company's influence on your next desktop procurement decision. Buy Macs, Linux machines, or new Microsoft PCs, your choice.

e [sic] Consider carefully your "Co-Existence Strategy" -- i.e., respect the difficulties of installing OpenOffice.org on a PC while also leaving MS Office installed on the same machine. The concern is not for software or operating system conflicts but for the difficulties of enforcing "User Compliance" and full use of the new APPLICATION. Users tend to revert to what they know very quickly at the slightest problem of software feature differential, workflow stoppage, printing issue or document incompatibility and they commonly do not use Help systems unless well-trained to do so or unless the Help system is actually helpful.

e) Consider carefully the resource and political challenges of migrating state agencies whose geographically de-centralized and autonomous CIO or IT leadership do not support the concept of office suite migration. Plan and implement goals, incentives and penalties (carrots and sticks) to align the IT organizations state-wide.

d [sic] There is a possibility that plug-ins may become available in the future to enable ODF (or its successor Universal Document Format) within existing installations of MS Office on Windows.

Existing proofs-of-concept demonstrated, acknowledged "best" and supported by Massachusetts ITD leadership are known to produce document CONTENT and LAYOUT fidelity equivalent to the Microsoft OOXML plug-in that is freely available in the OOXML Compatibility Pack. If such an ODF plug-in is completed and made freely available, it is by far the least disruptive way of implementing ODF within document authoring SOFTWARE APPLICATIONS which are already in place and familiar to users, and would obviate the need for a disruptive office suite migration.

### **Question 37. Re validity of cost savings studies?**

a) All cost savings studies of which I am aware are biased toward their respective sponsors. Studies favoring OOXML are all paid-for by Microsoft or ISV participants in the Microsoft ecosystem (this is well-documented). Studies favoring ODF are either sponsored by IBM (which provides most of the impetus behind the ODF Alliance) or by the personal un-paid sweat equity of genuinely independent members of the open source and Free Software communities. I'm afraid you'll have to trust your human instincts and rely on your judgment as to which side's motivations might be a shade closer aligned with software quality and the flexibility which may come from not being locked in by a monopoly.

b) At one time (at least up to 2005), BECTA in the UK produced objective and helpful cost analyses, but I understand that organization has lately lost its objectivity (<http://news.becta.org.uk/display.cfm?resID=15524&page=1661&year=2005&month=12>).

c) There are great productivity gains offered by both XML document format franchises; one offers relative flexibility in software APPLICATION and system procurement

throughout the stack and across business processes and the other locks an organization and related populations into large and long-term financial commitments to a single-source supplier which looks to be binding for the next 15 or so years. It's a common-sense call.

**Question 38. Re studies shown OOXML has lower costs?**

a) I am not aware of any such studies, although IDC Norway authored an opinion piece sponsored by Microsoft which made some ridiculous assertions along these lines ...

<http://www.google.com/url?sa=t&ct=res&cd=1&url=http%3A%2F%2Fdownload.microsoft.com%2Fdownload%2F1%2F6%2Ff%2F16fd06b3-7059-4e21-adf4-9fbdc9a2853%2FRoadOpenDocStand.PDF&ei=c8yQRf6F6iMeuTKhQc&usq=AFQjCNEBtiIUpZDBI5YIiKA8ED92xdEnhw&sig2=W0nYnWfBX06S5ejG90yRHQ>

b) It is quite an audacious gambit of logic to assert that one thing which is essentially difficult to migrate to but free is more expensive than something else which is equally difficult to migrate to and yet has software licensing fees.

c) Microsoft has not widely publicized or marketed its next generation stack; but if one looks at the new .Net 3.0 SDK, it becomes clear how many different Microsoft products are touched by OOXML, where OOXML's embedded and unpublished extensions residing with Microsoft APPLICATIONS (other than and including Office 2007) are waiting to function as lock-in mechanisms for users of OOXML. The costs of adopting OOXML in terms of follow-on expenses for retaining related features and services off of the desktop, elsewhere and throughout the Microsoft stack have not to my knowledge been compared to the cost of deployment of an equivalent stack of services using open source or Free Software components. Obviously, such an analysis would be helpful, but Microsoft can't be expected to make visible a cost picture which argues powerfully against that adoption of its XML document format.

**Question 39.** [cannot usefully comment.]

**Question 40.** [cannot usefully comment.]

**Question 41.** [cannot usefully comment.]

**Question 42.** [cannot usefully comment.]

**Question 43.** [cannot usefully comment.]

**Question 44. Re premature time for deciding office suite formats?**

a) It's possibly the most important question put so far. Kindly see my response to Question #11, above. Arguably all parties should wait 2 or 3 years for the Web applications to mature.

b) If other platforms in the Office 2.0/3.0 arena are gradually maturing into useful range, it would be quite easy to "inject" the specification of the desired FORMAT into these APPLICATIONS. The trouble, as discussed above, is APPLICATION behavioral differentials which are not very effectively handled by today's best agreed format –

though use of generic tags is a good way to handle this problem while APPLICATIONS standardize their feature-sets for the single winning format.

Over and above the Office 2.0/3.0 question – which is an APPLICATION question -- I would be more concerned with the risk of picking ODF prematurely if a better format with better and broader governance should become realistic. My colleagues and I, for instance, have already identified W3C CDF as the best available standard framework for containing all formats across all desktops and all devices; and we can create systems to work from Microsoft formats or ODF to the Web-ready CDF framework quite readily to unify both old fat-client APPLICATIONS and new Web SaaS Office 2.0/3.0 APPLICATIONS, as well as play hand-held devices and smartphones in, all working around the single agreed open format.

c) Stepping back a few paces, we need a single STRUCTURED (XML) document format for the Web and fluid Web applications. The general challenge is to get legacy documents, legacy formats to be compatible and legacy APPLICATIONS to interoperate with new Web APPLICATIONS and convert documents (with minimized loss of both CONTENT and sometimes LAYOUT too) to the new single format that functions really well with all the new Web APPLICATIONS. ODF is too APPLICATION-dependent (a lot like OOXML) to offer the very best single structured format for the Web era. This is why it may be too early: Web APPLICATIONS are too immature and we haven't been working on a Web-ready UNIVERSAL DOCUMENT FORMAT for sufficient time to meet the requirements of the new environment and also bridge getting from the old to the new.

**Question 45. Re ISO ratification and revisions as reasons for prematurity?**

a) Nether a) nor b).

b) There is a superior/senior reason why a format decision may be premature: neither format merits adoption even though XML is a compelling new development.

**Question 46. Re timeframe for recommending an office suite format?**

a) If New York State takes the passive view, then the right timeframe is whenever a good enough Web-ready Universal Document Format becomes available.

b) If New York State takes the pro-active view, New York State can assemble other state IT leadership with experts and create the suitable format. Answer being, as soon as you like.

**Question 47. Re lead time for State and vendors?**

a) For the State, 6 months.

b) For vendors, it will take Sun & IBM approximately 3 years to adjust OpenOffice.org and Lotus Symphony for the necessary interoperability extensions which would enable ODF to work sufficiently with legacy Microsoft formats. Do not wait for vendors; their business plans are in conflict with the goals of system interoperability and the idea of a Universal Document Format.

**Question 48. Re “Intellectual Property” concerns?**

- a) Concern is legitimate.
- b) My brief on the problems with Open Specification Promise: “Analyzing the Microsoft Office Open XML License” (Jan 10, 2007, [http://fussnotes.typepad.com/plexnex/2007/01/analyzing\\_the\\_m.html](http://fussnotes.typepad.com/plexnex/2007/01/analyzing_the_m.html)) was widely cited.

This article is referenced in the whitepaper, “OpenDocuments and Democracy: A Political Basis for Open Document Standards,” page 20, by Laura DeNardis & Eric Tam, Yale Information Society Project ([http://isp.law.yale.edu/static/papers/Open\\_Documents\\_and\\_Democracy.pdf](http://isp.law.yale.edu/static/papers/Open_Documents_and_Democracy.pdf))

**Question 49. Re threat to the State of “Intellectual Property” litigation?**

- a) Other issues include the problem of unfavorable patent encumbrances which accost most technology products which are touched by OOXML throughout the Microsoft catalog.
- b) The patent issues are vast and demand holistic patent system reform. The best way to accelerate fundamental reform of the patent system is to adopt Free Software and open standards which challenge the premises of monopoly abusers who threaten legal aggression to stifle competition and innovation.
- c) **\*\* THIS IS NOT LEGAL ADVICE\*\*** It is unlikely Microsoft will sue customers over their deployment of competitive alternative formats. Doing so would force them to declare specificity, which would permit the Free Software community to code around specific patents or it would open legal challenges to the validity of specified patents, thus ending the usefulness to the monopolist of the perception of patent infringement to stall the adoption of alternatives.
- d) Look for 3rd-party patent trolling firms to pursue legal actions against IT vendors.

**Question 50. Re avoidance of liability?**

- a) Use W3C technologies, where the most extensive precautions against IP contamination of standards is effectively practiced, enforced and embedded in process.

**Question 51. Re protect citizens from liabilities?**

- a) Build & deploy standards plans around W3C technologies.

**Question 52. Re implication of document & record standards for discovery?**

- a) The issues are the same for discovery as they are for public access in general; though the fundamental imperatives for using an open structured document format may be generally stronger.

**Question 53. Re is PDF/A acceptable?**

- a) Yes. PDF/A is the very best format we have for general document archiving where fixed document CONTENT and LAYOUT is required.
- b) Dual-format strategies for archival purposes add complexity and cost as well as increase risk of human and technological error. Such risks need to be balanced in context of the value of the identified service.
- c) It is possible to create a format (or modify ODF) to zip all files by default to include a PDF/A file internally along side the content, styles and other file components. This would increase average file size but the overall cost of such an umbrella solution could possibly

**Question 54. Re compatibility issues pertaining to the litigation context?**

- a) There are no special issues here that do not apply to the general case of document compatibility with APPLICATIONS or APPLICATION-to-APPLICATION interoperability.

**Question 55. Re other formats?**

- a) The W3C's Compound Document Format ("CDF") because ...
- conformance is not optional
  - it is thoroughly APPLICATION-independent
  - it is based around XHTML 2.0 + CSS 3.0, which are open Web standards that are suitable for implementation
  - it is a framework suitable for containing other formats across desktop, device and Web
  - implemented properly, it is a candidate for the single Universal Document Format around which file compatibility and document application interoperability may be successfully achieved

**Question 56. Re validity of criticisms against OOXML?**

- a) criticisms of OOXML vis its unsuitability as a de jure standard are sound
- b) OOXML's strengths are that it is Microsoft XML implementation of XML for documents and therefore Microsoft can make its APPLICATIONS work optimally with both new and old formats.
- c) OOXML's weaknesses are that it is designed and in fact will perpetuate lock-in.

**Question 57. Re validity of criticisms against ODF?**

- a) There is not enough criticism of ODF, since the open source community, being lead by IBM and Sun in this case have created a gulag mentality that is intolerant of discussing ODF's limitations ... ( <http://www.robweir.com/blog/2007/10/cracks-in-foundation.html> ).

A neutral consultant described the character-assassination effort as “IBM closing off the exits”.

b) ODF's only strength is that its native application is available as a free download

c) ODF suffers from a surprising and alarming number of the same weaknesses as OOXML, making it a new lock-in paradigm designed to benefit IBM and Sun's software business strategies without due regard for the interests of software using customers and the principles of interoperability.

**Question 58. Re definition of “best value”?**

a) There is NO VALUE to changing to a format that is not a SINGLE UNIVERSAL DOCUMENT FORMAT which offers the highest possible levels of interoperability.

**Question 59. Re factors in cost?**

a) cost factors include ...

- software licensing fees
- migration costs
- user training (often overstated)

**Question 60. Re factors of efficiency?**

a) efficiency factors include

- worker productivity gains from better document compatibilities;
- worker productivity gains from better system/software interoperability
- worker productivity gain from standardizing on a common free software VERSION'S user-interface design across the agency;
- enterprise cost and effort savings from termination of lock-in (software budgetary and procurement flexibility improves)

**Question 61. Re factors of response?**

a) Open source is better in most cases, though Sun's response to bug reports on OpenOffice.org is so poor and Microsoft-like that it supports the argument that OpenOffice.org is not managed as an open source project (I am the ex-Marketing Project Lead of OpenOffice.org).

**Question 62. Re criteria for procurement objectives?**

a) software should be ...

- free

- Free (GPLv3 or similar)
- unencumbered by patent, copyright or trademark threats
- meet interoperability objectives (MUST import and export HTML 5.0, XHTML 2.0, CSS 3.0 and export PDF/A)
- formats MUST conform to W3C standards (ignore ECMA, OASIS, ISO standards which are compromised by vendors)
- formats must be developed by open & public consensus and openly documented, though the value of this is put in doubt by both ODF and OOXML and their respective standards consortia.

**Question 63. Re other issues?**

no comment

**Recommendations**

**Question 64. Re open source, visibility of source code?**

a) Yes. Absolutely! In all circumstances.

**Question 65. Re escrow of code?**

a) Not necessary if a proper Free Software license is used. Free Software is its own form of escrow. Anything less is not acceptable.

**Question 66. Re conversion?**

a) This is an interesting idea worthy of further exploration of its implications in specific contexts.

**Question 67. Re flexibility to change?**

a) No. Decide on a single UNIVERSAL DOCUMENT FORMAT and then DISRUPTIVE change will be unnecessary because a single universal document format is so tolerant by definition of choice of applications.

**Question 68. Re State influencing vendors on features?**

a) Yes, but it is not necessary to do this through any other means than by pursuing an open document formats policy correctly. The identified problem is addressed perfectly in the case of adoption of a single UNIVERSAL DOCUMENT FORMAT wherein if the State needed software applications to address particular needs, it could fund open source development through bounties, which would exclude no one from addressing the requirement and would maximize competition.

**Question 69. Re encouraging more features?**

a) No. Do you want the Web or e-mail to become more feature-rich? We have enough basic features in the document authoring software now to establish a fine single UNIVERSAL DOCUMENT FORMAT. From there, feature improvements can be made through any means so long as the UNIVERSAL DOCUMENT FORMAT's development consortium (quite likely W3C) continues to preserve the primacy of INTEROPERABILITY over feature innovations. This is the whole point. Features may be added to applications so long as they can be accommodated within the application-independent facilities of the format.

**Question 70. Re defining best value in the procurement context?**

a) Best value is defined by INTEROPERABILITY not by APPLICATION performance or features. Those may be added to an interoperable framework which is established by the document format.

**Question 71. Re the “no change” recommendation?**

a) No. There needs to be change simply to exclude non-INTEROPERABILITY options (in this instance both ODF and OOXML) and to define the new requirement while keeping agencies from upgrading unnecessarily until the true INTEROPERABILITY solution (a single UNIVERSAL DOCUMENT FORMAT) becomes available.

**Question 72. Re is a pilot recommendation viable?**

a) Yes, such a recommendation is possible though not very appealing. It would be helpful for New York State to pilot ODF-ready (or CDF-ready) software applications only if NYS were willing to build and improve upon the approach used in Massachusetts ITD. Repeating the Massachusetts ITD pilot would be an unnecessary waste of time, since the problems encountered in Massachusetts are already documented there (under non-disclosure).

b) The funds used in an ODF pilot – on the other hand – could be spent on developing and achieving definitively a single UNIVERSAL DOCUMENT FORMAT, all necessary plug-ins and server-side software to enable leading Office 2.0/3.0 applications to deliver an INTEROPERABILITY solution that would solve the document problem non-disruptively for the New York State and the world.

c) I am able to produce a comprehensive document formats pilot plan which would touch a variety of desktop experiences across multiple New York State government agencies, if the CIO/OFT feels that would be useful. Such a pilot design would incorporate full learning from the Massachusetts ITD pilot experience.

**Question 73. Re maintaining a legacy hardware-software conversion complex?**

a) This question identifies a commonly raised solution to the (theoretical) problem of varieties of formats remaining in circulation indefinitely.

b) An ideal solution for an enterprise like the State of New York would be to outsource any such document conversion services to a new or old entity in the private sector. If the need is valid, then the amount of work would support the viability of such a service – which may supplement such other services as Gold-, Silver-, and Bronze-level certifications for large document conversions for the legal profession, for example. It would be a shame to create such a facility in-house (redundantly in many state government agencies) and find out the expense was not justified.

**Question 74. Re centralized records management system?**

a) In light of data center consolidation trends, this is a very interesting idea – not without risk -- which justifies further inquiry?

**Question 75. Re other approaches?**

a) I believe neither ODF or OOXML are solutions which aspire to innovation or address with sincerity the confounding problems associated with document interoperability.

b) I believe this impasse is a function of risk-aversion among IT vendors and a trained-in lack of imagination among IT customers.

c) I am certain in my heart that the IT buy-side of state governments are capable of providing the leadership and vision to achieve a truly interoperable single UNIVERSAL DOCUMENT FORMAT to satisfy requirements. This will require imagination as well as a committed circumnavigation of the mind-numbing complexity – imposed deliberately by the status quo -- which drives innovative people away from this problem.

d) I am also certain that state governments are not actually cornered into a set of no-win unsatisfactory options – unless they themselves believe in being force-fed by the vendor community.

e) New York State can have a single UNIVERSAL DOCUMENT FORMAT if it wishes. Presently we are very close to a CDF demonstration of concept that will unlock this impasse.

**INDIVIDUAL # 56:** [INDIVIDUAL'S NAME REDACTED]: *Friday 1/18/08 2:53 PM*

**Question 1. Contact info:** [REDACTED]

I am an independent computer consultant and author of several books, including The Internet For Dummies Quick Reference, E-Mail for Dummies and Switching to a Mac for Dummies.

**Question 2., 3., 4.**

The most important thing is that electronic record policy is not a static problem that admits a one-off solution. Increasingly, the business of government will be conducted electronically and new technology will introduce new problems. A long term process must be established to provide ongoing guidance and protect the state from attempts by vendors to make the state dependent on proprietary technology. Ideally this would involve close coordination with other states and governmental entities.

**Question 5.** The most important thing is to insist on open standards.

**Question 6.** The state must develop a classification system to assign priority and importance of documents from a preservation and access perspective. Trying to put all data on the same footing will result in a system that is so unwieldy as to be equivalent to destroying everything.

**Question 7.** Long term data must be stored in multiple locations. One or two sites is not enough.

**Question 8.** no opinion

**Question 9.** The state should be fully aware that software vendors consciously attempt to get customers to adopt proprietary data formats in the expectation that they will eventually be able to extract monopoly prices. A firm commitment to open standards may have short term costs, but they pale compared to the eventual cost of proprietary formats.

**Question 10.** The state should pay close attention to standardization efforts for specialized data. In the interim, data should be preserved in its original format AND in the best available interchange format. All RFPs for specialized data software and hardware should address the question of long term access, open standards and interoperability. and require demonstrations of data interchange as part of the evaluation matrix.

**Question 11.** See answer to question 9.

**Question 12., 13.**

no opinion

**Question 14.** The state should consider the issue of real-time event-processing data, such as surveillance camera video, SCADA systems, toll and fare collection systems, license plate readers, RFID systems, security access/badge-readers, biometric scanners, medical monitors, etc. These will produce vast amounts of data and raise both long term access and privacy concerns.

#### **Response to detailed questions.**

**Question 1.** I found the distinctions a little vague. I would want to see a distinction between final documents, work in process and incidental chatter, for example.

**Question 2.** There needs to be additional categories of access for electronic records that distinguishes between

- a. human readable, i.e. on screen or in paper or microfilm printouts
- b. electronically accessible, e.g to search programs
- c. live data, such as a stored database that can be made operational, or a CAD drawing that can be used to derive other drawings, say for a renovation or addition or accident analysis.

**Question 3.** While privacy and security must primarily be assured by an access control layer that is distinct from document storage format, the enormous complexity of OOXML raises both privacy and security concerns. There have been a number of recent incidents where the U.S. government has redacted classified Microsoft Word documents to an unclassified version and then released them to the public in electronic form, only to find that the redacted classified portions were recoverable from the electronic, thereby revealing state secrets. Such lapses are much more likely to happen with personal privacy related information, where the oversight is likely to much less stringent than with national security information.

**Question 4.** FOIL accessibility will be enhanced by adoption of ODF, particularly for groups with limited resources. Computers are becoming available price points less than the cost of Microsoft Office. The low cost is largely due to the use of free open software, including Linux operating system and the Open Office suite. In addition, volunteer groups can rejuvenate older PC using the same software and make them available at little or no charge to groups with limited resources. Such users may be unable to read OOXML files.

**Question 5.** Again, the state should consider the fact that software vendors consciously attempt to get customers to adopt proprietary data formats in the expectation that they will eventually be able to extract monopoly prices. The use of formats that are not truly open means the vendor, not the state, ultimately controls the data.

**Question 6.** The definition of "interoperability" is suitable.

**Question 7.** The definition of "open standard" is suitable.

**Question 8.** n/a

**Question 9.** I think Gartner is over optimistic. There is significant push back from vendors against open standards because they have very large economic interests at stake. So I think 2012 is a more likely time for ODF to achieve Gartner's projection.

**Question 10.** My crystal ball runs on proprietary software and is down at the moment.

**Question 11.** no opinion

**Question 12.** Office suite is the tip of the iceberg. The state will have to follow open standard development in any number of areas, including medical records, human resource data, event processing, geographic information, CAD, semantic web, and no doubt many others I am not aware of.

**Question 13.** The large number of existing formats suggests the importance of defining workable open standards for universal document formats. While the problem of writing programs that convert from each format to the standard format may seem daunting, it is far more manageable than writing converter between the millions of possible pairs of documents. In more technical terms, the problem grows in proportion to the number of formats rather than the square of the number of formats.

Furthermore the format proliferation problem emphasizes the need for an open standard that is comprehensible, so format conversion programs for less popular or historic format can be written by volunteers with limited time and resources. If a standard that is open in name only, such as the

massively complex OOXML, is adopted, the state will largely be dependent on the resources of large corporations, who will abandon all but the most commercially important formats.

I regret that I do not have time to respond to additional questions.

**INDIVIDUAL # 57:** [INDIVIDUAL'S NAME REDACTED]: *Friday 1/18/08 4:31 PM*

I heard about the call for public comment through the Government Documents listserv, GOVDOCS-L. My opinions are my own. [CONTACT INFORMATION REDACTED]

**Question 15. What is the “problem” that this study should be addressing? Please define with specificity exactly what the State should be trying to solve.**

This isn't a singular problem, it's a web of issues. I'd rather not go into immense detail here, since I worry about finishing this prior to your deadline. To me, some of the larger less obvious issues present include

- \* Base Connectivity - The Digital Divide still exists in NYS, which I applaud Governor Spitzer for recognising.

- \* The Implied Capabilities of an Ordinary Citizen - We ought to be getting to deeper issues like Hargittai's Second Level Digital Divide.

[http://www.firstmonday.org/Issues/issue7\\_4/hargittai/](http://www.firstmonday.org/Issues/issue7_4/hargittai/)

- \* Diffusion Of Innovations - Rogers and Rogers set forth in very clear terms why good new ideas often don't take root

- \* Resource Sharing

- \* Inter Agency Communication

- \* Change Management Within the State Government

- \* Disabled and Minority Rights

**Question 14. Is CIO/OFT's proposed focus for this study appropriate? (That is, conceptualizing three types of “access,” and focusing on office suite formats as an illustrative example). If not, please describe with specificity the approach which you recommend CIO/OFT should take.**

The focus is not well defined. The format is such that neither large picture, general comments are elicited, nor granular fine detail responses.

Issues are not arranged in a fashion that facilitates meaningful dialogue. The questions jump from overarching principles to razor fine syntactic discussion.

The priority of the office suite and the focus on State Employees seems so heavy that I'm baffled why a simple electronic survey format wasn't utilised internally.

If you're truly soliciting Public Comment, go out into the Public Square. If I weren't deeply involved as a citizen and an information engineer, I wouldn't know about this process.

There are far too many questions for ease of response. This could have been broken down into theme based parts that were relevant to individual State Agencies so that citizens could select concerns tailored to their interests and complete the entire process little by little if they so chose.

There is not a single mention that I recall of a child's right to accessible government information.

The whole approach is very dry and unnecessarily pedantic. The feel is that of taking a standardised test.

References between Part I and Part II should be hyperlinked, particularly the definitions.

Bias is very detectable in the line of questioning.

**Question 1. Are the distinctions described in Part I of this RFPC between the definitions of electronic data, documents, and records useful? Are there any specific elements or distinctions in those terms which CIO/OFT should be taking into account?**

Defining one's terms is certainly a wise step for facilitation of a dialogue, but I wonder if such a syntactic focus deters an ordinary citizen from answering this request. I also wonder if the distinction itself matters in the larger sense. Are we tackling syntax at the expense of dedicating valuable time to actual transition and real world experimentation? I think that on my spectrum of usefulness (and I'm fascinated by complexity and topics that would put normal people to sleep) the distinctions are less useful and less relevant than the definitions for access.

**Question 2. Is the description in Part I of this RFPC of three types of access needed for electronic records – day-to-day utility access; ancillary active record access; and historical access – a realistic and useful conceptualization of the main uses of electronic records? If not, please describe with specificity recommendations for alternative methods for conceptualizing the study's issues.**

Again, definitions are important. I think the usage of metaphor as advocated by Morgan in Images of Organisation and Nonaka and Takeuchi in the Knowledge Creating Company could have got the point across less professorially and with a higher degree of interest. That would have resulted in more feedback. The alternative method I'd propose would be to let people offer their opinions in a free form brief fashion through an interactive web survey. You could split these definitions and ask something far simpler such as "Are you interested in preserving our State's Historic Treasures?" A yes response could funnel a respondent to a deeper set of questions like "Do you find that historical documents you try to find aren't digitised?" The survey could be broken up into groups of 10 or so questions. The final question for each segment could be "Do you wish to continue?" and it could serve as a save point. Respondents completing a certain percentage could be phoned for additional feedback.

**Question 5. In terms of appropriate "government control" of electronic records, what factors or concerns should the State be addressing?**

Personal privacy should be respected when the law would protect certain types of information. Apart from that, the public should be granted as much access as possible. With the opposition of these concepts in mind, I'd prefer the State to err on the side of sunshine.

**Question 4. Will accessibility to electronic records through the FOIL process be affected by adoption of either format, and if so, how? Will the rapidity of response required by recent updates to the FOIL law be affected?**

I should think that if data that does not need to be secured were warehoused in an open format in a logical fashion, accessibility would be quicker. A single State government portal, like the one that already exists can be very helpful for an advanced user, but great pains should be taken to make it simple for any user to find the information their "distressing ignorance" (Rubin) demands. People would FOIL less frequently if they could just find the information in the first place on their own.

**Question 3. Does the use of particular office suite formats such as the Open Document Format (ODF) or Office Open XML (OOXML) raise any security or privacy implications and, if so, what are they?**

I don't understand why this question reads as it does. There's no question I recall that says "Does the status quo suite raise any security or privacy implications?" It heavily implies that a newer open format would somehow be less secure than an older proprietary one. Given the rampant security problems persistent with Microsoft's products not to mention privacy issues with its snoopware, I'd rather the State stay away from that trough.

**Item D. Many State agencies maintain large-scale information systems designed for specific purposes (e.g., maintenance of birth and death records, processing of Medicaid claims), and it may not be possible to specify a single interoperability standard for such diverse systems.**

Not in all cases, but that's no reason to not strive for interoperability when it's practical and helps a citizen or a State employee do his or her job better or faster.

**Question 6. Is this the correct definition of interoperability which the study should be using? If not, please provide a better, alternative definition.**

I almost said yes until plug ins and patches wandered through my mind. That "without modification" makes for a climate hostile to innovation. Why wouldn't you improve upon something if there were no monetary outlay for the improvement? We should strive for native or out of the box interoperability based on logical standards, but those standards come from some place. There was a problem, someone solved it, folks said "Hey! Great idea, can you always do things that way so we can share?" Many Libraries have different heavily customised catalogues, but they can exchange data through the NISO z39.50 functionalities built into different products. That sharing was born of a standard format for records, MARC, which was born of the catalogue. Any definition needs give leeway for improvement through innovation when that improvement saves the time or money of an Agency or citizen or improves the quality of service. So how about just "products and systems from multiple vendors that can be used together. Products requiring less modification, customisation, and training are more desirable."

**Question 7. Is this the correct definition of "openness" and "open standards" which the study should be using? If not, please provide a better, alternative definition.**

\*whistle\* That's one mighty fine definition ye got there, I reckon. I think I'd only strike "negligible fee" from it. My negligible fee and Mark Shuttleworth's would differ, methinks. I'm skipping 8.

**Question 9. Is Gartner's prediction correct? What predictions have been made about other formats?**

We'll have to wait til 2010 to find out. I hope that it bears out or is exceeded. Predictions have a terrible flaw of not always coming true; How to Lie With Statistics, Alan Greenspan's advice, and the recent New Hampshire poll debacle come to mind.

**Question 10. Will the usage of ODF among those individuals and entities with whom the State interacts be so great that failing to provide the NYS workforce with the capability of using ODF will cause NYS interoperability problems? If so, if the State did not adopt the ODF format, what would be the best method to ensure interoperability with ODF documents received by the State from others?**

This smacks of a competitive intelligence question, which is something rooted in for profit economics. I'm not sure why one would wish to relate that sort of issue to the non profit environment of the State. Ranganathan's Save the time of the user becomes save the time of the taxpayer. We should not hold hostage the State of New York to the whim of a for profit vendor. Corporations should take a back seat to individuals. Using an open standard is less likely to present an interoperability problem than the current stagnant quagmire of proprietary systems. The best method is to experiment with a product before one is forced to adapt it. Move left on the technology adaptation curve, NYS government luddites! :) You could ensure this happens somewhat smoothly by experimenting with a hypertext hierarchy like Nonaka and Takeuchi advocate. Task forces could investigate different problems. Given employee battles, some sort of neutral arbiter would need to select members for a given task force. Patronage and office politics are a problem!

**Question 11. For office suite formats, which governments have adopted ODF exclusively? Which governments have adopted OOXML exclusively? Which governments have adopted both formats? What other formats for office suite software besides ODF and OOXML have other governments adopted?**

I was paying attention. That would be South Africa. You cite wikipedia at me, and I countercite:

[http://en.wikipedia.org/wiki/OpenDocument\\_adoption](http://en.wikipedia.org/wiki/OpenDocument_adoption)

I raise you a more relevant question - What is being done to encourage cooperation in this endeavour with Massachusetts seeing as how we're neighbours?

**Question 12. Other than in the office suite context, in what other ways does the State need to be concerned about electronic records interoperability?**

Call me crazy, but maybe child sex offender registry data should be more readily interoperable than office suite content. Transportation department data? Certainly Library data can and should be interoperable at this stage of the game. It's scandalous that America doesn't have a National Library Catalogue for public Libraries, much less is struggling to get State Catalogues in place. Health records should certainly be of concern provided good security. All emergency

response data should be interoperable and particularly it should work both within our jurisdiction and across geographical boundaries. Voting data...

**Question 13. Given the existence of tens of thousands of e-data formats, the increasingly dynamic nature of electronic documents, and a preference toward more open formats in other realms besides office suite formats, what type of an approach or mechanism should be used within the State to further the existence of openness in all relevant formats? Please describe with specificity.**

This seems like a job for machine work! How about implementing something sorta like the cataloguing Authority Check function of the Xtensible Catalogue project at the University of Buffalo et al to go out, crawl for extensions, prioritise them by frequency, disambiguate, and then queue em up for review and work on down the list?

**Question 16. If determinable, what percentages of current formats do you have in your systems, e.g. what percentage of your digital data is in the common office suite formats, e.g. .doc format? .xls format? .ppt format? .rtf? .pdf? .html? .txt? .wpd? etcetera. To what degree have you already migrated to XML-based formats such as .docx, .xlsx, .pptx, .odt, .ods, or .odp, or what are your plans to do so? What tools do you use to determine the mix of formats being used within your systems? Anyone can respond, but we are particularly interested in learning the experience and current situation of governmental responders, and particularly from state and local governments.**

This is a plain old text file. I'm at home and a luddite, though. At my last job, my Library's Patron Public Access terminals all had open office. Some kids preferred creating textfiles. This saved us money and equipped folks to be able to take things with them or email stuff, et cetera. I'd say the overwhelming application was word processing. Kids would very frequently use gimp or paint to create custom backgrounds, too.

Historically, good ideas will jump the format chasm or ones that contain sentimental value.

**Question 17. Assuming this observation is correct, please provide a numbered list, with the greatest specificity and in the simplest terms possible without marketing verbiage or usage of ambiguous phrases, of exactly which customer requirements are best met by OOXML.**

Seeing as how it couldn't pass the ISO ballot and I was clear in advocating for standards based innovation, 0. I can't even claim that it would help in discovery since I recall the word count error with Word.

**Question 18. Assuming this observation is correct, please provide a numbered list, with the greatest specificity and in the simplest terms possible without marketing verbiage or usage of ambiguous phrases, exactly which customer requirements are best met by ODF.**

Since this did pass an ISO ballot, I would be comfortable with it for preservation's sake. This is particularly true with Australia's backing. I think it's too early yet to settle on just these two, and the testing and prioritisation burdens more properly rest with your Agency since each Agency will vary in its response. A populist approach should be taken with the needs of the agencies catering to the widest spectrum of the citizenry taken first (Libraries, Health, Education, Dormitory, Emergency Response, OGS, Parks, Courts et cetera. Not the old standby of the Governor, Legislature, Courts, et cetera.)

Well, it's almost the deadline, so off it gets sent :)

In closing, use LOCKSS for preservation!

[http://www.lockss.org/lockss/How It Works](http://www.lockss.org/lockss/How%20It%20Works)

As a citizen, I'd rather have an Agency work towards actual change and have to apologise later if there's a bump than have an Agency study something until it dies a terrible, horrible death by committee.

I regret that I do not have time to respond to additional questions.

**INDIVIDUAL # 58:** [INDIVIDUAL'S NAME REDACTED]: *Friday 1/18/08 5:40 PM*

In response to CIO/OFT Request for Public Comment # 122807

Dear Director Melodie Mayberry-Stewart,

I write to persuade you that the issue of "open formats" for Jumbo Office Suite software(s) is not centrally a question about fixed formal standards definitions. By formal standard I mean a standard promulgated by formal standards bodies, such as ANSI, ECMA, ISO, OASIS, and bodies like them.

My argument is in aid of my recommendation: that New York State never force OOXML upon its residents nor upon its workers. A formal document standard can be of some help in attaining the ends set forth in the Call for Comments. But my central thesis is that, for things like Jumbo Office Suite formats, only free software can defend open standards.

Claim: any fixed standard for Jumbo Office Suite document format will fail, on its own, to do much for interoperability, even if all parties producing software which use the formats, adhere strictly, formally, and fully, to the fixed standard. The reason is that Jumbo Office Suite documents are in principle and in practice substrates for many different kinds of computer programs, which do many different things, in service of many different human ends. For example:

1. a document might contain a spreadsheet which is updated over the Net, so that the numbers in the cells of the spreadsheet are kept up to date
2. a document might be subject to automatic translation between different human languages, and even between different sensory modalities, such translation being done, say, by a read out loud program, so that a blind person can hear the contents
3. a document might, in concert with remote cryptographic time stamping and authentication services, be available in all its historic drafts, and comments included with cryptographic signatures of the authors of the comments

These are but a few of the things which already we do with such flexible documents. In future, as you point out, we will do more. (Indeed it is clear that we should be careful to distinguish documents so susceptible to plastic deformation and old fashioned hard to modify things such as

sheaves of ink marked paper. I know, by your request for comments, you are acutely aware of the importance of this fundamental difference.)

Why cannot a full formal standard and full honest adherence to the standard by authors and vendors of Jumbo Office Suites come near to guaranteeing interoperability? It cannot because a standard for a Jumbo Office Suite format is not at all like a standard for wood screws, concrete mixtures, use of a radio band, ethernet cable, fatty cheeses, arterial stents, codes for computer storage of text, steel girders, electrical power cords, etc.. A department of the New York State government might need some wood screws, and the department might require them to meet some ANSI standard. A vendor offering some screws will have had certified the compliance of the screws to the standard, or in some cases, the department's inspector of materials will do the tests to assure compliance. Now, assuming the screws pass the tests, the screws are very likely to be adequate to the job. For this case, a formal standard is a powerful tool, a tool of human organization, in aid of efficient, ah, say, furniture repair. Costs of organization and coordination, often called "transaction costs", are reduced by the existence of a standard which buyers and sellers adhere to. In this case the reductions in cost are large.

But a wood screw is different in kind from a "document" in a computer. Such a document is not like a piece of paper with ink marks. It is rather a nexus for literally thousands of programs which may operate on it and which may take it as an input and which may use it as a place for output. And by application of the famous two laws of computer programming, that all programming difficulties may be solved by one more level of indirection, and that caching often really helps, the document itself will often contain executable code, or will be able to call upon programs in its environment. Some of this code and some programs in the environment will, on occasion, or even just about every time the document is used, be run. (Again, as the call for comments implies, the present culture of computers, in most businesses and government offices, does not sufficiently deal with the manifold problems of usability and security caused by such a rich and flexible document architecture.) And often office workers whose primary job is not computer programming will write executable code which will become part of the document: the paradigm example is spreadsheet code, but there are other examples, such as embedded TeX, automatic calls to external data sources, bits of Javascript, etc..

So a Jumbo Office Suite document is a central part of a large flexible programming system. The published format cannot capture with sufficient exactness the actual interfaces and actual behavior of any particular system claiming to adhere to the standard. Every programmer knows the phrase of art "bug for bug compatible". The phrase refers to this fact of incompleteness and inaccuracy of all formal published specifications for large software systems. If we have one team, call it the source secret team, which writes a big system which reads and writes documents to some specification, and copies of the software are sold under a restrictive EULA, in particular the software is source secret, then the source secret team has an enormous advantage over any competing team, call it the free software team, which offers their system under a free software license. The source secret team gets to debug both the output side and the input side of the document interface. But the free software team gets to see the documents output by the source secret software and that is all. The free software team does not get to see any of the code of the source secret team. Every programmer who has worked on a large system knows that being able to design and to debug on both sides of any interface is a large advantage. Thus the source secret software works well only with itself. The source secret software will not work well with other software and other software will not work well with it. "Vendor lock-in" comes about naturally as a consequence of the use of source secret software.

The free software system is radically open to improvement and to innovation, because any author or group can grab the code, study it, discuss it openly, and write better code, and write code that does new things. But the source secret system is under the full legal control of the entity with copyright on the code. You are not allowed to publish a bug fix to Microsoft's word processor program. You are not allowed to even dis-assemble the code and discuss it openly. In practice this means that only Microsoft can fix bugs. It would be absurd for New York to hand to Microsoft full legal and practical control over all of New York's government documents. If New York uses free systems, then, if the vendor does something the department does not like, well the department can hire another vendor, or in house people, to do what needs to be done. To repeat, this is legally prohibited in the case of any source secret system.

We have examples of solid standards which do not depend upon a central formal fixed standard. We have the GNU Compiler Collection of Project GNU, gcc is a real standard, and the Internet itself. Here is Project GNU on what freedom of source means to writing C code:

From the GNU info node "system calls" under the node "program basics" under the "Libc" node:

```
<blockquote  
  to-be-found-at="http://www.gnu.org/software/libc/manual/html\_node/System-Calls.html"  
>
```

```
The description of `syscall' in this section assumes a certain protocol for system calls on the various platforms on which the GNU C library runs. That protocol is not defined by any strong authority, but we won't describe it here either because anyone who is coding `syscall' probably won't accept anything less than kernel and C library source code as a specification of the interface between them anyway.
```

```
</blockquote>
```

The GNU Project with the Linux kernel, and with literally hundreds of other projects and many thousands of paid and un-paid workers has produced a complete operating system, indeed, a family of operating system, over the past twenty some odd years. There is no formal standards body like ANSI that assures interoperability. Yet interoperability is attained and maintained. What underlies this achievement is the openness of the whole project. There are no formal impediments at the level of being able to study, discuss, improve, extend, and publish code. Whether a proffered contribution is accepted is decided by the leaders, by other authors, and by the users. The Internet was built and is today maintained by a similar process. Both examples show that real interoperability does not come from an impossibly precise, impossibly accurate, and impossible to enforce, formal standard, but from the free cooperation and the free competition of individuals, tribes, organizations, companies, and even, government departments.

The formal ODF standard is powerful because it is a formal expression of a general intent to cooperate, backed by the existence of large free systems, a large number of users and authors, and some large companies. The formal ODF standard by itself is nothing much. But New York State may, by its choice of the ODF standard, and by its use of free systems which use the ODF format, do much to assure real interoperability.

I remain, as ever, your fellow user of free software, and fellow student of probability, [CONTACT INFORMATION REDACTED]

**INDIVIDUAL # 59:** [INDIVIDUAL'S NAME REDACTED]: *Tuesday 1/29/08 12:54 PM*

## **PART I - GENERAL QUESTIONS**

**Question 1: Contact Information** [REDACTED]

**Question 2: What mechanisms and processes should the State of New York establish for accessing and reading its electronic records in order to encourage public access to those records?**

The State of New York should make it a priority to make most electronic records accessible in a manner that is at least as transparent and universal as a paper record. State records that predate the US Constitution are available in the state archives today -- is it reasonable to assume that a Microsoft Word 2007 document will be readable in 2207? Would it be reasonable to expect a State archivist to read tape media from a 40-year old IBM System 360? The only requirements on the part of a citizen to read a paper record are the ability to see and ability to read. Electronic records should maintain the fundamental openness of paper documents while at the same time using technology to make documents more accessible to citizens with physical limitations and more semantically accessible by electronic indexing.

**Question 3: What mechanisms and processes should the State of New York establish for accessing and reading its electronic records to encourage interoperability and data sharing with citizens, business partners and other jurisdictions?**

Adopting open data formats where appropriate will facilitate interoperability and sharing by making it feasible to exchange information without the need for a common application platform. Instead of requiring all business partners, jurisdictions and citizens to use specific proprietary software packages, the state can simply require that all entities use software that can read an open data format.

Basing requirements on data standards rather than vendor/software standards is important because the application marketplace is a fundamentally unstable. In 1990, Wordperfect was clearly the dominant word processing application. In 2007, Microsoft Word is the market leader. In 2020, another product, perhaps one conceptually similar to the "online office" suites being marketed by Zoho, Google and others may become dominant. The State of New York doesn't drive the software marketplace and cannot predict the future. Therefore, it is important that records be maintained in formats that are fully documented, developed in an open, vendor-neutral process, and either not encumbered by patents or licensed under conditions that grant free use of patents.

**Question 5: What mechanisms and processes should the State of New York consider for encouraging choice and vendor neutrality when creating, maintaining, exchanging and preserving its electronic records?**

The State should define record policies based on the nature of the data and the requirements of the business, and not on specific vendors, hardware or software. Embracing open standards for data formats and open source software for the creation, maintenance, exchange and preservation of data where appropriate will guarantee long-term retention.

**Question 6: Are there mechanisms and processes that State of New York should establish that are specific to the management of its electronic records in its various life cycle stages?**

**Question 7: How should the State address the long term preservation of its electronic records? What should the State consider regarding public access to such archived content?**

As addressed in Questions 2, 3 and 5, using open formats for the storage of government records increases the. [sic]

The State should formally study whether storing records that require permanent retention electronically is a viable strategy.

Long term preservation of records that will have lasting historic value may preclude long-term storage in electronic format. Organizations with large collections of digital assets ranging from NASA to members of the Motion Picture Association of America<sup>1</sup> have found that storage of digital media costs an order of magnitude to store than traditional media, and may prove to be less reliable.

Public access to. [sic]

## **PART II - DETAILED QUESTIONS**

### **Question 2:**

Automated or programatic access to data should be considered as a distinct access category. With paper records, locating information is dependent on the quality of the index, external catalogs (both of which are manually created) and the skill of the user, librarian or archivist.

With electronic records, search and indexing technology has advanced to a point that software can allow for automated searches that take semantic factors into account and thus can make information more accessible.

The effectiveness of automated search and indexing technology is heavily dependent on factors such as data format and structure, and State policy regarding record creation, maintenance and archival should take those factors into account.

Question 3. Does the use of particular office suite formats such as the Open Document Format (ODF) or Office Open XML (OOXML) raise any security or privacy implications and if so, what are they?

Formats that are overly complex present serious security implications. The Office Open XML (OOXML) format, for example is over 6,000 pages long, and includes a variety of obscure and incomplete elements. autoSpaceLikeWord95 is a well-publicized example of an OOXML element of this type:

*2.15.3.6 autoSpaceLikeWord95 (Emulate Word 95 Full-Width Character Spacing)*

---

<sup>1</sup> Cieply, Michael; The Afterlife is Expensive for Digital Movies; The New York Times; December 23, 2007; <http://www.nytimes.com/2007/12/23/business/media/23steal.htm>

*This element specifies that applications shall emulate the behavior of a previously existing word processing application (Microsoft Word 95) when determining the spacing between full-width East Asian characters in a document's content.*

*[Guidance: To faithfully replicate this behavior, applications must imitate the behavior of that application, which involves many possible behaviors and cannot be faithfully placed into narrative for this Office Open XML Standard. If applications wish to match this behavior, they must utilize and duplicate the output of those applications. It is recommended that applications not intentionally replicate this behavior as it was deprecated due to issues with its output, and is maintained only for compatibility with existing documents from that application. end guidance]*

With hundreds of obscure and undocumented elements to implement in software, any application implementing the OOXML standard will have a large “surface area”, which makes it more likely for obscure security or other flaws to manifest themselves.

Complex standards also serve as a disincentive for the creation of new office applications. ODF is a capable, but relatively simple standard that has been implemented in dozens of commercial and open source applications. If one application has a security or privacy-related flaw, users can choose to adopt an alternate application. OOXML has only been fully implemented by Microsoft Office, so users do not have that choice.

**Question 4: Will accessibility to electronic records through the FOIL process be affected by adoption of either format, and if so, how? Will the rapidity of response required by recent updates to the FOIL law be affected?**

The Open Document Format (ODF) should improve the State's ability to respond to FOIL requests in a cost-effective manner, as the availability of open source code to read and manipulate ODF data means that a wide variety of competing vendors will be able to produce software to search archives of ODF data. Additionally, since ODF is a fully open standard, state programmers are able to create specialized search software to meet needs specific to the State of New York.

**Question 5: In terms of appropriate “government control” of electronic records, what factors or concerns should the State be addressing?**

The need for “government control” of electronic records should not conflict with the state's duty to preserve and archive data. Any controls (ACLs, encryption, etc) that are implemented should be based on open standards, and preferably be accessible from multiple vendors or software implementations.

For example, using a proprietary digital rights management tool that encrypts, restricts access or defines the actions that users may perform on a record that is specific to a particular software package or version may severely limit access to data down the road.

The experience that major software vendors have had with electronic music and video sales provides a frightening example of this risk. Microsoft created a DRM standard called “PlaysForSure” intended to protect the integrity copyrights for music sold over the internet. PlaysForSure is supported by a marketplace with a number of competing online music stores and vendors producing portable music players.

When Microsoft's marketing strategy shifted, the viability of that protected data was called into question. The Microsoft Zune player and music marketplace uses a DRM system that is not compatible with the previous PlaysForSure standard. Users who purchased music with the old system will not be able to access their data (music) as hardware and software becomes unavailable.

If the state was faced with a similar situation with electronic records, it would be forced to convert data en masse to the new format, or risk losing all access to the data.

**Question 6: Is this the correct definition of interoperability which the study should be using?**

Yes.

**Question 7: Is this the correct definition of "openness" and "open standards" which the study should be using?**

Yes, with one modification. I recommend changing "there are multiple implementations of the standard" to "there are multiple, independent implementations of the standard".

**Question 9: Is Gartner's prediction correct?**

Predicting the future is by nature more art than science, and I do not possess any special skills to enable myself to predict the future more accurately than Gartner. The Gartner prediction does, however, grant more credibility to the notion that the marketplace wants a platform independent format for storing data.

The State of New York is a major producer of documents and other content, and if the State officially adopted and promoted ODF among other jurisdictions and business partners, you would see a dramatic increase in ODF usage.

**Question 13: Given the existence of tens of thousands of e-data formats, the increasingly dynamic nature of electronic documents, and a preference toward more open formats in other realms besides office suite formats, what type of an approach or mechanism should be used within the State to further the existence of openness in all relevant formats?**

In large measure, the marketplace is mandating a shift to open formats in most market segments. The standardization of Email and the creation of HTML is widely recognized as factors for the revolutionary advances in data delivery and exchange that have taken place in the last fifteen years.

If the State wishes to shift towards increasing the use of open standards and formats, it will use a variety of approaches, such as:

- Add provisions to procurement processes to require agencies to financially justify the use of proprietary data formats.
- Engage in partnerships with the State University, system integrators and other stakeholders to educate State IT and business leadership about the advantages of adopting open formats.

- Establish a award/recognition program for project teams that implement solutions using open technology.

**Question 15: What is the “problem” that this study should be addressing?**

The problem is that the State of New York may not be fulfilling its duty to citizens to maintain records in an accessible and cost effective manner. A few questions that this study should answer:

- How much does it cost the State to enable employees and local government to exchange information using selected proprietary formats?
- If an open standard with multiple implementations, such as ODF, is available, what advantages does a proprietary standard deliver?
- What is the value of these advantages?
- Is the added cost of proprietary software recaptured by productivity or other savings?

**Questions 19-21: Regarding ODF vs. OOXML**

Open file formats are separate and distinct from applications. For example, Microsoft publishes a plug-in for Office to produce PDF documents. ODF plug-ins are available from Sun and Microsoft/Novell (upcoming) for Office 2007. OpenOffice can save documents in PDF and can load, save and edit legacy Microsoft Office formats.

**Question 22: Integrated Applications Microsoft Access, which is included with Microsoft Office, is a lightweight application and database management system, and not a document editor.**

Adopting an alternate office application would certainly break embedded ActiveX controls and VBA code. But using OOXML and using Microsoft Office will break these applications two years from now when the new version of Office is released. VBA and certain ActiveX elements are considered legacy technologies by Microsoft and will not be usable with future releases of Microsoft Office.

The situation that many Microsoft customers, including NYS are facing with respect to VBA is a perfect illustration of why adopting closed technology for mission critical applications is a bad practice.

**Question 25: For office suite software, would standardization by the State on the usage of a single format promote or stifle competition in the IT marketplace?**

Standardizing on a standard, truly open file format would promote competition in the IT marketplace. All vendors, including Microsoft, are able to access the specification and produce applications with adhere to it.

Standardizing on a somewhat-open file format, such as OOXML, which contains hundreds of references to proprietary code which is held only by Microsoft would stifle competition, as only one vendor (Microsoft) would be able to produce a fully-compliant application.

**Question 26: If standards were developed regarding the creation of electronic records in State government, how would they be enforced and who would be or should be responsible for enforcing them? Should NYS Archives be given enhanced enforcement authority?**

One potential solution would be for NYS Archives to devise a standardized, “free” method for state agencies to submit open/approved/preferred file formats to the archives. If other state agencies choose to adopt other formats, those agencies will need to convert the proprietary formats to meet standards defined by the Archives.

Another potential mechanism to encourage compliance would be to require agencies to budget for the cost to convert data to an open/preferred format during the Intent to Purchase phase of procurement.

**Question 35: To what extent does the WordPerfect to Microsoft Office transition serve as a viable migration model?**

Many aspects of the WordPerfect to Microsoft Office transition may be invalid as a migration model for several reasons:

- It was a transition between two commercial products, while most organizations implementing ODF are moving to OpenOffice, which is a free, open source product.
- The anti-competitive practices engaged in by Microsoft at the time make it difficult to establish what the actual cost of Microsoft Office was at that time.
- Many organizations who switched from Wordperfect to Office weren't completely converted to desktop computers at the time. In the early 1990's, depending on the agency, a substantial portion of the government workforce was operating with “green screen” terminals or on paper. Some New York County Social Services departments were mostly terminal-based as late as 2003.
- Email was not as pervasive during the early-mid 1990's, and Email/Groupware products were typically sold separately.

Other aspects of the Wordperfect to Microsoft Office transition are valid; the ancillary costs of integrating Wordperfect, Microsoft Office or both applications is and was minimal, with the exception of training.

**Question 40-41: Which format currently will better facilitate access to electronic records through the use of assistive technologies? Which is best positioned to provide such access in the long term? Would adoption of ODF be acceptable to conversion to other formats was available?**

Assistive technologies are not relevant to document format. Applications such as Microsoft Office, Apple iWork or OpenOffice provide an API to enable assistive technologies. If Microsoft Office provides a better interface for assistive technologies, than a user could use an ODF plugin to access documents in Microsoft Office.

Features such as assistive technology support are a valid reason to pick a product when choosing among proprietary applications or between proprietary applications and open source. That is

precisely why encouraging a competitive marketplace by using standardized file formats is a good idea.

**Question 42: Should the State be engaging in an initiative similar to that described in the Massachusetts MOU?**

Yes, but the Massachusetts ODF standardization effort did not adequately communicate the cost advantages inherent in adopting open standards and open software. If New York were to engage in a similar effort, the State should make an estimate estimate how much money is spent on the purchase, maintenance, and administration of Microsoft Office over a 3-5 year period.

**Question 44: Is it true that setting an office suite software format standard would be premature because other more interactive platforms will soon be surpassing in usage ODF- and OOXML-using software?**

No, while increased competition in the browser marketplace has made interactive, applicationlike websites more prevalent, a common data format is still relevant. Early "online office suites" such as Google Docs & Spreadsheets and Zoho Office provide Office-like functionality, including the use of documents in Microsoft Office, ODF, and PDF formats.

Adopting a truly open file format is even more relevant with online, interactive platforms, since the cost of switching from one web-based product to another is nil.

Consider how the NYS Department of Civil Service administers mail and web-based prescription drug providers for employees. Doctors use a legally mandated, common prescription format, and pharmacists who understand that format work for all pharmacies.

Therefore the business of filling prescriptions is now a commodity business. Every year the department request bids, and whomever comes in lowest gets a one-year contract. Why should office suites be any different?

**Question 45: Is it true that setting an office suite software format standard would be premature because (a) ISO- standardization has yet to fully play out for OOXML format, or (b) ODF format is undergoing revisions?**

Absolutely not! File formats are never static -- revisions are part of a natural process that takes place all of the time as products and requirements evolve. Consider the changes that have taken place between Microsoft Word 6.0 and Word 2007.

(a) Microsoft has already published OOXML as an ECMA standard, and it can be evaluated today.

(b) Both ODF and OOXML will continue to undergo revisions as long as they are viable formats. That revision/evolutionary process should not preclude either format from consideration.

**Question 61: To what extent should the State be concerned, or reassured, about the availability and reliability of maintenance and support from vendors of OOXML/ODF format-using software?**

In order to answer such a question, the State will need to determine what exactly its support needs are.

**Question 63: What other issues has this RFPC omitted which the State should be considering as it conducts this electronic records study?**

- The State of New York has a well-defined and understood approach to procuring products, including software. The State does not have a clear methodology for adopting no-cost open source software, which in the case of ODF, is tightly associated with open file standards.
- The RFPC doesn't provide sufficient background regarding how electronic records are currently retained. If the State has a system that works well, will choosing OOXML versus ODF affect that system negatively?
- The actual or reasonable estimates of the initial and ongoing costs of Microsoft Office statewide are completely unknown. The lobbies for ODF and OOXML have vested interests in overstating or understating the current costs, so the State should try to determine its actual cost.
- There may be compromise solutions such as using display formats such as PDF or the PDF/A format for archival and information exchange purposes may present a way to allow state agencies to continue to choose document creation and editing software, whether that software uses ODF, OOXML, or a legacy format.

\* \* \* \* \*

## Part III-C

### Public Comments Received: Government Responses

See Next Section: ➡