

**BACKGROUND ON MICROSOFT'S INTEROPERABILITY EFFORTS
AND SUPPORT FOR FILE FORMATS**

**Submitted by Microsoft Corporation to the Danish Competition Authority
in Relation to the DCA's Sector Inquiry Concerning Office Software
23 February 2009**

Microsoft is separately responding to the questions provided by the Danish Competition Authority ("DCA") in relation to its sector inquiry concerning office software as ordered by the Danish Parliament. Microsoft understands that this inquiry is principally focused on the nature of competition concerning office software, including interoperability between competing office suites.

Because the questionnaires do not specifically address interoperability issues, which we understand to be a central aspect of the DCA's inquiry, in this paper Microsoft provides additional background concerning what it has done to promote interoperability among office software applications.

In summary, Microsoft is doing this in a variety of key ways. With the next update to Microsoft Office 2007, users can view and create (1) an ODF document, (2) an Open XML document, (3) a legacy Office binary file format. Microsoft Office supports more than a dozen other formats, including PDF. With the standardization of both ODF and Open XML, and Microsoft's comprehensive documentation of its binary formats, vendors and users can achieve a high degree of interoperability across their office software products. Details on these and other efforts are provided below.

I. INTRODUCTION

Microsoft is committed to promoting interoperability between its office software products and third-party software products. We recognize the value our customers place on interoperability for our mission-critical products. Microsoft recognizes that many of our products, including Office, have become so important to customers' businesses that interoperability and data portability solutions are highly valued. To promote such interoperability and ensure the continued appeal of its products to developers and customers, Microsoft is committed to designing its high-volume products in a manner that ensures open connections to these products, supports key industry standards, and enables customers to transfer their data for use by others.

These commitments are at the root of, and are reflected in, Microsoft's publicly announced Interoperability Principles.¹

By way of general background: Microsoft collaborates with hundreds of industry participants to improve interoperability for its products through a variety of constructive engagement models. As one example, Microsoft helped form the Interop Vendor Alliance ("IVA") in November 2006 to enhance the way that diverse products interoperate with Microsoft's operating systems and applications. Major industry players belong to the IVA, including BEA Systems, NEC, Novell, Siemens, Sun, and nearly 60 other vendors.² Microsoft is also a founder and major supporter of the Document Interoperability Initiative ("DII"), a series of forums around the world where vendors address interoperability issues between different document format implementations and collaborate on the development of tools to enhance interoperability over time.³ Microsoft is also a major supporter of the development of open source tools that translate between different formats, such as: ODF to Open XML, binary formats to Open XML, and UOF to Open XML. Separately, Microsoft and Novell formed a precedent-setting collaboration that responds to the needs of customers who want to run both Windows and Linux. In particular, that agreement includes cooperation on document format interoperability. In yet another recent example, Microsoft agreed to license its ActiveSync protocol technology to Nokia for use in connecting its Exchange email server to Nokia's mobile phone software. Microsoft also licensed the ActiveSync protocol technology to Google for use in Google's servers (running software that competes with Microsoft's) that provide hosted services to synchronize calendar and contact information with mobile phone software (including non-Microsoft software) that also uses the licensed technology. These are just a few of the many active interoperability collaborations in which Microsoft has engaged recently.

"Interoperability by design" is another important aspect of Microsoft's interoperability approach and a key component of its product development efforts, including Office development. Microsoft designs its products to enhance interoperability across diverse systems, benefiting consumers and the industry at large. The Windows Principles released in July 2006 memorialize that commitment for Windows.⁴

Office 2007 incorporates the same design philosophy. For example, Microsoft developed and supported the standardization of a new file format, which is used by default in the core Microsoft Office 2007 applications (i.e., Microsoft Word, Microsoft PowerPoint, and Microsoft Excel), called Office Open XML ("Open XML"). Open XML is now an approved standard in both Ecma International and the International Organization for Standardization and International Electrotechnical Commission (ISO/IEC), enabling any vendor to use it in their own productivity

¹ See <http://www.microsoft.com/interop/principles/default.aspx>

² See <http://interopvendoralliance.org>.

³ See <http://documentinteropinitiative.org>.

⁴ See <http://www.microsoft.com/presspass/newsroom/winxp/windowsprinciples.aspx>.

and other applications. Microsoft has made any patents it may have that are necessary to implement all or part of Open XML available on a royalty-free basis.⁵ Indeed, Novell and Corel announced plans to support Open XML in their products, which will deliver format interoperability on both the Windows and Linux platforms.

Importantly, Office 2007 carries forward many other traditional Office features that enable support for a wide variety of alternative document formats. The next update to Office 2007, known as Service Pack 2 (“SP2”) and expected to release in the first half of 2009, will include built-in support for an OASIS-approved document format standard, OpenDocument Format (“ODF”) v. 1.1, XPS and PDF.

The history of file formats in Microsoft Office is one example of Microsoft’s commitment to interoperability. The level of interoperability that Microsoft has achieved in file formats is extensive, and it is the result of years of development in response to evolving customer preferences. The following background regarding file formats should be useful in the DCA’s inquiry. After reviewing the history of file format interoperability, this paper discusses challenges faced by all implementers of standardized file formats and lays out Microsoft’s holistic approach to addressing those challenges and enabling maximum interoperability for users.

II. HISTORY OF FILE FORMAT TECHNOLOGIES IN MICROSOFT OFFICE

A file format is basically a way to encode information for storage in a computer file. Common examples of file formats include those for pictures (JPEG, TIFF), media (MP3, Windows Media, Quick Time), fixed-format documents (PDF and XPS), and unicode text. For productivity applications, file formats are a way to store the attributes of a document prepared using the features of the application in a persistent state and to reproduce them when the file is re-opened.

File formats and the feature set of the applications that use them are mutually dependent. If an application does not implement a feature capable of using information stored in a format, the storage of the information itself will not provide that feature. As an example, you can use WordPad, a simple word processing feature in Windows XP, to open and read documents created in Microsoft Word 2003, but you cannot see attributes of features that WordPad does not include (*e.g.*, “blackline” revisions created by Word’s “Track Changes” feature). Conversely, rich document creation features available in an application are nearly meaningless if the file format is unable to store the data associated with that feature.

⁵ See <http://www.microsoft.com/interop/osp/default.mspx>.

As computing power and office application software have evolved, so too have file formats. The implications for interoperability are discussed below.

A. Binary File Formats

File formats can store data as a binary, meaning code made up of zeros and ones that can be read by machines but not by humans. Use of binary file formats was and is common in software development.

The earliest versions of the applications in Microsoft Office stored document information in proprietary binary file formats, as did most leading software products at that time. However, Microsoft has long recognized the need to be able to exchange text and graphics across applications and operating system platforms, and created a format called Rich Text Format (“RTF”) in the late 1980s that enabled easy transfer of documents between MS-DOS, Windows, OS/2, and Apple Macintosh applications. RTF is well documented and described for developers through specifications published on the Internet and through Microsoft’s developer site, MSDN. It was and continues to be widely implemented in applications that run on Windows, Apple OS X, and Linux.

As the functionality of Office improved dramatically over the years, the format technologies used to store document data became more complex. The Office binary file formats currently in use, including .doc (Word), .xls (Excel), and .ppt (PowerPoint), were designed in 1994 and first released as the default file format in Microsoft Office 97. When Microsoft created the Office binary file formats, the primary goals were efficient usage of disk space and quick processing in machines that were much less powerful than those available today.⁶ The Office binary file formats were very efficient and suitable for storing data in the computing environment of the day.

In general, in the 1990s, few application developers disclosed information about file formats. Binary file formats created during this era were complex and were designed to support advanced application features with optimum performance. Interoperability was not a key design objective in an era where computing power was more limited and networked document collaboration was less relevant. Documents saved in these formats were commonly created and used on a single machine, with a single productivity application. Manipulation of the binary formats by third-party application developers was a common cause of file corruption — which was incorrectly attributed to Microsoft and its Office software and resulted in support calls from customers experiencing problems. Microsoft, like the handful of other leading vendors of

⁶ Because the binary formats persist only in binary form, they take up less disk space than human readable source code files would. In addition, since the files are already in binary form, they do not need first to be compiled in order to be able to be read by the appropriate application. This cuts out an important step and results in improved performance, particularly on older, slower machines.

productivity software, addressed compatibility between software products through a combination of engineering efforts, licensing deals, and occasional disclosures.

It was not uncommon for companies to study competitors' file formats and design their own applications to enable conversion from those file formats for use in their own applications. In fact, Microsoft's support in Office for formats used by leading applications like Lotus 1-2-3 and WordPerfect was developed using techniques other than "public" file format documentation.

Even so, Microsoft has disclosed and licensed specifications for its binary file formats to hundreds of licensees since 1995. The extent to which these disclosures have been "public" has varied, but third-party developers – including major Microsoft competitors such as IBM, Adobe, Oracle and Corel – have routinely been provided with technologies and specifications that facilitate interoperability. Microsoft has accomplished this through programmatic binary file format offerings and by offering additional interoperability technologies or license rights related to its binary file formats to vendors, when requested. For example, Microsoft and Corel cross-licensed file format technologies in 1995. Lotus was provided with file format documentation in connection with Office 2000. And Microsoft has provided IBM with binary file format documentation as well as direct assistance in working with and supplementing this documentation. These and other licensing efforts are described below.

Microsoft's binary file format documentation was periodically available publicly, sometimes subject to limitations on use and sometimes not. For example, beginning in October 1995, Microsoft documented the Excel binary file formats and made this documentation available with no licensing restrictions via the Microsoft Developer Network ("MSDN") and through MS Press, a group at Microsoft that publishes books and materials on various topics including software development and design.

By 2003, however, Microsoft released a new, programmatic Office Binary File Format Agreement for governments and for software developers building products that would complement the functionality provided by Office. And by March 2006, Microsoft removed the complementary use limitation from its file format agreements. In April and May 2006, IBM, Adobe, and Corel (among others) signed that new agreement.

Based on a desire to make these specifications broadly available to anyone, the binary file formats used in Microsoft Word (.doc), Microsoft Excel (.xls) and Microsoft PowerPoint (.ppt) are available on Microsoft's website at: <http://www.microsoft.com/interop/docs/officebinaryformats.mspx>. They initially posted on February 15, 2008. Since that time, there have been more than 281,600 downloads of these specifications. The specifications are covered by the Microsoft Open Specification Promise, which means that both open source and commercial developers are free to use the specifications without the need to obtain a patent license from Microsoft. On June 30, 2008, Microsoft

released revised documentation for the Microsoft Office binary file formats for Word, Excel and PowerPoint (.doc, .xls, .xlsb and .ppt), available at <http://msdn.microsoft.com/en-us/library/cc313118.aspx>. This revised documentation maps more closely to the documentation format Microsoft uses for other specifications and addressed feedback we received from the community. There have been more than 26,600 downloads of these documents.

B. HyperText Markup Language (“HTML”) and Extensible Markup Language (“XML”)

At the same time in the mid- to late-1990s that Microsoft was releasing Office versions that used the binary file formats as the default, the computer industry experienced a significant shift with the increasing importance of the Internet and the associated explosion of technologies to support communication and collaboration among office workers. As products like Office adapted to new uses and consumer needs, Microsoft recognized the advantages of common standards for storing and presenting information. In a now-famous 1995 internal memo describing the “Internet Tidal Wave,” Bill Gates observed the power and flexibility of HyperText Markup Language (“HTML”) as a standard defining how information on the Internet — including text — would be presented and encouraged the Office team to build support for HTML directly into Office.⁷

Microsoft rapidly began to build this support into its applications. For example, Office 2000 offered users the option to save documents natively in HTML. This effort was soon supplemented by support for a more-powerful standard for creating interoperable documents, the Extensible Markup Language (“XML”).

XML is a widely accepted industry standard that was originally defined in 1996 by a working group of members that included representatives from Sun, Netscape, Microsoft, HP, Adobe, and others. It enables exchange of data between disparate systems by providing rules for defining the structure, content, and attributes of documents in a way that can be understood by any software familiar with XML. These rules include “tags” and “schema.”

XML tags define a document’s structural elements, such as the types of data it includes, and the meaning of those elements. XML tags can be used to define how information in a document looks, what it means, and how it relates to other information stored in the document. For example, a customer invoice might include tags associated with a customer’s address. XML tags could identify particular text, explain what it means, and define how it should appear in the document. Thus, `<city>Copenhagen</city>` would let other systems know that the text in

⁷ Memo dated 26 May 1995. Mr. Gates noted that Lotus was already shipping a product (the Internotes Web Publisher) which replicated Notes databases into HTML and observed that for Office customers, “[a]llowing for collaboration across the Internet and allowing people to publish in our file formats for both Mac and Windows with free readers is very important. ... Word could lose out to focused Internet tools if it doesn’t become faster and more WYSIWYG for HTML.”

question is a city, `<h2>City</h2>` would describe “City” as a second-level heading, and `City` could indicate that output devices should render “City” in bold text.

XML schema define the tags that can be used for a particular document type or information type, and the rules for applying them. Custom schema can be created by a company or industry to define information relevant to their business. For example, financial tags such as `<Amount>` or `<ROI>` will be used in a file format for financial documents. Custom schema can be created by software vendors to define the contents, behavior, and appearance of documents created in their application, *i.e.* to describe the file format.

In Office 2003, Microsoft implemented new XML-based file formats and made the schema for these new formats available royalty-free to other developers and customers for use with non-Microsoft computing environments.

C. Open XML

More recently, Microsoft created XML-based file formats called Office Open XML for use in Microsoft Office applications. The Danish Ministry of Science and Technology expressed a strong interest in using open file formats, and discussions with them in 2003 resulted in Microsoft making available a royalty-free license for the Open XML file formats. In May 2004, the European Commission’s agency for Interoperable Delivery of European eGovernment Services to public Administrations, Business and Citizens (IDABC) encouraged vendors to proceed with file format standardization activities through standards bodies. Microsoft’s Open XML format was subsequently standardized first by Ecma International, a well respected industry standardization body with nearly 50 years of operation, and then by ISO/IEC. In Ecma International, representatives from Apple Computer, Barclays Capital, The British Library, Essilor, Intel, Microsoft, NextPage, Novell, Statoil, Toshiba, and the U.S. Library of Congress worked together to help ensure that the Office Open XML Formats are highly useful and interoperable in a wide variety of scenarios. A record 87 national bodies participated in the standardization of Open XML in ISO/IEC with overwhelming support for the standard at the end of the process.

The Ecma Open XML standard is called “Ecma 376.” The ISO/IEC Open XML standard, which includes changes to Ecma 376 agreed to during the ISO/IEC standardization process, is called “IS29500”. The maintenance and evolution of Open XML now rests in the hands of the global community through their participation in the ISO/IEC process that has been set up for this purpose.

Ecma 376, the default file format in Word, Excel and PowerPoint 2007, is documented, and is available royalty-free for implementation by any application on any platform. The same is

true for IS29500, which will be supported in the next version of Office.⁸ This allows any software vendor to include support in their software for the default file format in Office and numerous developers are doing so, including:

- Linspire;
- Novell (OpenOffice);
- Apple;
- Neo Office;
- Adobe;
- IBM/Lotus;
- Thinkfree;
- Quickoffice;
- Sun;
- Corel; and
- PythonOffice – Open Source

See <http://www.openxmlcommunity.com/applications.aspx> (providing additional detail on vendors that support Open XML in their applications – or who have announced plans to do so). Applications supporting Open XML are available on multiple operating systems, including Linux, Mac OS and Windows.

Microsoft has also provided documentation and licensing for technologies that can be used by developers to create software utilities that convert Office formats, such as Office binary file formats, to other file formats for which a developer has designed support. For example, Sun has developed software that converts binary file formats to ODF. And Microsoft itself has created and licensed a utility that converts Office binary files to the standardized Open XML format, and vice versa. This utility, the Compatibility Pack, is available royalty-free, does not require an Office license, and has been downloaded more than 70 million times.

As part of its continued commitment to interoperability, Microsoft is funding an open source project, “Office Binary (doc, xls, ppt) Translator to Open XML,” to create software tools and guidance that will show how a document written using the Office binary file formats (doc, xls, ppt) can be translated into the Open XML format. Customers will be able to use the translator to migrate from the binary file formats to Open XML-based documents. The translator is developed as an open source project in SourceForge (<http://b2xtranslator.sourceforge.net/index.html>) under the open source Berkeley Software Distribution (BSD) license, which allows that anyone can use the mapping and code, submit bugs and feedback, or contribute to the project. Microsoft supports the development of similar

⁸ These materials will also be subject to Microsoft’s “Open Specification Promise,” which provides significant protections from potential IP claims. *See* <http://www.microsoft.com/interop/osp/default.mspx>

translators on SourceForge that focus on ODF-Open XML translation and UOF-Open XML translation.⁹

In Office 2007, the Open XML-based file formats have become the default format for files created in Word, Excel, and PowerPoint 2007. The formats are fully documented in a specification available to third parties. The formats were originally created by Microsoft and revised by Ecma as it developed Ecma 376 and then by ISO/IEC as it developed IS29500.¹⁰ Open XML is platform-independent and it is not tied to any programming language, operating system, or software vendor. Thus, any developer can use Open XML in its products to save files in a format that can be read by Office and to read and edit files that were saved in Office. Microsoft has made the patent rights necessary to implement all or part of Open XML available royalty-free to any implementer anywhere in the world. As discussed later in this paper, Microsoft has also created Implementers' Notes to assist other applications in maximizing interoperability.

Finally, for customers and vendors who choose to develop and/or support some format other than those shipped in Office, Microsoft has built an application programming interface (API) into Office 2007 Service Pack 2 and future versions. The API enables additional document formats to appear in the drop down menu of available formats and be selected by users as their default format should they choose to do so. Over time and with each successive release, Microsoft has built on the file format support provided in Office to improve the performance and interoperability of documents used with its applications. Today, users can choose between more than a dozen file formats (including many based on common standards) or add other formats of their choice.

a. File Formats in Microsoft Office Today

The history of file format technologies used in Office is a story of consistent improvements in format choice and interoperability. As a result of this steady march, in Office 2007:

- Users save documents by default in a file format, Open XML, that is documented, is an Ecma International standard, and is available royalty-free for implementation by any application on any platform. This allows any software vendor to build support into its applications for the default file format in Office. (In the next major release of the Office system, codenamed "Office 14," Microsoft will update the already substantial support it provides for IS29500, the Open XML specification that was recently approved by ISO/IEC members.)

⁹ UOF is a national standard document format in China.

¹⁰ For more details, see *Introducing the Office (2007) Open XML File formats* available at <http://msdn2.microsoft.com/en-us/library/ms406049.aspx>.

- Microsoft Word, Excel and PowerPoint 2007 already provide support for more than a dozen different document formats. These include cross-platform formats that allow documents created in Office to be read, saved, and (with some formats) edited, on platforms and devices that have no Microsoft productivity applications or no productivity application at all. These formats include implementations of widely available technologies like RTF, HTML, XML, XPS, and PDF. (For the last two, a code download is currently needed, but this will be changed with the release of Office 2007 SP2 in the first half of 2009).
- Users can choose to open, edit, and save documents in other formats that are provided by third parties, through an architecture built into Office 2007 Service Pack 2 (and later versions) that allows developers to include their formats in the drop down menu of available formats for Office users and, if they want to do so, set that format as their default.
- Users have the option of using ODF-Open XML, UOF-Open XML and binary-Open XML translators that have been developed and are continuing to evolve as open source projects on SourceForge that involve the collaboration of Novell, Microsoft and others.

In May 2009, Microsoft will further expand the range of document formats supported in its flagship Office productivity suite, by offering “out of the box” support for XML Paper Specification (XPS), Portable Document Format (PDF) 1.5, PDF/A and Open Document Format (ODF) v1.1. These major changes will be made with the release of Microsoft Office 2007 Service Pack 2 (SP2) in the first half of 2009.

Microsoft’s support for ODF v1.1 will enable users to read and write ODF v1.1-compliant documents and improve interoperability with products such as OpenOffice, Sun’s Star Office, KOffice, and IBM’s Symphony. Microsoft adhered to the following set of principles to guide it through the complex task of implementing ODF:

- **Implement the most widely implemented version of ODF (OASIS ODF 1.1).** While this is not the ISO/IEC standard version (that is ODF 1.0, IS 26300), it is the version used by the most popular products that implement ODF.
- **Adhere strictly to the standard.** This maximizes interoperability with other compliant implementations.
- **Do not extend the standard.** Microsoft did not extend ODF to support Office features that go beyond the current ODF specification because that might have created confusion in the market, forced other implementers to undertake additional work to match Microsoft’s implementation, and raised claims that Microsoft was “embracing and extending” ODF to take control of it or thwart its adoption.

- **Support as much of the ODF standard as we could.** Generally, Microsoft enabled use of an Office feature in its implementation of ODF if ODF supported it. Microsoft Office has many features that simply are not supported by the ODF standard (since that standard was based on the OpenOffice product and file format).

These principles guided our technical teams as they worked through many difficult issues and documented their efforts in the ODF Implementers' Notes. This approach was well received by the community during various interoperability testing and "plug fest" events that Microsoft organized (known as the Document Interoperability Initiative events, described further in Section IV.C below) because the community recognizes that these are difficult issues for everyone implementing standards, with no easy answers.

By enhancing the support for other file formats in Office and participating in collaborative efforts to improve interoperability, Microsoft is offering customers ever-greater choice and more flexibility among document formats and creating additional opportunities for developers and competitors. Specifically:

- When using Office 2007 as updated by SP2, customers will be able to open, edit and save documents using ODF 1.1 and save documents into the XPS and PDF fixed formats from directly within the application without having to install any other code. Office 2007 with SP2 will also allow customers to set ODF as the default file format. In order to also provide ODF support for users of earlier versions of Microsoft Office including Office 2000, Office XP and Office 2003, Microsoft is funding an open source community project and collaborating with software vendors in the development of the Open XML-ODF translator project on SourceForge.net.¹¹ As a testament to the usefulness of this open source translator, we note that Novell is redistributing it to OpenOffice.org / SuSe Linux customers.¹²
- In addition, Microsoft has defined a roadmap for its implementation of the recently ratified International Standard ISO/IEC 29500 (Open XML). IS29500, which was approved by ISO/IEC in March 2008, is already substantially supported in Office 2007. The company plans to update that support in Office 14.
- These steps are in addition to Microsoft's previous commitment to provide new APIs in 2007 Office (in Service Pack 2) that will allow developers to plug in other formats so they show up on various Office drop down menus and can be selected as the default formats by users.
- Microsoft joined the Organization for the Advancement of Structured Information Standards (OASIS) technical committee working on the next version of ODF and will

¹¹ See <http://odf-converter.sourceforge.net/>

¹² See <http://download.novell.com/Download?buildid=OabXVm-plcA~>

take part in the ISO/IEC working group when it is formed to work on ODF maintenance. Microsoft has submitted 15 proposals so far for improvements to ODF as OASIS develops the 1.2 version of the standard. Microsoft employees are also participating in the ISO/IEC working group that was formed to maintain Open XML (JTC1/SC34 WG4) and the ISO/IEC working group that was formed to improve interoperability between ISO/IEC-recognized document formats (JTC1/SC34 WG5). Microsoft employees participate in these working groups through their membership in various national bodies worldwide, as well as their membership in Ecma TC45.

- The company will also be an active participant in the ongoing standardization and maintenance activities for XPS and PDF and will continue to work with the IT community to promote interoperability between document file formats, including Open XML and ODF, as well as Digital Accessible Information SYstem (Daisy XML), the foundation of the globally accepted DAISY standard for reading and publishing navigable multimedia content.

In short, Microsoft is committed to providing Office users with greater choice among document formats and enhanced interoperability between those formats and the applications that implement them.

III. INTEROPERABILITY CHALLENGES OF STANDARDIZED FORMATS

Implementers of any standardized file format face a number of interoperability challenges. It is difficult to implement standards in a way that enables products from a range of vendors to interoperate without any issues. Because developers in different companies will not make identical decisions, a single ODF document may not look the same even in different ODF implementations of the same version of the *same* ODF standard. A similar situation can occur with different implementations of Open XML.

To illustrate the challenges all vendors face in designing interoperable products, Microsoft has included a series of screenshots in Appendix A. These screenshots show how, as of June 2008, three open source products, OpenOffice, KOffice, and IBM Lotus Symphony, *all of which were based on the ODF standard*, displayed the same document in very different ways. Each of the screenshots describes a scenario and explains the reason why the document view appears different across these programs.

A more recent blog posting by Doug Mahugh¹³ explores the “hidden text” feature of two ODF implementations. It demonstrates how hidden text in a document created with Sun’s

¹³ The blog post may be found at: <http://blogs.msdn.com/dmahugh/archive/2009/01/14/interoperability-challenges.aspx> .

OpenOffice may not be hidden when the document is saved and opened with IBM's Symphony. Likewise, it shows how a scatter plot may be substantially distorted when saved and opened in a different implementation of the same standard.

The distortions caused even by various open source implementations of the ODF standard illustrate the challenges faced by all implementers. The idea of perfect round-trip fidelity among implementations must be viewed in this perspective. When one speaks of interoperability of implementations of a standardized document format, or between implementations of different standardized formats, it is important to recognize that interoperability is a spectrum. For products requiring large complex standards, such office software with ODF and Open XML, interoperability can be "weak" or "strong," but is never "perfect" between different vendors' products.

There are a number of valid reasons why this happens.

- **The standard has gaps that each implementer must fill or choices that each implementer must make.** Clear examples of this point are the lack of formulae in the ODF spreadsheet specification and the fact that ranges can be established in a variety of different ways in Open XML. For example, the ECMA-376 specification states that tables can have any number of columns, but Word only supports up to 63 columns in a table. Each developer must, therefore, devise its own method of addressing these gaps or making these choices. Transparency is important because each implementer needs to understand the approach that other implementers have taken.
- **Some aspects of a standard are vague.** In many places, ODF is not sufficiently detailed for developers to choose the same implementation path. For example, ODF does not specify how data in a chart should be laid out. The Open XML standard similarly does not specify a maximum length for a header caption in a pivot table.
- **Existing applications do not conform to the standard.** Some existing applications that support ODF, including OpenOffice, deviate from the standard, either through inadvertent bugs or intentional product design choices. The same is true for applications that support Open XML. For example, the ECMA-376 specification states that charts embedded in documents should be auto-refreshed when the document is opened, but Excel does not refresh charts until the underlying data is changed. This is not intended to criticize any implementation, but rather to point out a reality of the marketplace.
- **Developers implement different parts of the specification.** Developers incorporate features into their applications based on market demands. Because various developers do not serve identical markets or view the same market in the same way, their products will not be the same. That is clearly the case when various implementations of ODF are compared. And again, the same is true for Open XML implementations. For example,

many applications in the market today do not implement all the optional elements of either ODF or Open XML.

- **A specification may not be designed to support all of the functionality that users demand.** ODF designers decided not to include the entire range of features demanded by users today. For example, the encryption algorithm required by ODF does not meet the requirements of many governments. In many cases, therefore, products that implement a standard will go beyond that standard to provide the functionality that they believe users demand in the marketplace.
- **Products and standards continuously evolve.** That is evident with both ODF and Open XML. ODF 1.0 was adopted by OASIS on May 1, 2005 and then became ISO/IEC 26300 on May 3, 2006. Subsequently, OASIS adopted ODF v1.1 on February 1, 2007, but this version has not been submitted to ISO/IEC. OASIS is currently working on ODF v1.2, and Microsoft employees are actively participating in that process. Open XML was initially standardized by Ecma, which then submitted it to ISO/IEC for further standardization. The Open XML maintenance process began in late January 2009 at ISO/IEC, and Microsoft employees have been active contributors, working through their national bodies and SC34 liaison organizations such as Ecma TC45. The evolution of these standards and the applications that implement them evolve asynchronously, creating gaps between versions of the standard and products in the marketplace.

In Microsoft's view, the fact that implementations of the same document format standard may not result in perfect interoperability is not an interoperability "problem." Rather, it is simply the result of the practical challenges presented by document format interoperability. The best way for an implementer to address these challenges and improve interoperability is through participation, transparency, and collaboration, as described in Section IV, below.

The last slide of Appendix A shows an example of how, using these means, open source office software applications have developed highly interoperable support for Microsoft's binary file formats. They have been able to do this because Microsoft is engaged and committed to efforts that ensure a high degree of interoperability across the office software from different vendors. That is demonstrated by its documentation of the binary formats, the standardization of Open XML, Microsoft's support in Office for ODF 1.1, its transparency in sharing information regarding its specific implementations of both formats, and its collaborative work side-by-side with other implementers to test user scenarios and address issues that are identified. The next section discusses in more detail Microsoft's approach to facilitating document format interoperability.

IV. MICROSOFT'S APPROACH TO INTEROPERABILITY IN STANDARDIZED FORMATS

Having reviewed the general history of file format and Microsoft's historic approach to interoperability and licensing, it is worth discussing where things stand today.

Microsoft recognizes that its customers want to be able to choose document formats that best meet their needs and to have interoperability between implementations of a single format and between implementations of different formats. Microsoft believes that the way forward on document interoperability involves three steps: (1) participating actively in maintaining and evolving the standard, (2) being transparent about how the standard is implemented, and (3) collaborating with other implementers and end users to address new challenges and needs as they arise. Applying this approach increases interoperability and preserves customer choice. Microsoft has encouraged others in the industry to adopt these same practices.

A. Participation

Participating and contributing positively to the maintenance and evolution of a standard is critical if it is to remain useful for vendors and customers alike. Every standard needs to evolve over time to correct errors, fill gaps, and include new technologies. Microsoft is committed to being a positive part of that process, as demonstrated by our high level of activity in maintaining ODF in OASIS and Open XML in ISO/IEC now that ISO/IEC IS 29500 has been issued. We are directly involved in the process in OASIS to develop the next version of the ODF standard, which includes making a number of positive technical contributions through our representatives on the technical committee, and to participate in the working group focused on interoperability between ODF implementations, which has only recently begun its work. We are actively working in ISO/IEC to maintain Open XML and to develop further the standard, which likewise includes working with the Ecma technical committee that edits the standard and submitting positive technical contributions to ISO/IEC JTC1/SC34 WG4 through the appropriate national bodies and liaison organizations. And we are participating in JTC1/SC34 WG5, which is focused on improving interoperability between different ISO/IEC-recognized document format standards.

B. Transparency

Being open and transparent about how file formats are implemented can greatly increase interoperability. On December 16, 2008, Microsoft published details about its implementation of ODF v1.1 support in the soon-to-be-released Microsoft Office 2007 Service Pack 2. The details were published in the form of Implementers' Notes, which Microsoft will update as future versions of Office are released. Likewise, on January 16, 2009, Microsoft published Implementers' Notes for its implementation of Open XML (ECMA-376 Edition 1). This documentation is publicly available to all and free of charge on the Document Interoperability

Initiative website, <http://www.documentinteropininitiative.org>. Microsoft's implementation notes include the following:

- **Details on implementation decisions.** When implementing a standard, an implementer may find the text ambiguous or more permissive than is appropriate for that particular implementation. In these cases implementers need to make a choice that best suits their application. This type of information enables developers to see the direction a vendor is taking and make informed decisions about their own efforts to interoperate. Therefore, the Notes describe in detail how Microsoft implements elements of the standard. For example, the Notes describe specific range maximums in our products and explain how we support “black and white” mode and graphic frames.
- **Details on additional data written into files.** File format standards typically allow additional application-specific information (such as certain user customizations) to be written to the file. By providing this information vendors allow developers to correctly interpret the additional data.
- **Details on implementation variances.** With every application there may be instances where an implementer cannot follow the standard exactly for one reason or another. For example, general industry practice may differ from what is in the specification or users may have made clear that they need something different. In such cases, it is important for vendors to document their approach so other vendors can make fully informed decisions about how they will approach implementation.
- **General Additional Information.** Some of the Notes simply provide additional information that we believe would help interoperability by being fully transparent as to how we work with Open XML. For example, the Notes explain how we apply bold fonts in circumstances not addressed by the standard and describe how we write Office 2007-specific data (such as information about ribbon User Interface customization) into Open XML files outside the scope of the standard.
- **Proposed Corrections and Updates to the Standard.** The Notes also include information that will help ISO/IEC maintain the standard over time. For example, the Notes identify typographical errors and other apparent editorial mistakes. Microsoft worked within the Ecma technical committee responsible for maintaining Ecma 376 to submit a number of these corrections to ISO / IEC SC34 WG4 in January and February.

The Notes Microsoft published on January 16, 2009, describe Microsoft's implementation of the Ecma International ECMA-376 open standard in Office 2007. The next version of Office – “Office 14” – will support the subsequently standardized ISO/IEC IS 29500 open standard. We will update the Implementers' Notes to address IS 29500 after we release Office 14. We do not expect many changes, however, because the standards are generally quite similar.

In that way, Microsoft's voluntary openness and transparency provides all parties with detailed information they can use to make decisions that might affect the interoperability of their products with Microsoft products. These Implementers' Notes have been highly praised by the developer and standards communities. Microsoft believes that all file format implementers should share the details of their implementations so that others can take that information into account when determining how best to proceed with their own work.

C. Collaboration

The goal of constantly improving interoperability requires that vendors collaborate to preserve customer choice, meet new challenges to interoperability, and develop tools to assess and improve interoperability over time. We are leading efforts with other stakeholders to do just that with respect to file format implementations.

Towards this end, Microsoft has undertaken a major Document Interoperability Initiative, a global program of technical vendor discussions, labs and solution enablement programs to develop solutions for effective data exchange between product implementations of file format standards. This initiative is open to any vendor who wants to collaborate with the community to identify and address interoperability issues between different implementations of file formats. They are encouraged to propose meetings, raise issues, and help set the agenda for the effort. Users of information technologies are also encouraged to participate to help prioritize and focus the efforts of vendors. The DII is community-based and open to all participants wanting to promote file format interoperability for the benefit of the user community. The DII's goals include:

- Testing interoperability between product implementations of file format standards in some common scenarios including enterprise server (back office, unstructured data, schema across supply chains) and data authoring, presentation, and exchange across a range of devices.
- Identifying systematic gaps that can be addressed via industry collaboration.
- Identifying issues with specific implementations or between implementations.
- Discussing how we as a community can work together improve the overall quality of file format implementations and improve interoperability.

In 2008, eight DII events were held in Brussels, Munich, Seoul, Beijing, Tokyo, and the United States. These events have been opportunities to surface and address interoperability-related issues between implementations, showcase tools and solutions from various vendors and projects, and begin to build tools to help create promote consistent implementations of Open XML. The events have been well received (see www.documentinteropinitiative.org) and will continue around the world. Additional events are planned for 2009.

As part of our commitment to working with a variety of stakeholders, Microsoft recently launched a collaborative effort with the Fraunhofer Society.¹⁴ Fraunhofer is a German research organization with dozens of institutes distributed throughout the country. It employs over 10,000 engineers and scientists and has a budget of over €1 billion, the majority of which comes from government.

Microsoft has provided initial financial support to Fraunhofer to kick off a new project to enhance document format interoperability between Open XML implementations. The Fraunhofer project has two elements. The first is the development of a document format test library that can be used by any Open XML implementer to test the quality of its implementation in a wide range of important user scenarios. The DII participants have identified the creation and use of a test library of this type as the most effective step the community of implementers can take to improve interoperability that provides real benefits to users. The second element of the project is the development of a “schema validator” that can be used by implementers to test whether documents produced by their software structurally conform to the requirements of the standard. The project is open to all implementers to contribute to and use and the fact that it is being led by the highly respected and independent Fraunhofer Society will ensure that it will rapidly advance the goal of interoperability between Open XML implementations, increasing choice and value in the marketplace.

V. CONCLUSION

As discussed above, Microsoft is committed to enabling a high degree of interoperability with its products and document file formats. Through its Interoperability Principles, its support for standardized file formats, its constructive engagement in standards organizations, and its hosting of collaborative DII events with the vendors of competing productivity applications, Microsoft has undertaken far-reaching efforts that have resulted in ever-increasing interoperability for the benefit of customers and users around the world.

¹⁴ Additional information about this joint initiative is available here:
http://www.fokus.fraunhofer.de/en/elan/projekte/international/laufende_projekte/Document-Interop_Lab/index.html

APPENDIX A:

Please see document entitled “Appendix A – Interop Testing Screens.pdf”.