

WHITE PAPER

Adoption of Document Standards

Sponsored by: Microsoft

Melissa Webster

Per Andersen

Thom Rubel

August 2007

IDC OPINION

An IDC survey of 200 organizations in the United States and in Europe shows that the use of XML-based document standards is increasing. Key findings are as follows:

- ☒ Organizations are looking to implement document standards to facilitate the exchange of documents externally and internally as well as increased interoperability with back-end applications. Also, the need to minimize the risk of long-term document archiving plays an important role.
- ☒ The uptake of XML-based standards seems stronger in Europe than in the United States, but in both geographies, the dominant XML standard deployed is Office Open XML. Open Document Format (ODF) is receiving some attention in the public sector but is not as widely used as Office Open XML even here.
- ☒ Organizations do not put emphasis on discussions about the "openness" of standards. Instead, more practical aspects are rated highly: Cost is very important as is the ability to have an easy transition of existing documents to a new standard. This is particularly true for large organizations and organizations in the public sector.
- ☒ The same pragmatic attitude is found in responses to the question about whether organizations aim for a single document standard or multiple document standards. IT managers favor managing just one standard, but line-of-business (LOB) managers generally see the need for multiple standards. Therefore, when LOB managers are looking at an XML standard such as Office Open XML, they see it as one of several standards deployed in the organization.
- ☒ In contrast to our initial thinking, organizations do not see major barriers to implementing document standards. Overall, the decision to move to a document standard is seen as of average complexity by most organizations and the cost factor is seen as a barrier for just a minority of organizations.

METHODOLOGY

To understand the document formats that organizations are using today and their strategies for future document standards, IDC conducted a survey. We interviewed 200 organizations — 100 in the United States and 100 in Europe.

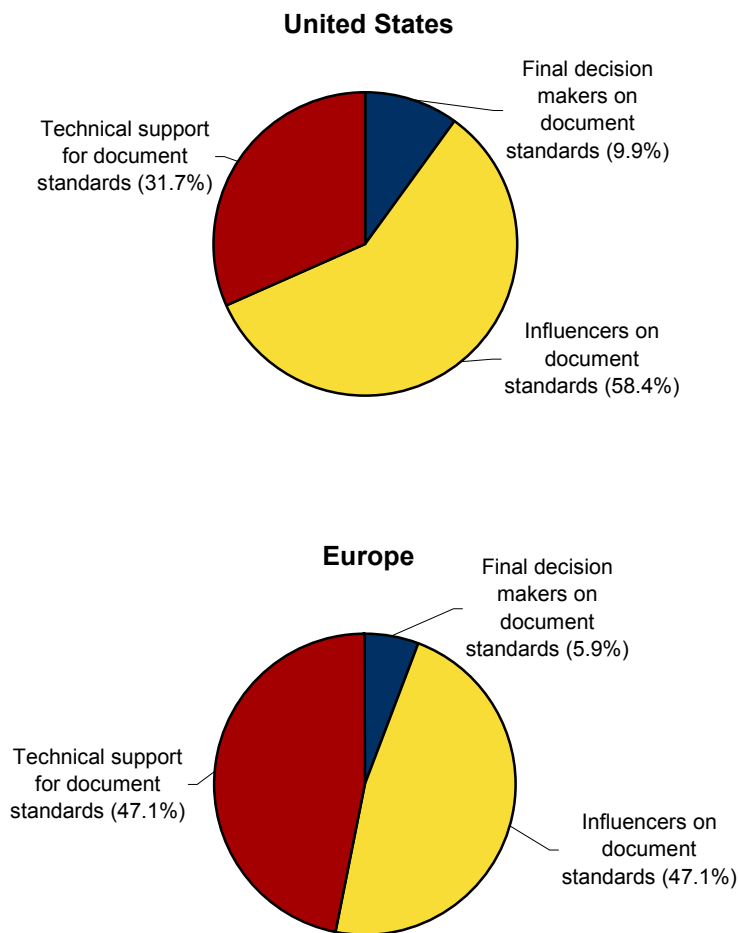
In each region, 50 organizations with more than 250 employees were selected from the public sector and 50 organizations with more than 250 employees were selected from the commercial sector.

Most of the respondents interviewed have a key influence on document standards or recommend document standards in their organizations (see Figure 1). The majority of the remaining respondents are staff providing technical support for the document standards in use.

FIGURE 1

Respondent's Role in Use of Document Standards

Q Which of the following best describes your role in your organization's use of document standards?



n = 200

Note: Data is weighted by organization's primary business.

Source: IDC, 2007

SITUATION OVERVIEW

As PCs and personal productivity tools emerged in the 1980s, enterprises and organizations were faced with significant barriers as they tried to exchange documents. These barriers existed when one person needed to share a document with another person in another company using a different system. But problems existed even internally. Because of the decentralized acquisitions of PCs, companies often ended up with different systems in different parts of their organizations.

In the early days of PCs and productivity tools, three barriers made life very difficult for PC users: The physical disk size was different from system to system (typically, 8in., 5.25in., and 3.5in.). Even if the disk size was the same, the disk format could be different (number of tracks, sectors, and sides). Even if the disk format was the same, different productivity tools used different file formats.

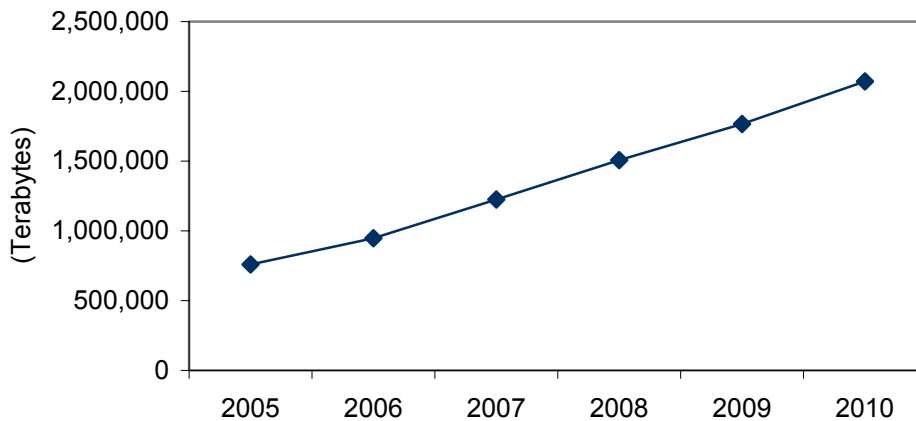
It is no wonder, then, that an industry emerged to help PC users convert a disk with a document from one system to a document readable on another system — although the process was rarely flawless.

Today, although we hardly think about it, most PC systems use compatible disk and ROM formats — and most documents in any case are shared through the use of a network. Thus, PC users and organizations today are faced with "just" the challenge of using different productivity tools and systems with different file formats.

The number of documents created or captured on PCs continues to increase, even decades after the installation of the first PC. IDC analysis shows that the number of PC documents created or captured will increase from about 1 million terabytes today to more than double this amount in 2010 (see Figure 2).

FIGURE 2

Worldwide Creation/Capture of PC Documents by Terabytes, 2005–2010



Source: IDC, 2007

No one can doubt that documents from PC productivity tools play an increasingly important role in communication today. Emails and document attachments have taken over the roles of the letter and the fax. The Internet has become the de facto communication network for the exchange of such documents. Between applications in organizations, electronic data interchange (EDI) has moved from proprietary networks to Internet-based communication.

The Need for Document Exchange

Today, the need for document exchange includes external parties. Companies need to exchange documents with customers and partners, and in the public sector, communication with citizens and corporations increasingly is electronically based.

The government sector has witnessed a blossoming of investments in information and communication technologies since the late 1990s. In Europe, national programs and European Commission action plans spurred attention toward deploying electronic communication channels. The eEurope 2005 Action Plan stated that "By end 2004, Member States should have ensured that basic public services are interactive, where relevant, accessible for all, and exploit both the potential of broadband networks and of multi-platform access." The sixth annual survey carried out on behalf of the European Commission highlighted that in 2006, 92% of analyzed services had a Web site with information and downloadable forms, and 48% offered full two-way transactional capabilities to citizens.

IDC believes that the short- to medium-term IT investments in the government sector will be driven by, among other factors, the need to increase customer centricity of service delivery, including increased use of electronic communication.

In the commercial sector, exchange of information between enterprises and other external parties and customers has increasingly moved to electronic formats, including areas such as eprocurement, electronic invoicing, and esourcing. In addition, supply chains increasingly automate supply- and demand-side business processes that bring to market a product or service, including multisite organizations involved in a complex supply chain process that includes logistics and inventory management. An example of an initiative in this area is RosettaNet, a nonprofit consortium aimed at establishing XML-based standards for the sharing of business information (B2B), primarily in the supply chain. Other organizations work on industry-specific XML-based standards; for example, CIDX focuses its efforts on the global chemical industry.

Electronic documents play an important role for the enterprise in communication with consumers and other external parties. A visit to the Web site of any consumer-oriented company will reveal a number of documents offered for download — from product instructions to support documents.

Internally, enterprises and organizations have the same and increasing requirements for the exchange of documents. Because most organizations today either are knowledge driven or have a significant knowledge element, the collaboration between employees becomes essential. They increasingly expect to share information and documents, for instance, through team workspaces, enterprise portals, and other business software offerings.

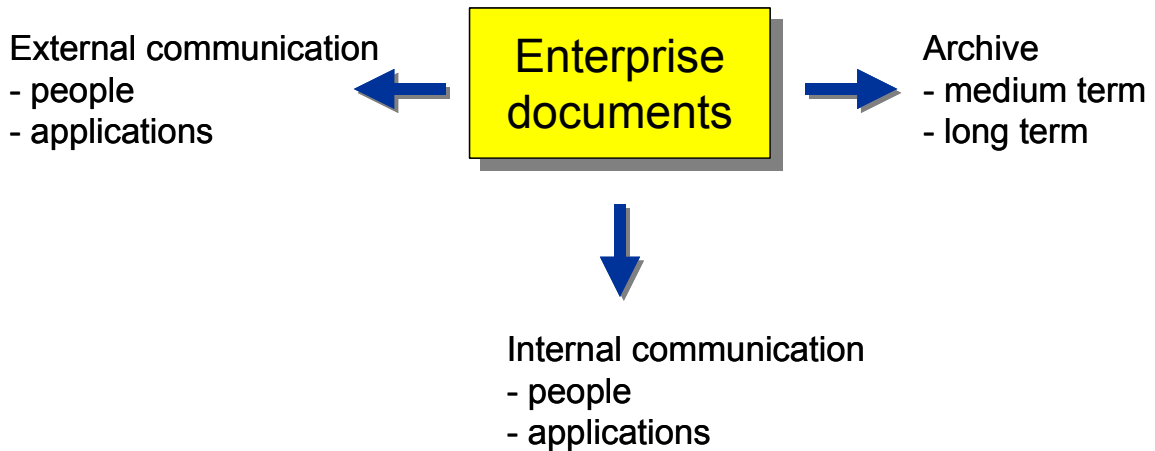
In both the commercial and public sectors, the use of automated document management and workflow systems is growing rapidly. The promise of business process automation is driving investments in the integration of content applications with other business applications in the organization. Document management and the integration of document formats with back-end applications as part of automating the process are among the challenges most organizations face today.

In the public sector, the use of document management in many countries has led to a mushrooming of activities with limited coordination across departmental silos and limited attention to establishing shared processes and information. To address these issues, organizations are launching new strategies to share back-office processes, IT management, and structured and unstructured information more extensively. A key investment driver in the public sector is therefore in solutions to enable cross-organizational collaboration through structured and unstructured information sharing.

In addition to these requirements (as shown in Figure 3), the need to archive and, more importantly, retrieve larger and larger repositories of documents adds a new dimension to document management. For many organizations, it is desirable to be able to store documents for a long time and make sure that if they have to be retrieved, the format of the document is still readable. For public archives, the "long-term" time horizon can be up to 100 years or longer!

FIGURE 3

The Need for Document Exchange



Source: IDC, 2007

Besides the challenge of retrieving old documents in readable formats, it is also important to be able to find certain archived information through search methods, which typically involve tagging and metadata. Therefore, besides the need for readable document formats, it is also important to have standard ways to embed metadata in documents when archived.

Similarly, for regulatory and compliance reasons, it is mandatory for most organizations to archive documents for an extended period of time. This comes in different forms and shapes, such as:

- Meeting regulatory records retention requirements from Sarbanes-Oxley and similar compliance regulations
- Having the ability to efficiently respond to agency audits and legal discovery
- Enforcing internal governance policies

Document Standards

As productivity tools emerged, so did the file formats used by these applications. Typically, different proprietary, binary file formats were used by a wide range of word processing, spreadsheet, and similar packages.

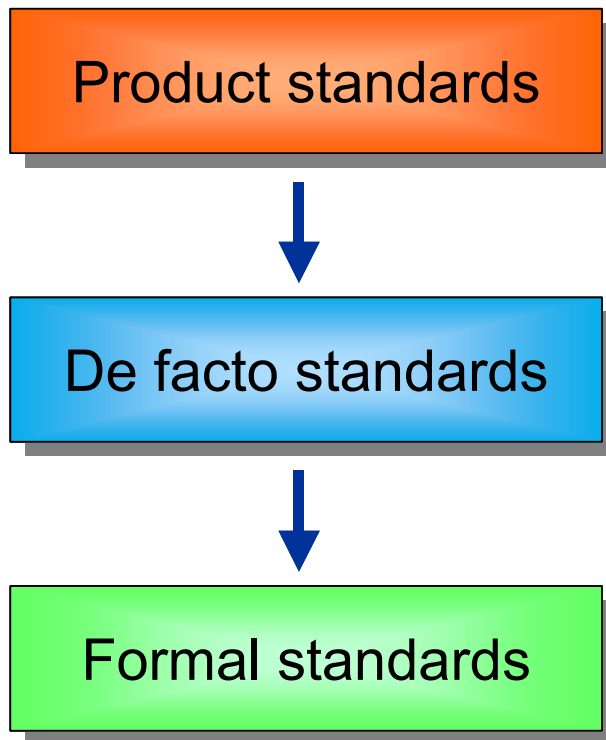
Widespread products such as WordPerfect, Microsoft Word, Lotus 123, and Supercalc all used their own specific file formats.

For all the reasons stated earlier, users required and vendors started to implement product-independent file formats, some of which have become de facto standards in the market (see Figure 4).

One example is the Rich Text Format (often abbreviated as RTF), which is a proprietary document file format developed by DEC in 1987 for cross-platform document interchange. It was purchased by Microsoft around 1990, and the company extended the standard. It is currently at version 1.9, released in January 2007. Most word processors are able to read and write RTF documents.

FIGURE 4

Development of Document Standards



Source: IDC, 2007

Another example is Symbolic Link (SYLK), a Microsoft file format typically used to exchange data between applications, specifically spreadsheets. From within a spreadsheet, data can be exported in the SYLK format. Comprising only ASCII characters, it is easily created and processed by other applications, such as databases.

The Portable Document Format (PDF) is a widely used file format created by Adobe Systems, in 1993, for document exchange. PDF is used for representing two-dimensional (2D) documents in a device-independent and display resolution-independent fixed-layout document format. Each PDF file encapsulates a complete description of a 2D document (and, with Acrobat 3D, embedded 3D documents) that includes the text, fonts, images, and 2D vector graphics that compose the document.

One of the limitations of these standards is that they are proprietary and therefore controlled by a single vendor. Still, PDF and other document formats are widely used, although these standards have limited functionality because they often are a least-common-denominator format or are tied to a particular type of document — or in the case of PDF, a publishing format.

As document standards that are robust, viable in the long term, and product independent are becoming increasingly important to many organizations, the past couple of years have seen a move toward formally approved document standards. PDF/A — a format for long-term document preservation — was approved by ISO in 2005, and Adobe is moving forward with ISO approval of PDF in general.

Other formal document standards are based on XML, which is a general-purpose markup language. It is classified as an extensible language because it allows the definition of context-specific tags. Its primary purpose is to facilitate the sharing of data across different information systems, particularly via the Internet.

XML is a strong foundation for defining standards, such as document standards. By leaving the names, allowable hierarchy, and meanings of the elements and attributes open and definable by a customizable schema, XML provides a syntactic foundation for the creation of purpose-specific, XML-based languages. The general syntax of such languages is rigid — documents must adhere to the general rules of XML, ensuring that all XML-aware software can at least read and understand the relative arrangement of information within them.

A number of XML-based standards have therefore emerged. They are found within areas such as EDI and industry-specific data interchange as well as within government for application-specific formats. The earlier cited CIDX promotes its Chem eStandards based on XML and developed specifically for the buying, selling, and delivery of chemical products.

Two XML-based general productivity document formats currently attract attention. One is the ODF, which was approved by ISO as a formal, international standard in 2006.

ODF originates from the German software company StarDivision, which in 1999 was working on a nonbinary, standardized document format for its word processing product when it was acquired by Sun Microsystems. In 2000, Sun decided to transfer the product into the open source software domain. The work to move it to a formal standard has been done through the Organization for the Advancement of Structured Information Standards (OASIS).

The other format is Office Open XML. Gradually, Microsoft has moved its Office document formats (.doc, .xls, .ppt) from binary formats to XML-based formats, starting with Office beta in 1998. With the release of Office 2007, XML-based file formats became the default in Word, Excel, and PowerPoint. The new file formats are extensions of the WordprocessingML and SpreadsheetML schemas introduced in previous versions.

Microsoft submitted Open XML to the industry standard organization Ecma, which approved the standard in late 2006. Subsequently, Ecma has submitted Office Open XML to ISO's fast-track approval process, which is ongoing.

FUTURE OUTLOOK

Key findings from IDC's survey on the document standards of 200 organizations in the United States and Europe are presented in this section.

Adoption of Document Standards

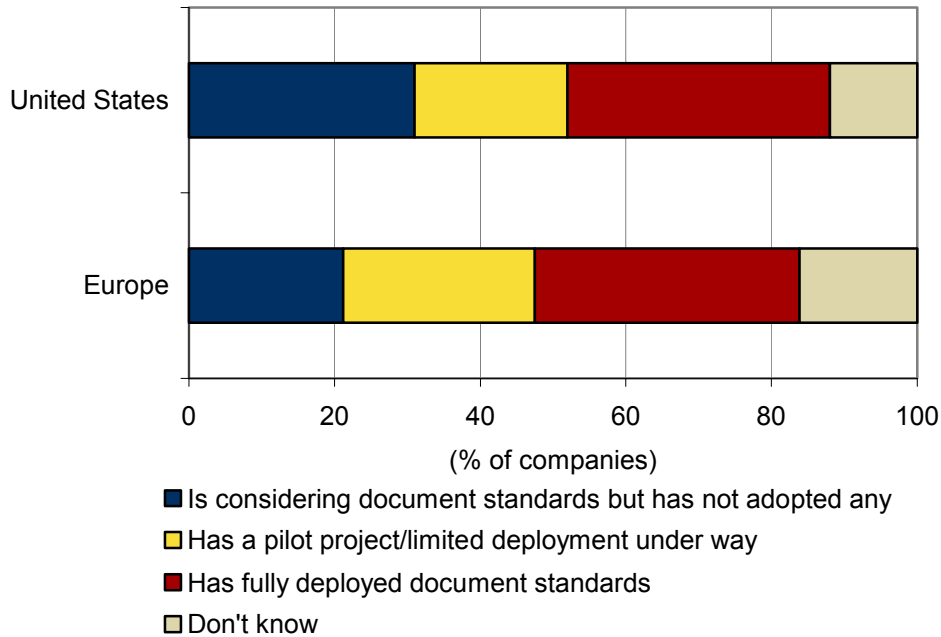
The sample for this survey consists of organizations that have stated that they have a strategy for document standards. Therefore, by definition, most are considering document standards or have implemented one or more standards already.

More than a third of the respondents have fully deployed document standards; the other two-thirds are split between considering document standards and having a pilot project/limited deployment under way (see Figure 5). There are no significant differences by organization size.

FIGURE 5

Status of Adoption of Document Standards

Q. Which of the following best describes your organization's current stage of document standards adoption?



n = 200

Note: Data is weighted by organization's primary business.

Source: IDC, 2007

Generally, more organizations in Europe than in the United States are in the pilot or deployment phase.

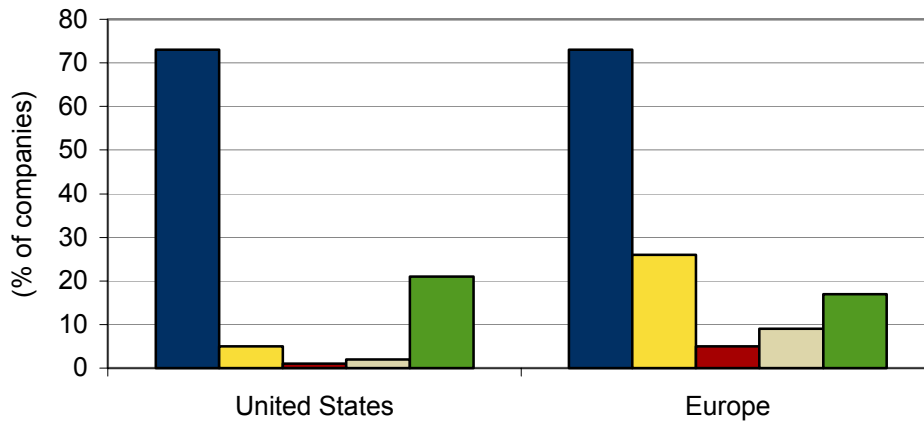
Overall, no major differences are found in the adoption pattern across industries, although business/professional services organizations appear to be more advanced as most organizations are fully deploying document standards. The public sector seems overall to be in a pilot/consideration stage rather than fully deploying document standards, particularly in the United States.

Not surprisingly, the dominant document standard today is PDF, according to more than 70% of organizations in both the United States and Europe (see Figure 6). Another general conclusion is that Office Open XML has created significantly more traction in the market than other XML-based standards such as ODF. "Other" standards include DATA and Docbook.

FIGURE 6

Document Standards in Use Today

Q. Which document standards are in use in your organization today?



- Adobe PDF
- Office Open XML
- ODF
- Other XML
- Others

n = 200

Note: Data is weighted by organization's primary business.

Source: IDC, 2007

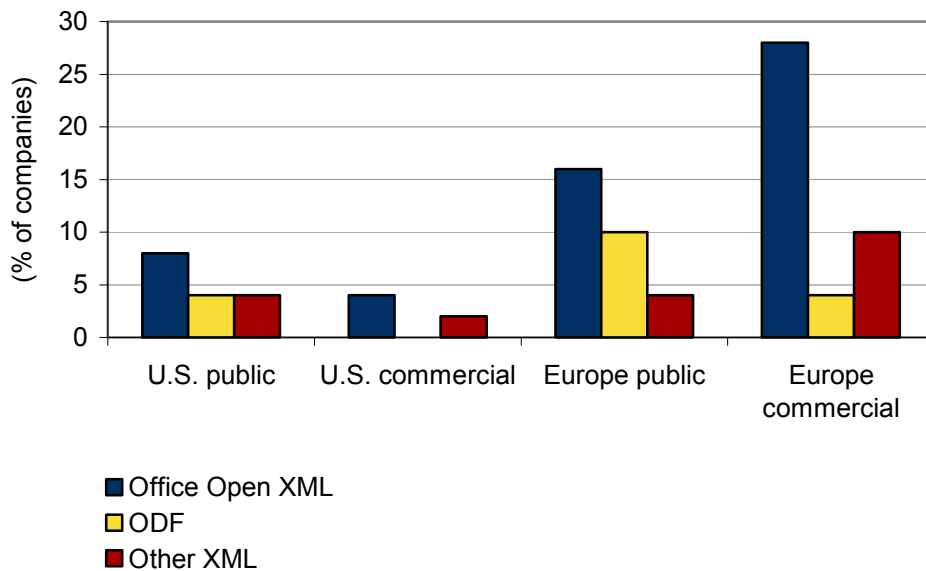
Adoption of XML-based document standards is particularly strong in Europe, where they are seen as additional standards to be used among several in organizations. This is true for Office Open XML as well as other XML standards. Comparing the benefits stated for document standards in Europe versus the United States, we noted that the key reasons for higher adoption in Europe could be the wish to be able to choose productivity tools more freely as well as more emphasis on long-term archiving.

We did not find differences by size of organization, but we found some key differences by industry sector when we zoomed in on the XML-based standards. Although Office Open XML is clearly preferred in both public and private sectors in the United States and in Europe, adoption of ODF is stronger in the public sector than in the private sector — both in the United States and in Europe (see Figure 7).

FIGURE 7

XML-Based Document Standards in Use Today

Q. Which document standards are in use in your organization today?



n = 200

Source: IDC, 2007

Satisfaction with Document Standards

Satisfaction with document standards was measured on a scale from 1 to 5, where 5 is very satisfied and 1 is very dissatisfied. Thus, 2.5 is average.

Overall, the satisfaction with document standards is quite high — generally 3.5 or higher. PDF scored quite high in satisfaction in both Europe and the United States (see Figure 8). It seems to be a good, documented, and very stable standard with good backward compatibility.

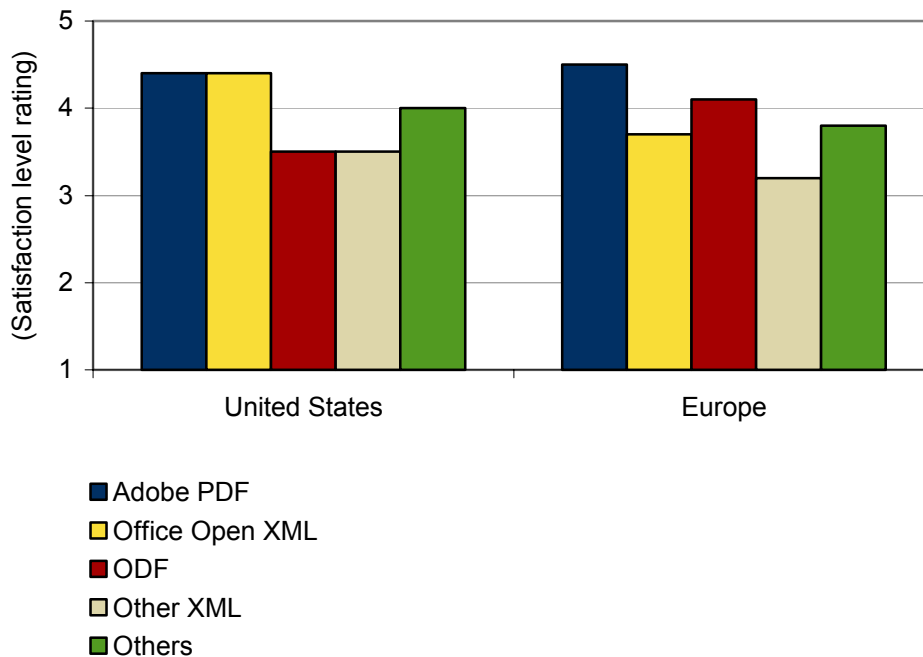
In the United States, Office Open XML is rated with the highest satisfaction among XML-based standards on par with PDF. In Europe, PDF is followed by ODF and Office Open XML.

In Europe, Office Open XML satisfaction is higher in organizations deploying multiple standards than in organizations deploying one standard (the reverse is true in the United States). Satisfaction levels for ODF and PDF are higher in Europe in organizations deploying just one standard than in organizations deploying multiple standards.

FIGURE 8

Satisfaction with Current Document Standards

Q. Using a 5-point scale, where 1 is very dissatisfied and 5 is very satisfied, please rate your current satisfaction with ...



n = 200

Note: Data is weighted by organization's primary business.

Source: IDC, 2007

Future Strategies for Document Standards

The overall picture is not expected to change much 12 months from now. PDF is showing the characteristics of a mature standard. Slightly fewer companies expect to have PDF fully deployed one year from now.

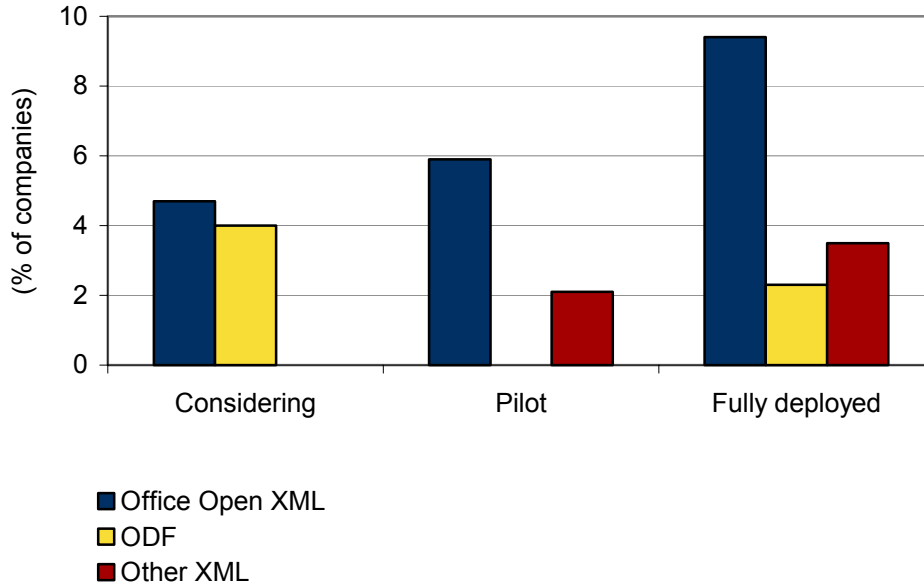
Office Open XML is the standard showing the most progress over the next year (see Figure 9). Among companies having an interest in Office Open XML, more will be piloting or fully deploying the standard in one year compared with today (this finding relates only to Europe because the sample of U.S. companies is too small for conclusions on this point).

Although ODF is a standard with a longer history in the market, it shows a different pattern. In one year, most of the companies showing an interest in ODF will still be in the consideration phase rather than the "pilot" or "fully deployed" phase (this finding relates only to Europe for the reasons noted earlier).

FIGURE 9

Status of XML-Based Document Standards in 12 Months (Europe Only)

Q. Which will be in use in 12 months?



n = 200

Source: IDC, 2007

With regard to the preferred number of standards, there is a significant difference between the United States and Europe (see Figure 10). In Europe, organizations generally prefer one standard (twice as many organizations prefer one standard versus multiple standards). In the United States, the number of organizations preferring one standard is the same as that preferring multiple standards.

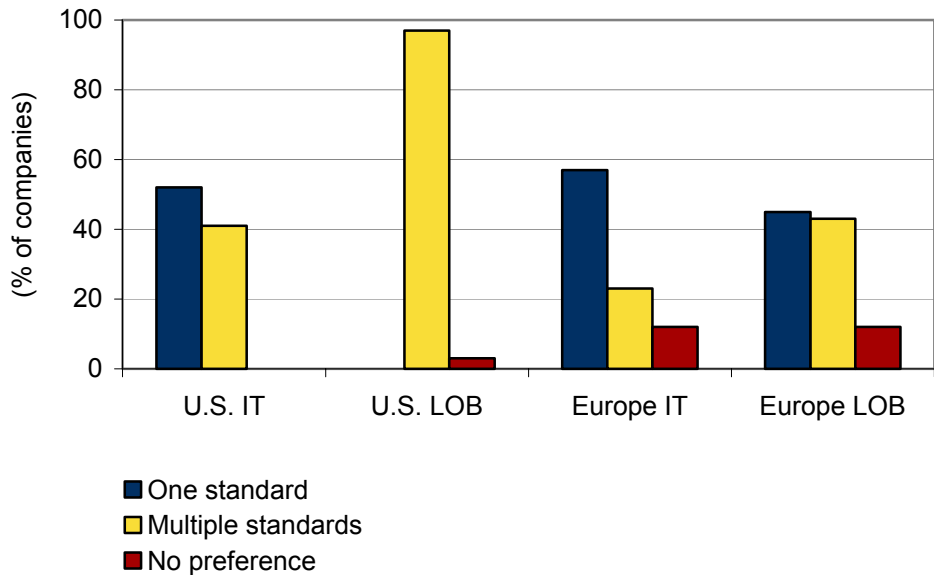
It is interesting that IT managers and LOB managers have different preferences. The highest share of respondents with a preference for one standard is among IT managers, while the highest share of respondents with a preference for multiple standards is found among LOB managers. This is true for both the United States and Europe. It seems that IT managers wish to avoid the complexity and cost of supporting multiple standards. On the other hand, LOB managers relate to the daily needs of the business and can see the necessity of more than one document standard.

Diverse business needs is most likely also the reason that large organizations have a preference for multiple document standards, while small organizations have a preference for one document standard.

FIGURE 10

Preference for One or Multiple Document Standards

Q. *Would your organization prefer to define one or multiple document standards across the organization?*



n = 200

Note: Data is weighted by organization's primary business.

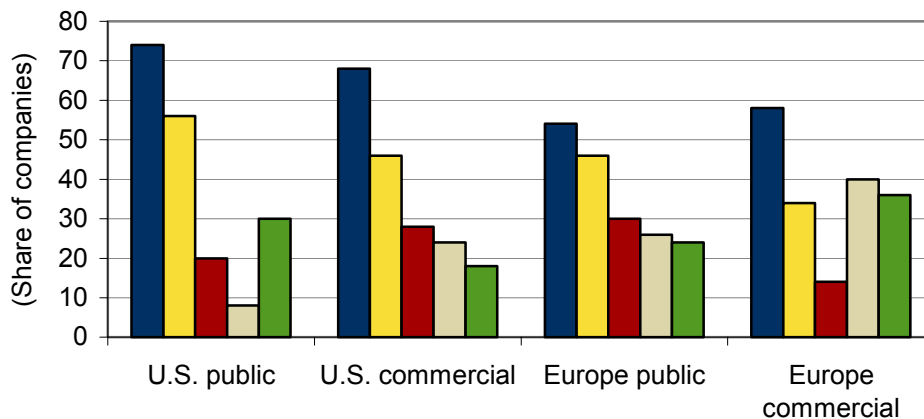
Source: IDC, 2007

Overall, the dominant benefits from document standards are seen as the improved ability to exchange documents internally and the improved ability to exchange documents externally — both in the United States and in Europe (see Figure 11). This finding supports the key points noted in the Situation Overview section.

FIGURE 11

Most Beneficial Aspects of Using Document Standards

Q. Please choose the two most beneficial aspects of using document standards.



- Easy document exchange internally
- Easy document exchange with external partners
- Interoperability between productivity tools and other applications
- Freedom in choice of productivity tools
- Long-term archive/access

n = 200

Source: IDC, 2007

In the commercial sector in Europe, there is relatively high focus on freedom of choice of productivity tools and long-term archiving, while the public sector in Europe puts relatively more emphasis on the interoperability between productivity tools and other applications (e.g., workflow systems).

Freedom of choice of productivity tools plays a much smaller role in the United States than in Europe. Instead, interoperability between productivity tools and other applications is seen as important, as well as long-term archiving in the public sector.

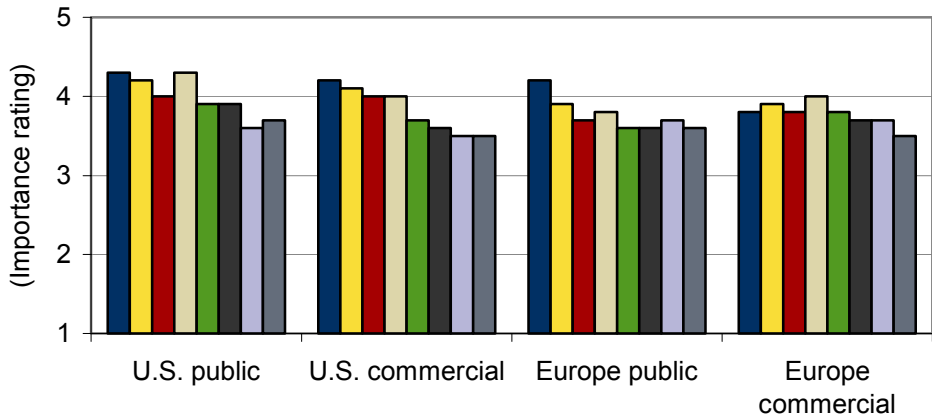
The least important benefits from document standards are political correctness and increased competition in the market.

As Figure 12 shows, respondents rated all of the characteristics of a document standard as fairly important (3.5 or higher on a scale from 1 to 5).

FIGURE 12

Important Factors When Implementing a Document Standard

Q. Using a 5-point scale, where 5 is very important and 1 is not important, please rate the importance of the following in terms of implementing a document standard.



- Allows easy transition of existing base of documents to new standard
- Has reasonable implementation costs
- Has complete functionality
- Is supported by a wide range of tools available on the market
- Is supported by your organization's productivity vendor
- Is platform independent
- Is truly open
- Has been approved by a formal standards body

n = 200

Source: IDC, 2007

Easy transition of the existing base of documents to a new standard is rated most important in all segments except the commercial sector in Europe. The ability to migrate the existing base of documents plays a particularly important role for large commercial organizations and organizations in the public sector. The other most important factor when implementing a standard is the cost.

The next level of key factors is the availability of a wide range of tools for the standard — particularly in the U.S. public sector and in the European commercial sector — as well as complete functionality.

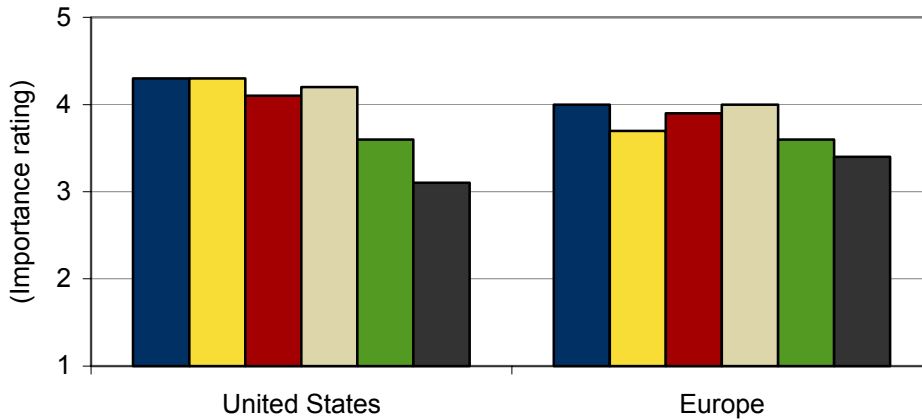
Least important is "truly open" and "approved by a standards body." These factors may be hot in the press, but they are not very important to end users.

Interoperability and integration are the most important concerns together with price/total cost of ownership (TCO) and functionality when purchasing productivity tools (see Figure 13). However, interoperability does not automatically translate into an interest in "open standards," maybe because as previously noted, "openness" is not seen as a mandatory requirement for a document standard.

FIGURE 13

Important Parameters When Purchasing Personal Productivity Software

Q. *Using a 5-point scale, where 1 is not important and 5 is very important, please rate the importance of the following when purchasing personal productivity software for your organization.*



- Allows easy integration and interoperability
- The total cost of ownership
- The price
- The product functionality and capability
- Allows the possibility of customization
- Uses open standards

n = 200

Note: Data is weighted by organization's primary business.

Source: IDC, 2007

Open standards play a role, particularly in the public sector in the United States and in Europe.

Cost Factors

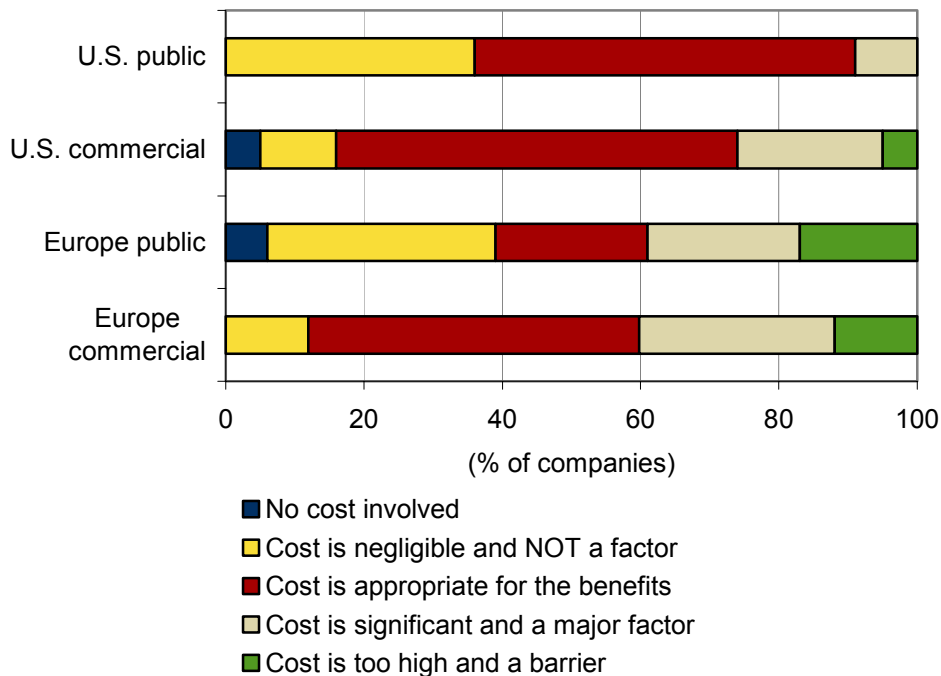
Most organizations believe that the costs of moving to document standards are appropriate for the benefits (see Figure 14). The costs of moving to document standards are less of an issue for public sector organizations than for commercial organizations: In the public sector, a significantly higher percentage of organizations stated that costs are negligible and not a factor in the decision of moving to document standards.

Some organizations find costs a barrier to implementing document standards. They are primarily found in Europe, but constitute less than 20% of organizations.

FIGURE 14

How Big a Factor Is Cost in Moving to Document Standards?

Q. Which of the following best reflects your opinion of the costs of moving to a document standard?



n = 66

Note: Data is weighted by organization's primary business.

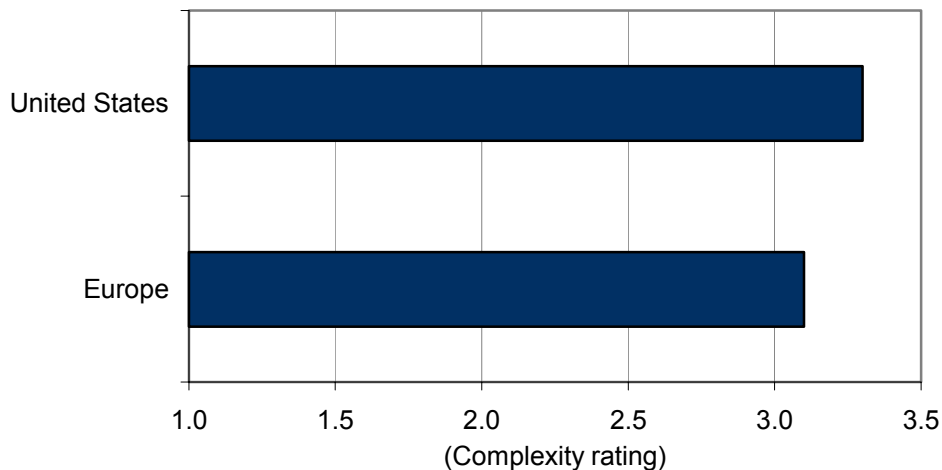
Source: IDC, 2007

The same absence of major barriers is reflected in the complexity of making a decision to move to document standards. The decision to move to a document standard is generally seen as not very complex in the United States and in Europe (see Figure 15). There are no major differences by size of company or by sector (public versus commercial).

FIGURE 15

Complexity of the Decision to Implement Document Standards

Q. *Using a 5-point scale, where 1 is not very complex and 5 is very complex, please rate the complexity of the decision to implement document standards in your organization.*



n = 200

Note: Data is weighted by organization's primary business.

Source: IDC, 2007

CHALLENGES/OPPORTUNITIES

Although XML-based document standards have been developed and deployed over the years, we are still in the early days of wide-scale use of these document standards. Many organizations are still piloting and exploring how to use nonbinary document formats, primarily XML.

Multiple XML-based document standards seem to emerge. Vendors and open source communities are building translators and converters between the alternative standards. Still, some users are critical of the complexity of having alternative and overlapping document standards.

XML document standards provide a document format that is independent of vendor platforms and vendor productivity products. But the format also provides functionalities that organizations generally still need to explore, including the use of metadata in document files that can prove useful in an organization's information life-cycle management (ILM).

Another key feature of XML is schemas. An XML schema can be used to express a set of rules to which an XML document must conform to be considered "valid" according to that schema. However, unlike most other schema languages, XML schema was also designed with the intent of validation, resulting in a collection of information adhering to specific data types, which can be useful in the development of XML document processing software.

XML schemas allow for the development of custom-defined schemas. Organizations can use custom-defined schemas to develop special, context-specific versions of XML document standards for specific purposes.

Long-term viability and stability of a standard should always be of concern to organizations. In the area of document standards, it is even more important because these standards will also be the basis for long-term archiving of documents. And long term could mean really long term, which is a nontrivial issue for both private companies and public organizations such as national archives.

This puts extra emphasis on the aspects of a document standard being a formal standard (e.g., nonproprietary and governed by an independent standards body). Although organizations in this survey did not put particular emphasis on whether document standards are controlled by a standards body, IDC believes it should be a priority for organizations.

CONCLUSION

The number of documents created by productivity tools is increasing exponentially, with no flattening of the curve in sight. At the same time, the use of networks — the Internet in particular — has become the primary channel for communication and exchange of information. This is true both internally in organizations and externally with customers, partners, and suppliers.

This situation has created an increasing need for product-independent document standards with a high level of accuracy between applications. XML-based standards are growing in importance in the market, with the benefits of XML being a nonbinary W3C standard that is relatively human-legible and enables the sharing of information over the Internet. XML is also flexible and allows users to specify tags and implementations with specific specifications and constraints.

Productivity document standards such as Office Open XML and ODF have emerged based on XML, driven by the need for interoperability between different productivity tools and other applications, preferably without loss of any of the extended and advanced functionality found today in productivity tools. XML is a language that is flexible enough to accommodate complex functionality and at the same time provide the possibility of reinforcing constraints. It is also a language that allows developers and users alike the possibility of defining custom extensions (called schemas) to meet the need for specific, context-related versions of a specification.

A standard does not live just in the present. The importance of time, both past and future, is essential, not least when working with productivity documents. Billions of documents already exist that users will want to preserve for the future. The emerging document standards provide the promise of low-risk standards for long-term archiving.

A survey of 200 organizations confirms these directions. It is clear from the survey that an increasing number of organizations are moving to XML-based standards for productivity documents. The interest in such standards is somewhat more profound in Europe than in the United States, but it does span both public and private organizations. Among the XML-based document standards, Office Open XML seems to be creating the most traction in the market.

Among the biggest advantages organizations find in moving to document standards is increased interoperability (e.g., easier exchange of documents with internal as well as external parties). When organizations implement standards, cost and the ability to move existing documents to the new standards are very important. Whether the standard is "open" is of less importance to the organizations.

The survey has not revealed any major barriers for companies moving to document standards. The decision to move to a standard is seen as of average complexity by most organizations, and the cost factor is seen as a barrier for just a minority of organizations. However, IT managers do seem to favor just one standard, probably for manageability reasons, while LOB managers prefer multiple standards. Therefore, when LOB managers look at an XML standard, they see it as one of several standards deployed in the organization.

Copyright Notice

External Publication of IDC Information and Data — Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2007 IDC. Reproduction without written permission is completely forbidden.