Chapter IX
Improving Access to Government Information with Open Standards for Document Formats

Rajiv C. Shah
University of Illinois at Chicago, USA

Jay P. Kesan
University of Illinois College of Law, USA

ABSTRACT

An important element of transformational government is improving the access and use of government information. Effective use of government information requires government to move beyond traditional proprietary document formats because these formats limit access and use of information stored digitally. In this chapter, we show how document formats based on open standards and the Extensible Markup Language (XML) can revitalize citizens’ access and use of government information. We also offer an improved definition of open standards to assist governments in selecting which document formats to adopt. A critical part of this new definition is an emphasis on multiple implementations of an open standard, which is also known as running code.

INTRODUCTION

The rise in the use of digital media by governments has created vast amounts of digital information. Much of this information is stored in various document formats, such as DOC, XLS, and PPT. The developers of these formats hold proprietary control over them and therefore limit users’ widespread access to them. A new set of document formats, however, based on open standards and
Improving Access to Government Information with Open Standards for Document Formats

the Extensible Markup Language (XML) has the potential to allow easier dissemination and access to this information. In this chapter, we explain how these formats function and provide governments with guidelines for selecting document formats that will allow a greater level of access and use of government information.

Document formats are specifications for storing electronic information. A myriad of document formats exist for various tasks from word processing to architectural drawings. Most citizen interaction with digital government is with documents formatted as Web pages. While the Web has greatly enhanced citizen interaction, government Web pages cover only a small portion of the information developed and collected by government. Virtually every piece of information on a Web page began in another document format. As such, most of the existing document formats make it difficult to search for this information as well as repurpose it for other uses.

The next generation of document formats are based on open standards. Open standards refer to specifications that are publicly available and developed through a public process. They stand in contrast to standards that are proprietary and restricted to certain parties. A significant benefit of open standards is that anyone can implement the standard. The result may be a proliferation of competing implementations that leads to economic and technological benefits for adopters.

This next generation of document formats is based on XML. XML uses markup tags to add structure to documents. The power of these tags is that it fosters more powerful searching and the ability to access and manipulate information within documents. For example, the U.S. Patent and Trademark Office (USPTO) uses an XML-based system for processing applications. By receiving XML-based documents, the USPTO can avoid the cumbersome process of scanning in paper documents, especially because the submittals to the USPTO start as electronic documents. The XML format uses a set of tags to mark aspects of the documents’ procedural data (names, addresses, fee data, title of invention, etc.) and application data (abstract, description, claims, figures, patent citations, and drawings). With this information “marked” in a document, the patent office can automatically process a document. For example, the patent office could extract information manually regarding the inventor and document. Without an XML-type format, it would be necessary to read each document and then “cut and paste” the relevant information.

The first section of this chapter provides background on various document formats, including the most widespread formats developed by Microsoft, two next generation XML-based document formats, ODF and OOXML. The second section reviews the benefits of open standards, which are a critical reason why governments are moving to document formats based on open standards. The final section focuses on the criteria which government should use in the selection of the next generation document formats.

BACKGROUND ON DOCUMENT FORMATS

The most recognized document formats are Microsoft’s Word DOC format for word processing, Excel’s XLS for spreadsheets, and Powerpoint’s PPT format for presentations. However, there are countless other formats that are used by government. Information produced and used within government is created and modified within document formats. The ability to edit, search, and analyze this information directly can transform how government manages itself as well as its relationship with its citizens. Consider the example of Thailand:

December 26, 2004. 7:58 am. A thirty-foot-high wall of water—a tsunami—slams into the famed resort islands off Thailand’s southern coast. In one tragic moment, thousands of lives are lost, and thousands more are missing.