

Report to WIPO SCIT Plenary International Priority Document Exchange Primer



February 1, 1999

Standard	Description	Product	Comments
	<p><u>Character Coded Files</u> Document files where the document information is in machine-readable character codes. Character coded files can be searched by search tools for the existence of character coded search terms.</p>	<p>Text editors, SGML editors, word processors and many other applications typically create character-coded files.</p>	<p>Most character-coded files contain the document data and other characters used to format the data for presentation on a page. SGML files tend to exclude formatting information.</p>
	<p><u>Directory Structure and Filenames</u> This refers to conventions used to exchange electronic documents consisting of multiple files.</p>		<p>The Trilateral Partners have agreed to put document files in a directory or folder that is named to identify the type of documents contained. Filenames follow the convention "filename.type", where filename is an 8 character alphanumeric identifier and type is a 3 character specifier.</p>
	<p><u>Document Certification</u> The process by which an organization endorses the official status and content of a document. Paper documents may be certified by permanently binding numbered pages and adding a signed official cover sheet. An electronic certified document may carry a digital signature and the digital certificate of the endorsing organization.</p>	<p>Email software and electronic commerce software are beginning to offer digital signature features.</p>	<p>The digital signature seals the content of a document so that if the value of one bit of data changed the signature would not validate correctly. A digital certificate is used when creating a digital signature. The certificate establishes the identity of the signer to the extent that it can be trusted.</p>

Standard	Description	Product	Comments
	<p><u>Document Image Files</u> Document files containing a bitmap image (facsimile) of one or more document pages. Image files are created by scanning a paper document. The document image is not machine-readable and cannot be searched directly and efficiently.</p>	<p>Commercial document scanners and scanner software are used to create document image files.</p>	<p>Group 4 facsimile format is a popular standard for storing page image data. ST.33 and TIFF are image file standards used in the patent and trademark business.</p>
<p>DTD</p>	<p><u>Document Type Definition</u> A formal description (written in the SGML syntax) of the allowed tags and structure of a particular SGML document type. For example: USPTO has just released its DTD for Published Patents (ST.32/US/Grant) that will be produced as SGML documents by the data capture contractor.</p>	<p>Commercial off-the-shelf tools are available for creating DTDs.</p>	<p>USPTO has created four DTDs to be used for IPDE.</p>
<p>SGML</p>	<p><u>Standard Generalized Markup Language</u> An international standard (ISO 8879) for defining markup used to indicate the structure and content of documents. SGML tags are inserted with document text to indicate structural parts and content elements. For example: <FNI>Thomas A. Edison</FNI> would be a valid SGML tagging of the first named inventor.</p>	<p>Commercial off-the-shelf SGML authoring tools and SGML enabled word processors are available to create SGML documents.</p>	<p>IPDE documents are SGML documents. In the near term, while priority document images are exchanged, small SGML documents are created for descriptive data and the document images are referenced by the SGML document as external entities.</p>

Standard	Description	Product	Comments
ST.33	<p><u>ST.33 WIPO Standard</u> A WIPO standard defining the formats to be used for the data exchange of patent information in facsimile form. The coding scheme for the image data is based on the Modified READ II data compression technique for CCITT Group 4 facsimile described in recommendation T.6. The ST.33 image data is preceded by a character coded record prefix that contains image layout information and patent bibliographic data.</p>		<p>The ST.33 standard is used as an exchange standard for images in the intellectual property community. Some intellectual property image handling systems use the ST.33 format directly while others extract the bibliographic data and convert the image and image metadata to TIFF format.</p>
TIFF	<p><u>Tagged Image File Format</u> A widely-used image file format for data storage and interchange. TIFF accommodates several compressed image file formats including CCITT Group 4. The general nature of TIFF allows it to be used in any operating environment, and it is found on most platforms requiring image data storage. The TIFF Standard is maintained by Adobe Software.</p>	<p>Most scanner and graphics software can output image files in TIFF format.</p>	<p>TIFF has been recognized as an interchange standard in the intellectual property community by making it an option in the ST.35 standard.</p>

Standard	Description	Product	Comments
XML	<p><u>eXtensible Markup Language</u> A new markup language based on SGML that is optimized for web-based publishing of structured data. A working group within the World Wide Web Consortium has developed the XML standard. XML documents resemble SGML documents with text data, tags, and markup for elements, attributes and entities.</p>	<p>Xmetal from SoftQuad BladeRunner from Interleaf XML Notepad from Microsoft</p>	<p>XML is a streamlined subset of SGML. The inventors of XML created it by removing the little-used features of SGML and adding new rules that make it easier to build software for processing XML. XML documents do not have to have a DTD provided their element usage conforms to a strictly hierarchical structure. The potential for XML is that is suitable for electronic patent documents and it is expected to be popular on the World Wide Web.</p>