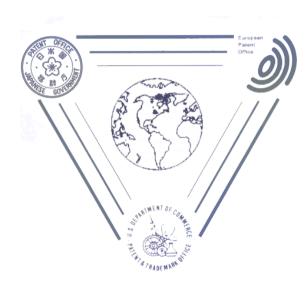
Report to WIPO SCIT Plenary International Priority Document Exchange Primer



February 1, 1999

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

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

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

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