Formats for long term preservation, e-Archive for long term preservation and access to electronic records. Based on rules issued by the Swedish National Archives, RA-FS 2009:2

A file format is a way of encoding information within a computer file. A program or application must be able to recognise the file format in order to access the data within the file.
Files in proprietary formats usually must be opened by the software in which they were created. Without a license to the software, is not possible to access the file.

Data formats that conform to an agreed international standard are less likely to become obsolete, because a variety of software applications should be able to read them.

The table below specified recommended archival formats for different files:

Name	Standard	Format		Comment
Databases and register				
Databases and register, text. encoding				
Sequence files	SS-ISO/IEC 8859-1:2004 8-bits Part 1: Latin no 1 (ISO/IEC 8859-1:1998, IDT), or ISO/IEC 10646:2003 Universal Multiple-Octet Coded Character			
Sequence files, comma separated				
Extensible Markup Language (XML)				
Format description for text based on mark-up language				
	SS-ISO 8879 text processing systems, structure (SGML) -			
	ISO/IEC 15445 Information technology - Document			

	description and processing languages – HyperText Markup Language (HTML),			
	Extensible Markup Language (XML)			
	Extensible HyperText Markup Language (XHTML)			
Schemas, Document Type Definition (DTD) and style sheet should be preserves electronically				

Textual formats

	SS-ISO/IEC 8859-1:2004 8-bits PartI 1: Latin no. 1 (ISO/IEC 8859-1:1998, IDT)			*Majority of West European languages*
	UTF-8 and UTF-16, Unicode Transformation Formats (as ISO/IEC 10646:the Universal Character Set)			*All alphabets and different symbols; comprised as 8-32 and 16/32* *Note! Do not use BOM (Byte Order Mark, control characters); may be interpreted in different ways*

Electronic signatures

	IETF RFC 2315 PKCS #7: Cryptographic Message Syntax Version 1.5			
	IETF RFC 2315 PKCS #7: Cryptographic Message Syntax Version 1.5			

E-mail

database or register	As former			

Text with mark-up language	XML or as textual formats			

Still image, scanned

Photographical	SS-ISO 10918 – Digital comprising and encoding (JPEG)			
	ISO 12639:2004 TIFF/IT – Tag image file format for image technology, Avsnitt 7.7 Binary picture (BP) file.			ITU-T (CCITT) T.6 Facsimile Coding Schemes and Coding Control Functions for Group 4 Facsimile Apparatus as for data comprising
	ISO/IEC 15948:2004 Information technology – Computer graphics and image processing – Portable Network Graphics (PNG): Functional specification			
Scanned	SS-ISO 19005-1:2005 – Del 1: Use of PDF 1.4 (PDF/A-1) (ISO 19005-1:2005, IDT)			

Geospacial

	SS-ISO 19136:2007 Geographic information (GML) (ISO 19136:2007, IDT)		GIS	
	ISO 24517-1:2008 Document management – Engineering document format using PDF – Part 1: Use of PDF 1.6 (PDF/E-1)			
Geospacial, raster format	ISO 24517-1:2008 Document management – Engineering document format using PDF – Part			

	1: Use of PDF 1.6 (PDF/E-1)			

Web

	ISO/IEC 15445 Information technology – Document description and processing languages - HyperText Markup Language (HTML)			
	SS-ISO 19005-1:2005 Text processing- Del 1: Use of PDF 1.4 (PDF/A-1) (ISO 19005-1:2005, IDT)			
	XML			
	Extensible HyperText Markup Language (XHTML)			Schemas, Document Type Definition (DTD) and stylesheet should be preserves electronically
Characters and codes	ISO/IEC 10646			

File formats not mentioned in RA-FS 2009:2

PDF/E+U3D			2/3D (CAD)	
WebCGM			2D (CAD)	
MPG			Sound and moving image	
LPCM WAVE			Sound	

Standards and formats examined by the National Archives

eARD (e-archive project) (recommendation)				XML-files and XML-schemas from eARD http://xml.ra.se/e-arkiv/

Flat files and databases	**ADDML**			http://xml.ra.se/addml/
Business systems/ economy	**BTLA**			http://xml.ra.se/economics/
Archival metadata (creators)	**EAC (**Encoded Archival Context)			http://xml.ra.se/eac/
Archival description	**EAD (Encoded Archival Description)**			http://xml.ra.se/ead/
Information packages in OAIS-model	**METS (Metadata Encoding and Transmission Standard)**			http://xml.ra.se/mets/
Still images	**MIX (NISO Metadata for Images in XML)**			http://xml.ra.se/mix/
Preservation metadata	**PREMIS (Preservation Metadata:Implementation Strategies)**			http://xml.ra.se/premis/
Xlink schema	xlink			http://xml.ra.se/xlink/

Summary of information about data formats and standards (may be completed)

Data formats

	ISO 11179 ”Specifica ion an Standardization Of Data Elements, Parts 1-6			
Representation of dates and times	W3Cs begränsning utifrån ISO 8601;1988 (E)	YYYY-MM-DDThh:mm:ssTZD		
Names, physical persons		Sure name, first name , i.e Andersson, Per Andersson, PER Arne		
Security number		YYYYMMDDXXXX	12 characters	

Organisation ID		NNNNNNMMMMMM		12 characters
Postal address	SS 61 34 01 "po ervice", Sweden	Address c/o-name Post code City Country		
Geographical coordinates	ISO 6709	"+27.5916+086. 5640+8850 " (Mount Everest)		Latitude and longitude, different combination
	WGS 84 (Word Geodetic System 1984)	La 59°19´37"N, Long 18°4´18"O		
	Sedan 2007 SWEREF90 TM	N=X E=Y		maps; x and y coordinates
Language codes	ISO 639 Two-Letter Language Code	sv		
	ISO 639-2	swe		Three characters
County codes, Sweden	ISO 3166 Two-Letter Country Code	SE SE-K/SE-10		
Currency	ISO 4217	SEK, EUR		
Bank identifier code	ISO 9362			Bank identifier codes, BIC-/ SWIFT-code: 8 till 11 characters
Bank account no.	ISO 13616 (IBAN)	SEkk BBBB AAAA AAAA AAAA AAAA		B=bank/clearing number Kk=control characters as in ISO 7064 modula97-10
Structure unit , archival description	RA-FS 2008:4	4.1.1		Decimal, dot notation
File and catalogue names		Not åäöÅÄÖ Spacing as _ Dot as _ or p None special characters as ?&*+ or likewise		