

Preservation Action Plan: Software/Code National Archives and Records Administration (NARA)

DRAFT 20190801
Template: 201907

Electronic Record Types and Associated Formats

Software comprises a set of instructions that instruct a computer to perform one or more tasks. There are three general categories:

- System software includes device drivers, operating systems (OSs), scripts, compilers, disk formatters, text editors and utilities helping the computer to operate.
- Programming software is a set of tools to aid developers in writing programs.
- Application software is intended to perform certain tasks. Examples of application software include office suites, gaming applications, database systems and educational software. Application software can be a single piece of code or a collection of programs.

Software used to *create* government records are not considered permanent records. Software or code that is an integral part of a record, such as something required for the functionality of a dataset that is a permanent record, or that was created as part of the work of government, have the potential to be permanent records.

Essential Characteristics of this Record Type

Appearance

Name	Definition	Function Description
Rendering	Rendering are those factors which determine the recreation of the "message," which in the case of software is how the software is executed in a manner that approximates its original mode of operation/interaction.	What is required for the software to run/render as closely as possible to the original design/look and feel/function/intent.

Structure

Name	Definition	Function Description
------	------------	----------------------

Content	Uncompiled code, compiled binaries, and accompanying documentation.	The components of the software
Software Composition/Component Structure	Documentation of the names and number of components and files; relationships and structure; test cases/suite; compiler; build instructions/scripts; and installers.	Everything needed to physically reconstruct the software object.

Behavior

Information which describes interaction with external stimuli.

Name		
Functional Description	Purpose, release notes, variant notes, manual(s), tutorial(s).	Documentation of the functional behavior of the software to inputs and outputs.
User Interaction	The user interaction model.	Model for the human-computer interactions, such as complex screen layouts, non-standard displays (such as touchscreens), audio, non-keyboard controllers, haptic interaction; anything outside the standard monitor-keyboard-pointer arrangement.

Context

unctional description.

Name	Definition	Function Description
Functional Description	Input specifications, output specifications, error handling documentation, API documentation.	

Provenance and Ownership	Package name, version identifier and license terms. Name of licensee if appropriate. License key value if needed. Name of creator (if created for US government).	
Software Environment	Required environment for the software to operate, including platform; operating system; external software dependencies; required settings; and hardware, graphics card, and controller device dependencies.	
Software Architecture	Documentation of the software architecture, programming language, and internal software package dependencies.	
Operating Performance	Specification/required performance for CPU processor; minimum required and typical memory usage for the CPU and on disk; specification/required performance for peripherals, including monitor and graphics card, audio card, etc.	

Current NARA Transfer Guidance for this Record Type

None.

Current NARA Public Access/Reference Format(s) for this Record Type

This Plan references existing public access file formats for electronic records at NARA, determined with a survey of the available public access formats in the National Archives Catalog. These references do not represent recommended public access formats under NARA policies. They are intended for informational purposes only.

Reference Format: Native/original format(s).

Public Access Format: Download from the National Archives Catalog if available, or request access through the reference staff in the appropriate custodial unit.

Comments and Notes

Software should be transferred in its native/original format, but should be provided as both raw source code and a compiled runtime version, if it exists. It should also be accompanied by release notes, readme files, and user/administrative documentation, if they exist.

File Formats Associated with This Record Type: See Appendix

Proposed Preservation Plan

Retain file format in its existing format.

Transform file to a new format.

Selected Format: _____

Procure/develop tools to preserve, manage and provide access to records of this type in their existing form.

Procure/develop tools to transform the format to the preferred normalized form.

Provide Additional Information so that the record type remains understandable/usable over time.

Explore Additional Options

Justification: Any format migration for compiled or binary code creates a guaranteed outcome that the code will no longer function, as will extracting files from binary container archives files such as .cab.

All uncompiled code such as JavaScript or Configuration files are plain text ASCII or Unicode and do not require format migration.

Preferred Processing and Transformation Tool(s)

For ASCII/Unicode-based code:

- Current Supported text reader

For Compiled Code:

- Do not open to review; preserve as transferred

Preferred Viewer/Access Software

Provide code to researchers in the format in which it was transferred, accompanied by any associated documentation.

Appendix: Formats, Risks, and Details

Format ID	Risk Rating	Risk Priority	Risk Level	Format	Extension	Compiled, Binary, or Plain Text
NF00357	-34	High Risk	Moderate Risk	PGP Encryption Key	pgp	Binary
NF00182	-31.00	-31	High Risk	Executable file	exe	Compiled
NF00499	-31.00	-31	High Risk	Windows New Executable file	exe, dll, cpl, drv, scr	Compiled
NF00380	-27.00	-29	High Risk	Portable Executable file	pe, exe, dll, cpl, efi, ocx, scr, sys	Compiled
NF00169	-26.00	-36	High Risk	Dynamic Link Library	dll	Compiled
NF00135	-21.00	-21	High Risk	Binary file	bin	Compiled
NF00424	-19.00	-19	High Risk	Temp file	tmp, temp	Indeterminate, depending on the original source
NF00351	-12	-14	High Risk	OS/2 Change Control File	cin	Plain Text
NF00257	-11.00	-11	Moderate Risk	Microsoft Active Server Page	asp, aspx	Plain Text
NF00411	-7	-9	Moderate Risk	Checksum File	sum	Binary
NF00110	-6.00	-6	Moderate Risk	Apple Macintosh Desktop Services Store	DS_STORE	Binary
NF00111	-3	-3	Moderate Risk	Compressed Archive File	arc	Binary
NF00166	-3.00	-3	Moderate Risk	DVD Data Backup File	bup	Binary
NF00167	-3.00	-3	Moderate Risk	DVD data file	dvd	Binary
NF00168	-3.00	-3	Moderate Risk	DVD Data File	ifo	Binary
NF00199	-2.00	-2	Moderate Risk	Generic library linker file	lib	Binary
NF00427	-2.00	-2	Moderate Risk	TrueType Font	tff	Binary

NF00276	-2.00	-10	Moderate Risk	Microsoft FrontPage Binary Tree Index	btr	Binary
NF00159	-1	-3	Moderate Risk	Custom Dictionary File	dct, zdct	Plain text
NF00158	-1	-3	Moderate Risk	Dictionary Exclude File	lex	Plain text
NF00151	1	-4	Moderate Risk	Data File	dat	Binary
NF00134	1.00	1	Moderate Risk	Batch file (executable)	bat, cmd, btm	Plain Text
NF00226	1.00	1	Moderate Risk	Lock File	lck; lock	Plain Text
NF00445	1.00	1	Moderate Risk	Windows Cabinet File	cab	Binary Archive Container File
NF00343	4	2	Moderate Risk	Microsoft OLE Compound Format	ole, ole2	Binary
NF00451	4.00	4	Moderate Risk	Windows Shortcut	lnk	Binary
NF00301	4.00	4	Moderate Risk	Windows Internet Shortcut	url	Binary
NF00146	5.00	5	Moderate Risk	Configuration File	conf, cnf, cfg, cf, ini	Plain Text
NF00277	7.00	5	Moderate Risk	Microsoft FrontPage Server Extension Configuration File	frontpg.ini; frontpg.cnf	Plain Text
NF00423	9	7	Moderate Risk	TAR Archive	tar	Binary Archive Container File
NF00242	9.00	9	Moderate Risk	Macromedia Flash 6	swf	Binary
NF00142	13.00	13	Moderate Risk	Cold Fusion Markup	cfm, cfc	Plain Text
NF00458	13	11	Moderate Risk	ZIP archive	zip	Binary Archive Container File
NF00243	13.00	13	Moderate Risk	Macromedia	swf	Binary

			Risk	Flash 7		
NF00450	14.00	14	Moderate Risk	Windows Setup Information File	inf	Plain Text
NF00440	16.00	16	Moderate Risk	Windows Autorun Configuration File	autorun.inf	Plain Text
NF00217	17.00	17	Moderate Risk	JavaScript	js	Plain Text
NF00386	19.00	19	Moderate Risk	PostScript font	pfb	Binary
NF00204	20	20	Moderate Risk	GZIP	gz; tgz	Binary Archive Container File
NF00138	22.00	22	Moderate Risk	C# Code	cs	Plain Text
NF00144	24	16	Low Risk	Common Data Format Toolkit	cdf	Binary
NF00387	24.00	24	Low Risk	PowerShell	psd1	Plain Text
NF00389	25.00	25	Low Risk	Python code file	py	Plain Text
NF00358	27	27	Low Risk	PHP Script	php	Plain Text
NF00215	28.00	28	Low Risk	Java Archive File	jar	Binary Archive Container File
NF00216	31.00	31	Low Risk	Java Object Code	java	Binary
NF00227	32	32	Low Risk	Log File	log	Plain Text
NF00211	32.00	32	Low Risk	Information or Setup File	info	Plain Text
NF00190	45	45	Low Risk	eXtensible Style Language	xsl, xslt	Plain Text
NF00416	45	40	Low Risk	Structured Query Language	sql	Plain Text