

文化財デジタルデータ長期保存のためのファイル形式

高田祐一（奈良文化財研究所）

File Formats for the Long-term Digital Storage of Cultural Heritage Data

Takata Yuichi (Nara National Research Institute for Cultural Properties)

・ファイル形式／File format・テキスト／Text・ドキュメント／Document
・ラスタ画像／Raster images・ベクタ画像／Vector images・3Dデータ／3D data

1. 文化財デジタルデータの要件

平成16年度に公表された文化庁『行政目的で行う埋蔵文化財の調査についての標準（報告）』において、「デジタル技術は急速に進歩し普及しつつあり、それを導入した報告書のあり方についても、今後、検討する必要がある。」と提起された（文化庁2004）。平成22年刊行の『発掘調査のてびき』「整理・報告書編」では、発掘調査報告書の3要件として、以下を求めている（文化庁2013）。

- ①将来にわたって保存されること。[保存性]
- ②相応の精度を有すること。[精度]
- ③公開・活用のための形態・方法が適切であること。[利活用]

てびきでは、この報告書3要件（保存性・精度・利活用）の観点からデジタルデータを「記録媒体自体の変化のほか、媒体の規格変更や製造中止など、いくつかの問題が指摘される」と保存性に関する懸念を示した。平成29年8月公表の『埋蔵文化財保護行政におけるデジタル技術の導入について2』（報告）においては、報告書3要件を充足する印刷物の形態こそが報告書としている（文化庁2017）。報告書PDFについて、高精度PDFは「印刷物の発掘調査報告書の、いわば「版下」に相当するデータであり、印刷物の発掘調査報告書そのものや掲載した図面や写真のバックアップ」で、「当該発掘調査報告書を作成した組織が保管する必要がある」と位置付け

ている（文化庁2017）。であるならば、まずはオリジナルの図面や写真を組織が保存したうえで、「図面や写真のバックアップ」を高精度PDFが担うという整理になる。

他方、このPDF形式については、作成フローや技術的見地からすれば、PDF生成のもととなるオリジナルファイルとPDFと一緒に保存することが推奨されている。結局のところ、行政的位置付けおよび技術的見地からすれば、オリジナルの図面・写真データと一緒にPDFを保存することが必要となる。

近年、図面や写真についてはデジタルデータが多くなっている。デジタルデータの長期保存を実現するには、デジタルデータに関するリテラシーが必要となる。デジタルデータの寿命に影響するものは、記録媒体とファイル形式の問題がある。記録媒体について、てびきが示す「記録媒体自体の変化のほか、媒体の規格変更や製造中止など」という懸念には、本誌別稿 高田祐一「文化財デジタルデータ長期保管の実務」・高瀬史則「デジタルデータ長期保存における記録メディアの選択」が参考となる。本稿では、ファイル形式の問題について取り扱う。

データの長期保存によって、将来にわたってデジタルデータの再利用性を高めることができれば、調査成果の情報発信強化・当該組織の業務効率化・調査研究の高次化を図ることができる。

2. ファイル形式選択の観点

イギリスの Archeology Data Service とアメリカの Digital Antiquity が協力して作成した「Guides to Good Practice」では、データ作成と保存について考え方が整理されている（Archeology Data Service / Digital Antiquity）。長期保存を可能にするファイル形式を検討する観点は3つである。

- ①仕様がオープン・標準化されていること。仕様がオープンであれば、複数のアプリケーションで使えるため、持続可能性が高まる。
- ②プレーンテキストファイルであること。ASCII プレーンテキストまたは XML であれば、人間にとって可読性が高い。それは当該ソフトウェアが使用不可となった場合に移行の可能性が高まる。
- ③可逆圧縮であること。不可逆圧縮の場合、データが失われ「品質低下」を引き起こす可能性がある。

3. ファイル形式の一覧

デジタルデータの長期保存に参考となるよう

「Guides to Good Practice」に掲載されている各ファイル形式を著者が要約した。2019年時点日本の状況を踏まえ、著者の責任において一部加筆修正を行った。

なお、ファイル形式の一覧の作成において、野口淳、堀木真美子、中村一郎、山口欧志各氏の助言を賜った。

【註】

文化庁 2004『行政目的で行う埋蔵文化財の調査についての標準（報告）』埋蔵文化財発掘調査体制等の整備充実に関する調査研究委員会

文化庁 2013『発掘調査のてびき 各種遺跡調査編』（報告）』

文化庁 2017『埋蔵文化財保護行政におけるデジタル技術の導入について 2』（報告）埋蔵文化財発掘調査体制等の整備充実に関する調査研究委員会

Archeology Data Service / Digital Antiquity 「Guides to Good Practice」(<https://guides.archaeologydata.service.ac.uk/g2gpwiki/>) 2019年12月2日参照

表1 テキストとドキュメントに関するファイルフォーマット

フォーマット	属性・説明	標準規格	データ保存に関する推奨事項
.doc	Microsoft 社の Word 用ファイル。バイナリ形式。Word1.0～6.0・95・97～2003で使用		データ長期保存に不適合。
.docx	Microsoft 社が作成した Office Open XML (OOXML) 形式。Microsoft 社 Office 2007 でリリース。独自仕様の OOXML を開発。XML の文書ファイルと関連ファイルを圧縮して構成している。Microsoft Office はオフィスアプリケーションのデファクトスタンダードとなっている。	ECMA-376 および ISO/IEC 29500-1:2008	データ長期保存に適合。非圧縮で保存する必要がある。埋め込みの画像などのコンテンツは、それぞれ適切な形式で個別に保存する必要がある。
.rtf	RTF (Rich Text Format)。Microsoft 社が開発。人間が読めるプレーンテキスト形式であるため保存に適する。ただし互換性や .doc・.pdf・.odt よりもファイルサイズが大きくなる問題がある。		データ長期保存に適合。ただし .docx や .odt の方がコンパクトで互換性があるため、.docx や .odt の方が良い。
.odt	ODT (Open Document Text)。OpenDocument 形式。 .docx 形式と同様に XML の文書ファイルと関連ファイルを圧縮して構成している。	ISO/IEC 26300:2006	データ長期保存に適合。非圧縮で保存する必要がある。埋め込みの画像などのコンテンツは、それぞれ適切な形式で個別に保存する必要がある。
.sxw	OpenOffice/StarOffice のバージョン 1.0～2.0 で使用された OpenOffice.org XML 形式。 OpenDocument 形式に変更された。		データ長期保存に適合。ただし ODT 形式の方が良い。
.pdf	PDF (Portable Document Format)。Adobe 社が開発した形式。PDF reader は無償で利用可能であり、広く普及している。ラスタ画像・ベクタ画像・javascript・3D データを埋め込んだファイルを作成できる。	ISO 32000-2 : 2017(PDF2.0)	オープン標準であるが、理想的なデータ保存の形式ではない。DPC の 'Preserving the Data Explosion: Using PDF' (Fanning 2008) において、PDF を保存に使用する課題が報告されている。
.pdf/a	長期保存のオープン標準を提供するために PDF 形式から開発された。PDF/A-1 : PDF1.4 に基づいて開発された。ファイルへの音声・動画・javascript などの埋め込みや暗号化を禁止する必要がある。PDF/A-2 : PDF1.7 (ISO32000-1) をベースに開発された。PDF/A 準拠のファイルを埋め込み可能となった。PDF/A-3 : ISO32000-1 に基づいている。PDF/A 非準拠のファイルも埋め込み可能となった。	ISO 19005-1 (PDF/A-1) ISO 19005-2 (PDF/A-2) ISO 19005-3 (PDF/A-3)	長期保存の形式として広く普及している。他のファイル形式からファイルを作成する場合、もとのファイルと PDF/A ファイルを一緒に保管することを推奨する。PDF/A 準拠のファイルを埋め込む場合は PDF/A-2、PDF/A 非準拠のファイルを埋め込む場合は PDF/A-3 を使用するのがよい。
.wpd	WordPerfect 用ファイル。バイナリ形式。WordPerfect の日本語版は 2019 年現在出荷されていない。		データ長期保存に不適合。
.txt	プレーンテキスト形式。多くのプラットフォームおよびソフトウェアと互換性がある。		データ長期保存に適合。ただしシンプルな形式に限定される。

表2 ラスタ画像に関するファイルフォーマット

フォーマット	属性・説明	標準規格	データ保存に関する推奨事項
.tif .tiff	TIFF Version 6。デファクトスタンダードであり、アーカイブ用の画像を保存するために非圧縮形式で広く使用されている。撮影情報を埋め込む EXIF メタデータとジオレファレンス用の GeoTIFF メタデータをサポートしている。		TIFF (Version 6) は非圧縮のデータ長期保存に適合する。ただし LZW 圧縮アルゴリズムを使用している TIFF ファイルは、長期保存に不適切である。
.png	PNG (Portable Network Graphics)。W3C が開発。GIF 形式の代わりに開発された。EXIF メタデータをサポートしていない。インターネット利用を目的としているため、RGB 以外の色空間をサポートしていない。	ISO/IEC 15948:2004。 RFC2083。 W3C 勧告。	可逆圧縮が必要な場合、GIF 形式より PNG 形式を使用した方がよい。JPG 形式より利点があるものの、写真画像で PNG 形式を使用するのは非推奨である。
.jpg .jpeg	Joint Photographic Expert Group が開発した。写真画像用に開発された。効率的な非可逆圧縮が特徴。	ISO/IEC 10918	非圧縮のオリジナルデータの長期保存には不適合。TIFF 形式および PNG 形式よりも大幅にファイルサイズを圧縮できるものの非可逆圧縮のため。ただし規格の公開と普及の観点から圧縮データの長期保存には適合。
.jp2 .jpx	JPEG 2000 は JPG 形式の替わりとなる形式。JPG 形式に比べ普及していない。	ISO/IEC 15444。ITU-T Rec.T.800	可逆圧縮と拡張性の点から、保存に適する形式か調査されている。

フォーマット	属性・説明	標準規格	データ保存に関する推奨事項
.gif	GIF (Graphics Interchange Format)。CompuServe 社が開発した。Webにおいて静止画像とアニメーション画像で広く使用されている。可逆圧縮を提供している。		長期保存には TIFF 形式を推奨。圧縮が必要な場合は PNG 形式を推奨。
.bmp	BMP (Bit-Mapped Graphics)。Microsoft 社が開発した。Windows アプリケーションで標準的に使用されている。		データ長期保存に不適合。
.psd	Adobe 社の Photoshop 用ファイル形式。画像編集の業界標準。		データ長期保存に不適合。サードパーティーがサポートできない。
.cpt	CPT (Corel Photo-Paint image)。Corel 社が開発した画像編集ソフトのファイル。Corel Photo-Paint は、Adobe 社 Photoshop の競合製品。		データ長期保存に不適合。サードパーティーがサポートできない。
.dng	DNG (Digital Negative)。Adobe 社が開発。オープンフォーマット。RAW ファイル形式。Adobe 社は、RAW ファイルから DNG に変換するソフトを公開している。仕様をオープンソースで公開している。		データ長期保存に不適合。
raw (various extensions)	各メーカー固有の形式である。		データ長期保存に不適合。TIFF を推奨する。

表3 ベクタ画像に関するファイルフォーマット

フォーマット	属性・説明	標準規格	データ保存の推奨
.ai	Adobe 社の Illustrator 用ファイル形式。他のアプリケーションで使用できない。		データ長期保存に不適合。
.cdr	CDR (CorelDraw)。Corel 社が開発したベクタ画像編集ソフトのファイル。ラスターデータとベクターを保持できる。		データ長期保存に不適合。
.ps	PS (PostScript)。Adobe 社によって開発された。		PDF/A の方が長期保存に適合。
.eps .epsf .epsi	EPS (Encapsulated PostScript)。		データ長期保存に不適合。
.pdf .pdf/a .pdf/e .pdf/x	PDF (Portable Document Format)。Adobe 社によって開発された。pdf/a: 長期保存目的。pdf/e: 技術文書の交換目的。pdf/x: 印刷目的。	ISO 24517-1 (pdf/e) ISO15930 (pdf/x)	—
.svg	SVG (Scalable Vector Graphics)。W3Cが開発。XMLベースのオープン標準。Javascriptと一緒によく Web 上で使用される。	W3C 勧告	データ長期保存に適合。
.cgm	CGM (Computer Graphics Metafile)。CGM のサブセットである WebCGM は、Web 用に W3C が開発した。	ISO / IEC 8632	SVG を使う方が良い。
.wmf .emf	WMF (Windows Metafile)。Microsoft 社が開発。GDI (Graphics Device Interface) が画像描画等の処理を行う。		データ長期保存に不適合。
.wpg	WPG (WordPerfect Graphics Metafile)。		データ長期保存に不適合。
.pict .pic .pct	Apple 社が開発した。ビットマップ形式とベクタ画像形式を保存できる。		データ長期保存に不適合。
.swf .fla .swd .flv .swc .swt .flp	Web 配信用のアニメーション形式。Flash は Macromedia 社が開発したが、Macromedia 社を Adobe 社が買収した。		データ長期保存に不適合。
.af	Macromedia 社 Freehand 用。Macromedia 社 Freehand の開発とサポートは中止されている。		データ長期保存に不適合。
.drw .dsf	Micrografx 社の Designer 用のファイル形式。Micrografx 社は Corel 社に買収された。		データ長期保存に不適合。

表4 動画に関するファイルフォーマット

フォーマット	属性・説明	標準規格	データ保存に関する推奨事項
.mpg .mpeg	MPEG 1。オープン標準となっている。VHS テープ相当の音声/ビデオ再生を提供する。MPEG-1 Audio Layer III は、MP3に相当する。	ISO/IEC 11172	データ長期保存に適合。
.mpg .mpeg	MPEG 2。オープン標準となっている。DVD および TV 用。	ISO/IEC 13818	データ長期保存に適合。
.mp4	MPEG 4。オープン標準となっている。	ISO/IEC 14496-14:2003	データ長期保存に適合。
.divx .avi	DivX。ビデオコーデック。インターネットで映画配信する際に使用されている。DivX コーデックは無料でダウンロードできるが、エンコードのためにはライセンスが必要。		データ長期保存に不適合。
.avi .xvid	Xvid。MPEG-4 をベースにした GNU 一般公的使用許諾		データ長期保存に不適合。
.mj2 .mjp2	Motion JPEG 2000。可逆圧縮と不可逆圧縮の両方でビデオを保存できる。	ISO/IEC 15444-3	データ長期保存に適合。
.mkv	Matroska。マルチメディアコンテナフォーマット。		データ長期保存に適合。
.flv	Flash Video。YouTube 等で活用されている。		活用には適している。
.avi	Audio Video Interleave。Microsoft 社が開発した形式。		活用には適している。
.mov	Quicktime。Apple 社が開発した形式。		活用には適している。

表5 3Dデータに関するファイルフォーマット

フォーマット	属性・説明	標準規格	データ保存に関する推奨事項
.x3d	Web3D コンソーシアムが開発した。Web3D コンソーシアムは、かつて VRML (Virtual Reality Modeling Language) を開発したが、いくつかの課題から X3D を開発し継承している。	ISO/IEC 19775	データ長期保存に適合。
.dae	COLLADA (Collaborative design activity) は、XML ベースの交換用ファイルフォーマット。Khronos Group コンソーシアムが開発した。	ISO / PAS 17506 : 2012	データ長期保存に適合。
glTF	glTF (GL Transmission Format)。Khronos Group が開発した。Khronos Group は大手ハード・ソフトウェア企業のコンソーシアムであり、ロイヤリティフリーの標準規格の策定を目的としている。glTF は、'JPEG of 3D' (3D における JPEG) ともいわれる。Web3D に活用される。glTF 2.0 の仕様は github で公開されている。		—
.obj (also includes optional .mtl and .jpg files)	Wavefront Technologies 社が開発した。ファイル形式が公開されており、ユーザーコミュニティによってサポートされている。Obj ファイル (ascii もしくは binary)、mtl ファイル (material/texture)、image ファイル (texture) によって構成される。		ワイヤーフレーム (wire frame) またはテクスチャモデルのデータ保存に適合している。保存の際は ASCII 形式を推奨する。
.ply	3D スキャンデータ用に Stanford 大学が開発した。ASCII もしくはバイナリのシンプルな形式。		データ長期保存に適合。ASCII で保存する必要がある。
.vrmf .wr .wrz	VRML (Virtual Reality Modelling Language) X3D の前身。最新バージョンは 1997 年に VRML97 として公開された。	ISO/IEC 14772	データ長期保存に適合。ただし現在は X3D に置き換わっている。
.u3d	U3D (Universal 3D)。データ交換用の圧縮形式。3D Industry Forum が開発した。2005 年に Ecma International (ECMA) が標準化した。PDF ファイルに U3D ファイルを埋め込むことができる。	ECMA-363	データ長期保存に不適合。

フォーマット	属性・説明	標準規格	データ保存に関する推奨事項
.stl	STL (Stereolithography or Standard Tessellation Language)。3D Systems 社が開発。3D プリンターおよびデジタルファブリケーション (Digital Fabrication。工作機械などによって素材から切り出し成形する技術) で普及している。ASCII バージョンとバイナリバージョンがある。ASCII バージョンではテクスチャがなく 3D モデルジオメトリのみ保存する。バイナリバージョンは、拡張機能によって色情報を保存できる。		ASCII の基本的なデータセットの保存に適している。サーフェイスのみ出力するため、フォトグラメトリー (photogrammetry) には適していない。
.dxf	DXF (Drawing Exchange Format)。Autodesk 社の CAD ソフトである AutoCAD で使用するファイル形式。CAD のデータ互換形式。ASCII バージョンとバイナリバージョンがある。CAD ソフトウェアで作成された 3 D コンテンツにのみ使用した方がよい。		CAD のデータセットの保存にのみ保存に適している。
.fbx	Autodesk 社が保有する 3D データ交換用のファイル形式。3 D ソフトウェアである Maya や 3DS Max などのソフトウェア間のデータ交換をサポートすることを目的としている形式。		データ長期保存に不適合。
.3ds .max	Autodesk 社の 3DS Max で使用するファイル形式。バイナリ形式。		データ長期保存に不適合。
.skp	Google 社の SketchUp で使用するファイル形式。		データ長期保存に不適合。
.blend	Blender Foundation が開発した。Blender で使用するファイル形式。バイナリ形式。		データ長期保存に不適合。
.prc	PRC (Product Representation Compact)。Association for Information and Image Management が開発した。3 D 設計データを高度に圧縮する形式。CAD・CAD・CAM にて使用されている。3 D データを PDF に埋め込むことができる。	ISO 14739-1 : 2014	データ長期保存に不適合。
.pdf	PDF (Portable Document Format)。Adobe 社が開発した。U3D・PRC 形式によって PDF に 3D データを埋め込む。特殊なソフトウェアを使わずに、データを表示できるため便利である。しかし、データを取り出せないため、データが 'dead end' (行き止まり) となるファイル形式である。	ISO 32000-2 : 2017 (PDF2.0)	データ長期保存に不適合。
.nxs	Nexus のファイル形式。CNR-ISTI が開発した。オープンソース。大きな 3 D メッシュを Web ベースで視覚化することができる。		データ長期保存に不適合。