

## 文書とデジタルテキスト

### セクション1. 文書とテキスト概説

ドキュメントやテキストファイルは、考古学的調査の結果として作成される最も一般的なファイルタイプです。フィールド調査、机上での評価、放射性炭素年代測定など、どのような作業を行っているかに関わらず、圧倒的に多くのプロジェクトでは、テキスト文書の形で何らかの最終報告書が作成されています。また、報告書だけでなく、物理探査やデータベースの文書化など、プロジェクトの他の要素に関連するプロセスやメタデータを記録するためにテキスト文書が作成されることもよくあります。

このガイドでは、考古学プロジェクトで一般的に作成されるバイナリおよびプレーンテキスト文書の主な種類を概説します。一般的なファイルタイプやアーカイブフォーマットの説明に加え、テキスト文書の重要な特性としてどのような要素を考慮すべきか、作成方法の違いがどのように影響するか、これらの特性を確実に残すためにはどのようなアーカイブ戦略を採用すべきかについても説明します。

#### 1.1 文書とテキストとは？

簡単に言えば、テキスト文書の大部分は従来の出版物のデジタル版であり、そのサイズと複雑さは、非常にシンプルなレポートや短い論文から、論文や書籍のような充実した文書までさまざまです。これらのファイルは、主に構造化されたテキスト（文、段落、ページ、章）で構成されていますが、画像、図、表データなどの他の要素を含むこともあります。

デジタルテキストはさまざまな方法で作成できますが、ほとんどの場合、Microsoft Word や OpenOffice などのワープロソフトではじめから作成します。しかし、Microsoft Word 2007 や OpenOffice などの最近のパッケージでは、.docx (Office Open XML [1] 形式の一部) や .odt (OpenDocument [2] 形式の一部) など、人間可読な xml ベースの形式や規格への移行が顕著になっています。ドキュメントがもともと作成されたフォーマットに加えて、最終版のテキストドキュメントの多くは、共通の交換フォーマット、特に Adobe の Portable Document Format [3] (PDF) で保存され、公開されることがあります。このフォーマットは、さまざまなプラットフォーム間でドキュメントのフォーマットと構造を一貫性を保つことができる一方で、利用者による編集可能性を排除することができます。

ワープロソフトで作成された文書に加えて、デジタル化プロセスの結果、広範なテキスト文書が作成されます。学術雑誌のデジタル化は、通常、デジタル化されていないコレクションの保存や公開のために行われますが、ワープロ以外で作成されたデジタルテキストの最大の供給源となることも多くあります。このプロセスは、通常、印刷物のページのデジタル画像から、光学式文字認識 (OCR) を使って画像を「実際の」(編集可能、検索可能など) テキストに変換するために

処理されます。画像や図表を含む最終テキストは、主に PDF ファイル形式で保存されますが、特に動的なオンライン配信が必要な場合は、xml ベースの形式が使用されることもあります。

一般的なワープロ形式や PDF ファイル以外にも、テキストはプレーンテキストや、SGML、HTML、XML などのマークアップされた形式で存在する場合があります。これらのフォーマットについては、Oxford Text Archive Preservation Manuals [4] で詳しく説明されていますが、以下では他のフォーマットと一緒に簡単に説明します。

#### 1.2 現在の問題点と懸念事項

##### ファイルフォーマット

以下では、特定のファイルフォーマットの課題について説明しますが、一般的には、アーカイブにとって懸念される、テキスト文書全体に関連する2つの課題があります。1つ目は、多くのファイルフォーマットに関連して見られるように、ワープロパッケージで使用されるフォーマットが継続的に変化していることです。今はないソフトウェア(例: Wordstar)で作成されたファイルを受け取る可能性以外にも、現在よく使われているワープロパッケージのフォーマットも継続的に開発・強化されるため、古いバージョンのファイルと現行バージョンのソフトウェアの間に互換性がないことがよくあります。前述したように、.docx や .odt といった XML ベースのオープンスタンダードなフォーマットへの移行は、これらのフォーマットを標準化し、異なるソフトウェアパッケージが各ソフトのネイティブではないフォーマットを読めるようにするための試みです。同様の問題は PDF フォーマットでも明らかになっていますが、ここでもオープンスタンダード (PDF/A [5]) に向けた最近の動きが、長期的なアクセス確保の課題に対処する試みとなっています。

##### 組み込みオブジェクト

ファイルフォーマットそのものに加えて、テキスト文書にコンテンツを埋め込むことができるかどうか、また、そのようなコンテンツを元の文書フォーマット内で長期的に保存することができるかどうかについても、広く懸念されています。最も一般的な埋め込みコンテンツは画像ですが、特に Microsoft Word や PDF などの特定のフォーマットでは、スプレッドシートやビデオなどのより複雑なコンテンツを、テキスト文書と一緒に保存することができ、多くの場合、別個に保存するかアーカイブするフォーマットになっています。一般的には、埋め込みだけでなく、コンテンツを別個に保存し、アーカイブすることが推奨されています。そうすることで、コンテンツのオリジナルの品質(画像の解像度など)を維持し、テキストコンテンツとは別のアーカイブ戦略に対応することができます。

[1] <http://www.ecma-international.org/publications/>

- standards/Ecma-376.htm  
 [2] <http://xml.openoffice.org/>  
 [3] <http://www.adobe.com/products/acrobat/adobepdf.html>  
 [4] <http://ota.ahds.ac.uk/documents/index.xml>  
 [5] [http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=50655](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=50655)

- ・文書に外部リンクや動的コンテンツを含めないようにしてください。
- ・文書の最終版をPDFに保存する場合は、元のフォーマット（Word.doc、OpenOffice.odtなど）を別途保存するようにしてください。

さらに、PDFに保存する場合は、そのファイルが可能ならばPDF/AまたはPDF 1.4の規格に準拠していることを確認してください。確認事項には、ファイルが保護されていないこと、javascriptやビデオ、またはオーディオコンテンツ、圧縮が含まれていないこと、すべてのフォントと画像が正しく埋め込まれていること、ファイルに正しく「タグ」が付けられていることが含まれます。

より一般的な検討事項としては、作成者は可能な限り、文書の内容が完全に自明なものであることを保証する必要があります。

## セクション2. 文書とテキストの作成

### 2.1 一般的な検討事項

テキストファイルや文書は、一般的に大規模なアーカイブの一部であり、それ自体が大規模なプロジェクトの計画や要件を文書化したものであることが多いです。ワープロ文書を作成する際には、プロジェクト全体の計画自体はほとんど必要ありませんが、作成に際して、いくつかの注意点があります。

- ・前述したように、画像や表計算ソフトなどの埋め込みコンテンツは、文書とは別に保存するようにします。これらのコンテンツを元のフォーマットやアーカイブフォーマットで安全に保存できるだけでなく、これらのファイルを別個に保存することで、元のサイズや解像度のまま保存することができます。

### 2.2 ファイルのフォーマット

これらの一般的な考慮事項に加えて、それぞれのアプリケーションやファイルフォーマット（多くは同じアプリケーションから作成されます）には、別の利点もあります。以下の表では、テキスト文書の作成に使用される一般的なフォーマットと、それに関連するアプリケーション、および長期保存のための潜在的な利用を紹介しています。

フォーマット	プロパティ／技術	詳細	推奨
.doc	Microsoft Wordの商用(プロプライエタリ)バイナリフォーマット	一般的なフォーマットで、MS Wordの1.0から6.0、95、97-2003のすべてのバージョンでデフォルトとなっている。Wordでの使用に加えて、OpenOfficeでも読むことができ、簡単に.pdfに変換できる。これまでWordのバージョン間での後方互換性はかなり高かったが、最近Microsoft Office Word 2003にサービスパック3が追加されたことにより、バージョン2.0以前のサポートが終了した。2008年からは、マイクロソフトのバイナリファイルフォーマットの仕様書が、マイクロソフトのウェブサイトや大英図書館で公開されている[1]。	アーカイブでの保存や公開には適さないが、広く利用されているフォーマットであり保存には便利である。
.docx	マイクロソフト社が作成したOffice Open XML (OOXML) フォーマットの一部。ECMA (ECMA-376 [2]) および ISO (ISO/IEC 29500-1:2008 [3]) の標準規格。	Microsoft社の比較的新しいフォーマットで、Office 2007とともにリリースされた。Microsoft社は、以前のMS Wordファイルフォーマットとの後方互換性を高めるために、既存のODF国際規格(ISO/IEC 26300:2006、下記ODT参照)を使用せず、独自の仕様(OOXML)を開発した。OOXMLは、人間が読めるXMLファイルとその他のコンテンツを1つのZIPファイルにまとめたものである。	埋め込みコンテンツは別個に保存する必要があるが、登録、公開、保存に適している。最終ファイルは基本的に.zip圧縮されたアーカイブであり、そのファイルを再度圧縮しないで保存するべきである。
.rtf	RTF(Rich Text Format)は、Microsoft社が開発したタグ付きテキスト形式。	人間が読めるプレーンテキスト形式であるため、公表と保存両方に適しているが、異なるワープロパッケージでファイルを開いた場合、版面表示(テキストボックスや表など)に関	保管や保存には適しているが、.docxや.odtなどの新しいフォーマットは、よりコンパクトで互換性のあるフォーマットなので、.rtfよりも優先して使用すべきであろう。

フォーマット	プロパティ/技術	詳細	推奨
		する互換性の問題がある。また、.rtf ファイルのファイルサイズは、同等の .doc、.pdf、.odt ファイルよりもはるかに大きくなる。	
.odt	Open Document Text は、XML ベースのオフィス文書フォーマットの ISO 規格 (ISO/IEC 26300:2006 [4]) である OpenDocument Format の一部。	.docx フォーマットと同様に、.odt ファイルは基本的に、スタイル、テキスト (XML として)、埋め込みコンテンツ (画像など) の各ファイルを含む圧縮された ZIP ファイルである。	オープンな XML ベースのフォーマットである ODT は、登録にも保存にも適しているが、後者の場合、ファイルは圧縮されていない状態で保存する必要がある。また、文書に画像やその他のコンテンツが含まれている場合、これらは適切なフォーマットで個別に保存すべきであろう。
.sxw	OpenOffice/StarOffice でバージョン 1.0 からバージョン 2.0 まで使用された OpenOffice.org の XML フォーマットの一部である。OpenDocument Format に引き継がれている。	このフォーマットは ODF に取って代わられたが、ODF と同様の構造 (例:zip 形式の XML ファイル) になっており、OpenOffice.org 2.0 で読むことができる。	保存に適しているが、可能な限り ODT を使用するべきである。
.pdf	PDF (Portable Document Format) は、Adobe 社が作成した、オープンな標準交換形式である。このフォーマットは誕生以来何度も進化しており、PDF/A を筆頭にいくつかのバリエーションがある。	PDF は、クロスプラットフォームでの文書交換を目的としたフォーマット。PDF は商用 (プロプライエタリ) のバイナリフォーマットだが、文書の版面表示を保持するように設計されており、リーダーは無料で広く利用できるため、公開に非常に適している。PDF 文書は、標準的なテキストに加えて、ラスターおよびベクター画像、JavaScript、さらには 3D データなど、さまざまな埋め込みメディアやリンクメディアを保存することができる。また、PDF ファイルは、編集や印刷を制限するセキュリティが設定されている場合がある。	オープンな標準形式ではあるが、PDF は商用 (プロプライエタリ) バイナリフォーマットであり、保管や保存には適していない。あるレポート「Preserving the Data Explosion: Using PDF」(Fanning 2008) は、DPC が作成したレポートで、保存のために PDF を使用する際の問題点を多数取り上げている。ほとんどの場合、PDF ファイルは他のフォーマット (例:.doc) から作成されており、この元のソースを保存、移行することが望ましい。これが不可能な場合は、PDF/A への移行がしばしば最良のルートとなるが、埋め込み要素やセキュリティが移行を阻害しないか確認する必要がある。
.pdf/a	PDF/A フォーマットは、Adobe 社が作成した PDF フォーマットをベースに、長期的なアーカイブのためのオープンな標準形式の提供を目的としている。	PDF バージョン 1.4 をベースにした PDF/A は、長期アーカイブのための堅牢でオープンな標準形式の提供することを目的としており、ISO 規格として公開されている (ISO 19005-1:2005 [5])。PDF/A フォーマットでは、長期的に信頼性の高いアーカイブを行うために、特定の文書要素 (フォントや色空間) を指定したり、ファイルに埋め込むことを要求する一方で、その他の要素 (javascript、オーディオやビデオ、暗号化) を禁止している。PDF/A フォーマットには 2 つの準拠レベルがあり、基本レベル (PDF/A1-b) では、文書を確実に視覚的に再現できることを保証し、PDF/A1-a では、タグ付け / 文書構造を含めて構築している。PDF/A フォーマットの有効的な概要は、ミシガン大学の Bentley Historical Library [6] によって書かれている。	PDF/A は、長期保存に適したフォーマットとして広く受け入れられており (Library of Congress [7] など)、デジタル保存コミュニティでも広くレビューされている (Fanning 2008 など)。しかし、標準的な PDF ファイルと同様に、他のフォーマット (.doc や .odt など) からファイルが作成された場合は、それらを保持し、PDF/A ファイルと一緒に保存することが推奨される。
.wpd	WordPerfect のバイナリ形式の独自フォーマット。	WordPerfect は、1980 年代初頭にリリースされた Microsoft Word の台頭により普及率が著しく低下した。.wpd (初期のバージョン) では .wp	保存や公開には適していない。WordPerfect のネイティブファイルだが、.wpd ファイルは Microsoft Office Word や OpenOffice でも読む

フォーマット	プロパティ/技術	詳細	推奨
		または .wp5 など)が主なフォーマットだが、最近のバージョンでは、幅広いインポートおよびエクスポートオプションをサポートしている。	ことができる。WordPerfect Office の最新バージョンは、ODF と OOXML ファイルの両方のエクスポート(およびインポート)をサポートしているので、後者の XML ベースのオープンスタンダードを使用することを勧める。
.txt/プレーンテキストファイル	シンプルなプレーンテキスト文書である。プレーンテキストは、マークアップテキスト(後述)の基礎にもなる。	プレーンテキストファイルは、「テキスト情報の最小公倍数のエンコーディング」(Wynne & Yeates 2004)であり、多くのプラットフォームやソフトウェアとの互換性がある。しかし、版面表示をほとんどサポートしていないため、非常にシンプルな文書にしか使用できない。すべてのプレーンテキストファイルには、エンコーディング(通常、US-ASCII または UNICODE)を指定する必要がある。プレーンテキストファイルについては、AHDS Preservation Manual on Plain Text [8] で詳しく説明している。	データの移動(インジェスト)、保存、公開に適しているが、極めて単純なファイルに限られる。

### マークアップテキストのフォーマット

マークアップテキストについては、AHDS Guide 'Creating and Documenting Electronic Texts' (Morrison et al 2001) および AHDS Preservation Handbook 'A Guide of Good Practice' (Morrison & Wynne 2005) で詳しく説明されて

います。報告書形式の文書ではあまり使われないものですが (Web ページでは HTML が、データ交換では XML がよく使われる)、一般的なマークアップテキスト形式を以下に簡単に紹介します。

フォーマット	プロパティ/技術	詳細	推奨
.sgml	Standardised Generalised Markup Language (SGML) の略。認証された ISO 規格 (ISO 8879:1986 SGML [9]) によるメタ言語。	SGML は、HTML や XML などの他のマークアップ言語を定義するために使用されるメタ言語である。	文書が有効であるならば、保存や公開に適している。
.html/.xhtml	Hypertext Markup Language (HTML) は、SGML のサブセットとして開発されたプレーンテキストベースのマークアップ言語で、W3C によって管理されている。	HTML は、ウェブページの作成によく使われるマークアップ言語である。HTML ファイルのプレーンテキストコンテンツ(インラインまたはリンクされたスタイルシートを含む)以外に、ウェブサイトは一般的に、リンクされたさまざまなメディア(画像、動画、音声、文書など)で構成されており、これらは別個のオブジェクトとして扱われるべきものである。	保存や公開に適した文書は、有効な DTD と文字エンコーディングを遵守(および指定)する必要がある。CSS スタイルシートを使用する場合は、文書内で指定するか、別のファイルとして提供する必要がある。画像やその他のメディアは、他のガイドと同様に、個々のオブジェクトとして扱われるべきである。
.xml	Extensible Markup Language (XML) は、W3C によって作成されたプレーンテキストベースのオープンスタンダードである。	XML は、SGML のサブセットとして開発され、一般的にウェブやシステム間(データベースなど)のデータ交換に使用されている。	文書は、有効な DTD/スキーマと文字エンコーディングに準拠(および指定)する必要があるが、保存や公開に適している。

ライフサイクルの観点から見ると、ほとんどの場合、テキストファイルは作成中も同じフォーマットのままです。ただし、PDF フォーマットは例外で、最初から PDF で作成された文書はほとんどなく、大半はワープロファイル (Word や OpenOffice など) から作成され、オーサリングプロセスの最後に配布するために PDF フォーマットに保存されます。

- [1] <http://www.bl.uk/aboutus/stratpolprog/ccare/introduction/digital/formats/index.html>
- [2] <http://www.ecma-international.org/publications/standards/Ecma-376.htm>
- [3] [http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=51463](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=51463)
- [4] [http://www.iso.org/iso/catalogue\\_detail](http://www.iso.org/iso/catalogue_detail)

- htm?csnumber=43485
- [5] [http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=38920](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=38920)
- [6] <http://bentley.umich.edu/uarphome/bestprac/pdfafaqs.php>
- [7] <http://www.digitalpreservation.gov/formats/fdd/fdd000125.shtml>
- [8] <http://ota.ahds.ac.uk/documents/index.xml>
- [9] [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=16387](http://www.iso.org/iso/catalogue_detail.htm?csnumber=16387)

## セクション3. ドキュメントとテキストの保存

### 3.1 アーカイブする内容の決定

前述したように、文書はその作成期間中、ほとんどが単一のフォーマットとして維持されます。ただし、PDF ファイルは、公開目的でプロセスの最後に作成されることが多いでしょう。このような場合には、元の作業文書と PDF ファイルの両方を保持することが推奨されます。さらに、画像などの埋め込みコンテンツは、テキスト文書とは別に保存し、これらのファイルが別の適切な保存経路をたどることができるようにすることも推奨されます。

その他のより一般的な考慮事項としては、保持・保存されるファイルが文書の最終版であることを確認することが挙げられます。これには、単に作業用ドラフトを注意深く追跡することも含まれますが、「変更履歴」などの特定のワープロアプリケーションの機能をオフにすることや、コメントやメモを最終版から削除することも含まれます。

### 3.2 アーカイブする方法の決定

文書を長期的に保存するための最適なフォーマットを決定する際には、文書の重要な特性を保持すると同時に、さまざまな一般的なアプリケーションから容易にアクセスし、必要に応じて移行できるフォーマットを提供することを考慮すべきです。

#### 文書の重要な特性

一般的に、保存する必要のある文書やテキストの重要な特性は次のとおりです。

- ・ 文書内の単語とその順序
- ・ 文書の階層構造（例：異なるレベルの見出し）
- ・ 文書内の書式（例：太字や斜体のテキスト）
- ・ 文書のページ番号。これは、文書が出版済みまたは未出版の報告書や論文である場合に特に重要です。ユーザーが文書を引用参照する場合、正しいページに正しいテキストを保持することが重要です（特に、ファイルがその後に移行・変換される場合）。
- ・ 画像やデータテーブルなど、テキスト以外のあらゆるコンテンツ。これらのコンテンツは別個に保存するのが理

想的です。

以下は、一般的にあまり重要視されていないプロパティです。

- ・ フォントタイプとフォントサイズ（ページネーションやフォーマットが大幅に変更されない限り）。
- ・ 変更履歴機能。

重要なプロパティは、保存されている文書の正確な性質に応じて変わる可能性があります。すべてのファイルは、上記のうちどれが関連するかを判断するために、ケースごとに評価されるべきです。

#### ファイルフォーマット

保存、長期安定的な保管のためには、Microsoft の OOXML (.docx) や OpenOffice ODF (.odt) など、現在広く普及している標準化された XML 形式のいずれかで文書を保存・保管することが望ましいとされています。JISC Technology Watch Report, 'XML-based Office Document Standards' (Ditch 2007) では、これらの仕様を詳細に検討、比較しています。両フォーマットの主な利点は、国際的にオープンな標準形式として認められていることと、(バイナリファイルではなく) テキストベースであるため、人間可読なことです。両フォーマットは、それぞれのネイティブアプリケーションと、Google Docs などの多くのサードパーティアプリケーションによって相互に受け入れられています。両フォーマットは、それぞれのファイルを構成する個別のコンポーネントを格納するために ZIP 形式のアーカイブを利用するという点でも類似しています。

しかし、ODF は他の既存のオープンスタンダード（例えば SVG）をより利用しています。また、ODF の仕様は、OOXML よりもはるかに短く（そして潜在的にはより完全で）、このフォーマットに対するサードパーティのサポートがより容易になる可能性があります。しかし、Microsoft 社の OOXML は、MS Word の旧バージョンからの移行を良くサポートしています（後方互換性は設計上の主要な目標の1つでした）が、MS Word から ODF に変換する際には、版面表示は完全に正確には再現されません。

これらの XML ベースのフォーマットに加えて、PDF/A も潜在的な保存フォーマットとして考慮する必要がありますが、これは主に元の文書が PDF フォーマットでしか存在しない場合に限られます。PDF/A はバイナリ形式ですが、自由に利用できるリーダーを備えたオープンスタンダードであり、サードパーティのサポートも拡大しています。PDF 文書から他のフォーマットにコンテンツを抽出したり移行したりすることは問題ですが、PDF/A は、バイナリではあるものの、認識されたオープンスタンダードフォーマットで既存の PDF コンテンツを正確に保存するための効果的な手段を提供します。

保存形式	要件
.docx	埋め込みコンテンツは別途保存する必要があるが、登録、公開、保存に適している。最終ファイルは基本的に ZIP 圧縮されたアーカイブなので、ファイル自体は圧縮せずに保存すべきである。

保存形式	要件
.odt	登録と保存の両方に適していますが、ファイル自体は圧縮せずに保存すべきである。また、ドキュメントに画像やその他のコンテンツが含まれている場合、それらは適切な保存フォーマットで個別に保存することが理想的である。
.pdf/a	長期保存に適している。他のフォーマット(例:.docまたは.odt)からファイルを作成した場合は、それらを保持し、PDF/Aファイルと一緒に保存することを推奨する。
.txt / プレーンテキストファイル	データの移動・移行(インGEST)、保存、公開に適しているが、極めて単純なファイルに限る。
.sgml	保存および公開に適しているが、ドキュメントが有効であるか確認する必要がある。
.html / .xhtml	保存および配布に適しているが、文書が有効な DTD および文字エンコーディングに準拠(および指定)している必要がある。CSS スタイルシートを使用する場合は、文書内で指定するか、別のファイルとして提供する必要がある。画像やその他のメディアは、他のガイドと同様に、個々のオブジェクトとして扱われるべきである。

### 3.3 メタデータと文書

多くの文書は自己説明的であることが多いですが、リソースの探索や証明のために、個別または一連の文書に対していくつかの基本的なメタデータを記録することが推奨されます。メタデータの標準規格には、さまざまなコミュニティから生まれた長い歴史を持つものが数多くあり(例:図書館セクターにおける MARC [1])、それぞれに長所と短所があります。以下に説明するメタデータセットは、(textMD [2])

のようなフォーマットで記録される技術的なファイルのメタデータではなく)リソースの探索と所在の記録という文脈において、最低限のことを記録し、他の既存の標準(特に Dublin Core Metadata Element Set [3] の15の要素)と簡単に対応できるようにすることを目的としています。特に文書については、プロジェクトレベルのメタデータと、プロジェクトの結果として作成された文書(プロジェクトの報告書など)を記述するメタデータとが重複する傾向があります。

要素	内容
標題	文書の標題
概要	文書の短い説明
出版年	出版年
刊行状態	シリーズ/ジャーナルのタイトル、号数、巻数などと、出版されたバージョンの完全な所在地の詳細。開始と終了のページ番号または総ページ数も記録すべきである。
出版社	出版社の名前と所在地
ISBN	ISBN[4]番号(該当する場合)。
DOI	Digital Object Identifier (DOI) [5] (該当する場合)。

上記に加えて、その他の多数の要素を文書レベルで記録する必要があるかもしれません。前述のように、これらの要素のほとんどは、プロジェクトで作成された一連の文書に一括して適用されるかもしれず、したがってプロジェクトレベル

のメタデータでカバーされます。そのため、ここではプロジェクトで作成された文書の範囲が異なる場合に使用されるオプションの要素として概説しておきます。

要素	説明
プロジェクト名	関連するプロジェクトの名称
対象	モニュメント、遺跡種別、考古学文化、オブジェクト、資料のキーワード。ダブリンコアによる対象との対応。
調査種別	
空間的な範囲	遺跡の範囲、最小・最大の緯度・経度を定義する単一のポイントまたは一連のポイント
時期・年代	時期についてのキーワード、開始・終了年代、C14年代

### 3.4 アーカイブの構造化

ファイルのメタデータや文書を管理するだけでなく、文書や関連ファイルに可能な限りアクセスできるようにするために、いくつかの簡単なステップがあります。まず、プロジェ

クトレベルのメタデータが記録されていること(「プロジェクト・ドキュメンテーションとメタデータ」のセクションを参照)、ファイルを作成する際に適切なファイル命名規則に従っていること(「ファイル命名規則」のセクションを参照)

が重要です。また、主となる文書から削除されるなど、主となる文書とは別に保存された埋め込みオブジェクトが、該当するファイルに関連づけられるようにすることも重要です。例えば、画像用のサブフォルダを作成するなど、アーカイブを構造化することは、抽出された要素のファイルをソースにリンクする1つの方法となります。また文書レベルのメタデータは、文書と、他の場所に保存されている一連の関連ファイル/画像との関係を明示します。

- [1] <http://www.loc.gov/marc/>
- [2] <http://www.loc.gov/standards/textMD/>
- [3] <http://dublincore.org/documents/dces/>
- [4] <http://www.isbn-international.org/>
- [5] <http://www.doi.org/>

## 参考文献

- Ditch, W. (2007) *XML-based Office Document Standards*. JISC: Bristol. <http://www.jisc.ac.uk/techwatch>
- Fanning, Betsy A. (2008) *Preserving the Data Explosion: Using PDF*. DPC Technology Watch Series Report 08. <http://www.dpconline.org/docs/reports/dpctw08-02.pdf>
- Morrison, A., Popham, M. & Wikander, K. (2001) *Creating and Documenting Electronic Texts: A Guide to Good Practice*. AHDS. <http://ota.ahds.ac.uk/documents/creating/cdet/>
- Morrison, A. & Wynne, M. (2005) *AHDS Preservation Handbook: Marked-up Textual Data*. AHDS. [http://ota.ahds.ac.uk/documents/preservation/preservation\\_markup.pdf](http://ota.ahds.ac.uk/documents/preservation/preservation_markup.pdf)

# データベースとスプレッドシート

## セクション 1. データベースとスプレッドシート概説

厳密に言えば、データベースとスプレッドシートは非常に異なる機能を持っていますが、多くの場合、考古学では、どちらも似たような方法（レコード/行とフィールド/列）でデータを収集・保存するために使用されているでしょう。アーカイブの観点からは、重要な特性を考慮すると、両者の類似性がより明らかになります。データベースまたはスプレッドシートの形式でデータを保存する場合に重要なのは、データ値そのものと、このデータが保持されている構造（テーブルまたはシート）の両方です。この観点から、データベースまたはスプレッドシートいずれのオブジェクトも、同様の方法で処理（およびアーカイブ）することができます。

このガイドは、考古学研究の一環として作成されたデータベースとスプレッドシートの最も一般的な特徴を保存するための概要とガイドの提供を目的としています。このガイドでは、両者の共通点と、同様に処理できる方法を紹介するとともに、相違点や、追加の要素、特性、プロセスを文書化すべき範囲についても紹介します。しかしこのガイドは、データベースやスプレッドシートの設計について、その保存に影響を与えると考えられる範囲を超えて、詳細なガイダンスを提供することを目的としていません。

### 1.1 データベースとスプレッドシートとは？

データベースと表計算ソフトは、表形式のデータを列と行に分けて保存するという点では似ていますが、それぞれの用途に応じて区別することができます。表計算アプリケーションは、紙の会計用ワークシートをベースにしており、会計などの数学的データを扱い、その場で計算や処理を行うことを目的としています。一方、データベース・アプリケーションは、さまざまな種類のデータを保存し、そのデータに対して複雑な検索やレポート機能を提供することを目的としています。

## スプレッドシート

スプレッドシートは、データベースよりもシンプルなフォーマットで、一般的に表形式のデータを含む単一または複数の「シート」で構成されています。データ自体は、スプレッドシート内で、数式を使って追加の値（列の合計など）を作成したり、グラフやチャートを生成したりするのに使用されます。これらのグラフやチャートは、シート内に配置したり、シート自体として存在したりすることができます。セルやセル内の値のフォーマットも、スプレッドシートの重要な要素であり、意味を伝えたり、特定の要素を強調したりするために使用されることがよくあります。また、保護されたセルやロックされたセル、セル固有のフォーマット（小数点以下を四捨五入したり、通貨形式で表示したりするなど）を使用することで、スプレッドシートのデータ入力や使用をある程度コントロールすることができます。

## データベース

基本的な設計やアプローチが似ている表計算ソフトとは異なり、データベースはいくつかのタイプ（モデルやアーキテクチャと呼ばれる）に分けられます。考古学で最も一般的に使用されているのは、フラットファイル・データベースとリレーショナル・データベースの2つですが、オブジェクト指向のデータベースモデルの使用も徐々に増えてきています。フラットファイルモデルは、表形式のデータを、レコードを表す横方向の行と、記録する値や属性の種類を表す縦方向の列やフィールドに整理するという点で、スプレッドシートによく似ています。フラットファイル・データベースでは、データの定義や記録の仕方に本質的なルーズさがあり、レコード間で情報のセットが大幅に重複している場合があります。リレーショナルモデルでは、データ構造を事前に定義し、関連する属性グループを別々のテーブルに分割して、キーフィールド（主キーまたは外部キー）によってリンクすることで、フラットファイル・データベースの問題を解決しています。表計算ソフトや多くのフラットファイル・データベースとは対照的に、ほとんどのデータベース・アプリケーション

ンでは、フィールドの長さやデータの種類（数値など）など、記録するデータの種別を厳密に指定することができます（実際に必要です）。

表計算ソフトのデータから作成されたグラフと同じ様に、データベースはデータ値以外のものから構成される可能性があります。データの入力やクエリの実行に使用されるフォームは、多くのユーザーがデータベースと対話する唯一の手段であることが多く、データそのものとは異なるデータベースの内容の一部と見なすことができます。同様に、ユーザーの操作によって得られるクエリや結果、レポートも、データベースの「非データ」構成要素とみなすことができます。

## 1.2 現状の問題点と懸念事項

### 埋め込みオブジェクト

ワープロソフトと同様に、データベースや表計算ソフトの多くは、ファイルに他のメディア（特に画像）を埋め込むことができます。データベースでは、ファイルそのものではなく、外部ファイルへのリンクを保存することが多いのですが、Microsoft ExcelやOpenOffice Calcなどの表計算アプリケーションでは、データから作成したグラフやチャート、その他の画像を埋め込むことができます。テキストファイルと同様に、このようなコンテンツは別個に保存し、アーカイブすることが望ましいでしょう。そうすることで、コンテンツのオリジナルの品質（画像の解像度など）を維持し、テキストコンテンツとは別のアーカイブ戦略を採用できます。

### データの一貫性と文書化

コード化されたデータや不正確に入力されたデータは、データベースとスプレッドシートの両方において、データ再利用の際に問題となります。コード化されたフィールドやデータは適切に文書化してデータベースやスプレッドシートと一緒にアーカイブし、フィールドやデータの意味が時間の経過とともに失われないようにする必要があります。一貫性のないデータ入力は、データの意味を失わせ（例えば、「A」は実際には「a」と等しいのか）、データセットを照会する際に問題となります（これはスプレッドシートよりもデータベースの方が制御しやすいです）。

### データ以外のコンテンツ

上述したように（詳細は後述）、データベースとスプレッドシートはともに、表データ、グラフ、画像以外のものから構成されていることがあります。特に表計算アプリケーションでは、データやそのデータを含むセルに多くの書式（フォントの色やスタイル、セルの色、境界線のスタイルなど）を適用す

ることができます。一般的に、このようなスタイルは、データの特定の様相（列の合計、負の数値など）を強調して、何らかの意味を伝えるために使用されることもありますが、それらはデータの移行（特にプレーンテキストへの移行）の際に失われることがあります。

## セクション 2. データベースとスプレッドシートの作成

### 2.1 一般的な考慮事項

前述のように、本ガイドは、スプレッドシートやデータベースによるデータセットの保存に大きな影響を与える事項についてガイダンスを提供するもので、それ以上のものではありません。以下は、データの一貫性と再利用のしやすさを確保し、保存のライフサイクルを通じて状態を維持するための一般的なアドバイスです。

- スプレッドシートとデータベースのいずれについても、データを入力する際には、可能な限り統制された語彙や確立された単語リストを使用すべきです。
- 一貫性のある意味のあるテーブル/シート、フィールド/列（カラム）の命名規則を遵守しましょう。テーブルやシートが1つのファイルにまとめられていないこともあることにも注意してください。また、アプリケーションによっては、特定のフィールド名の使用が制限されている場合があります（例：ORACLEでは、数字で始まるテーブルや、「desc」や「date」という名前のフィールドは使用できません）。フィールド名や列名にスペースを使用する場合もある。ただし、アプリケーションによっては許可されているスペースの使用ですが、将来のデータ移行の際に問題となる可能性があるため、避けた方がよいでしょう。
- スプレッドシートでは、スタイルやフォーマットを使用して意味を伝えないようにしてください。

### 2.2 ファイルの種類

以下の表は、スプレッドシートやデータベースの作成に使用される一般的なフォーマットの概要を示しており、関連するアプリケーションや長期保存のための潜在的な用途について説明しています。Microsoft Office、OpenOffice、WordPerfect Officeなどの多くのアプリケーションは、Office Open XML (OOXML) と OpenDocument Format (ODF) の両方のファイル形式をサポートしていることは注目に値します。

### データベース

フォーマット	プロパティ・技術	説明	推奨
.accdb	Accessのデータベースで使用されているMicrosoftの商用(プロプライエタリ)形式(2007年以降)	.accdb形式は、Access2007で初めて導入され、Access2010でも導入され、従来の.mdb形式よりも機能を強化するために開発された。堅牢なデータベース設計の観点からすると、多項フィールドや添付ファイルなどの追加機能や拡張機能は、データベースの保存を困難にするだけである[1]。また、Access 2007および	.accdb形式は長期保存には推奨できないが、Accessはいくつかの適切な形式での出力をサポートしている。



フォーマット	プロパティ・技術	説明	推奨
		2010 ではこのフォーマットがデフォルトとなっているが、Access 2010 で作成されたファイルは Access 2007 と完全な互換性がないことにも注意が必要である [2]。accdb 形式は、従来の .mdb ファイルと同様に、引き続き Jet Database Engine をベースにしている。	
.mdb	Access データベース(2003 年以前)で使用されていた Microsoft 社の商用(プロプライエタリ)フォーマット	現在は .accdb フォーマットに置き換えられている。mdb 形式は、Access のインターフェースとデータベースエンジン(Jet Database Engine)をパッケージ化したものである。必然的に Jet のアップデートが行われるので(Access 2.0 はバージョン 2.5、Access 95 はバージョン 3.0、Access 97 はバージョン 3.5、Access 2000-2003 はバージョン 4.0)、その結果、.mdb ファイルのフォーマットは変更されてきた。Jet のバージョン 2 と 3 の間には特に互換性に問題があり、ジェットのバージョン 3 以下を使用している Access のバージョンは、おそらく現在アクセスできない可能性がある。	.mdb フォーマットの保存は推奨されないが、Access はいくつかの適切なフォーマットでの出力をサポートしている。
.odb	OpenOffice Base で使用される XML ベースのデータベース形式	一般的には OpenOffice スイートの一部として使用されているが、.odb 形式は元々の OASIS Open Document (v1.0、ISO および IEC 標準)の仕様には含まれていない。しかし、ODF フォーマットと同様に、.odb ファイルは、内容が「コンテンツ」xml サブドキュメントに含まれ、追加のサブドキュメントには、スタイル情報、ドキュメントのメタデータ、およびアプリケーション固有の設定が含まれる、一連の圧縮された XML ファイルで構成されている。また .odb は Access 形式と同様に、インターフェース(外部のデータソースにアクセスするため)とデータベースエンジン(ここでは HSQL データベースエンジン [3])から構成されている。	.odb 形式は長期保存には適していないが、この形式を使用する多くのアプリケーション(特に Open Office)は、いくつかの適切な形式での出力をサポートしている。
.dbf (.dbt、.ndx)	元々は dBASE で使用されていたが、他企業の製品でも採用されている(例:ESRI の ArcInfo)。	このフォーマットは、1980 年代に dBASE 用に開発されたが、多くの(しばしば互換性のない)バリエーションを持つ汎用フォーマットとして登場し、「xBase」と呼ばれている。ファイル構造は単純で文書化されているが [4]、ファイルの作成に使用されたソフトウェアアプリケーションを特定しておくことが重要である。	このフォーマットは長期保存に適しているが、標準のバリエーションのため、ファイルがどのように作成されたかを正確に知ることが重要である。dBase は、ほとんどのアプリケーションと同様に、より適切なフォーマットへの出力をサポートしている。

スプレッドシート

フォーマット	プロパティ・技術	説明	推奨
.xls	Microsoft Excel で使用されている商用 (プロプライエタリ) のバイナリ形式 (Excel 2003 まで)	Microsoft の独自形式ではあるが、Excel の .xls 形式は広く使用されており、多くのサードパーティアプリケーション (OpenOffice や Google Docs など) で読み込むことができる。2003 年以降の Microsoft Office では、このフォーマットは XML ベースのフォーマット (特に .xlsx) に置き換えられている。他のオープンアプリケーションとの互換性はあるが、xls フォーマットは Excel ファイル (下記の .xlsx を参照) のデフォルトフォーマットではなく、長期保存には推奨されない。	Microsoft Excel で使用されている独自のバイナリ形式 (Excel 2003 まで) Microsoft の独自形式ではあるが、Excel の .xls 形式は広く使用されており、多くのサードパーティアプリケーション (OpenOffice や Google Documents など) で読み込むことができる。2003 年以降の Microsoft Office では、このフォーマットは XML ベースのフォーマット (特に .xlsx) に置き換えられている。他のオープンアプリケーションとの互換性はあるが、xls フォーマットは Excel ファイル (下記の .xlsx を参照) のデフォルトフォーマットではなく、長期保存には推奨されない。
.xlsx	Microsoft 社が作成した Office Open XML (OOXML) フォーマットの一部。ECMA [5] および ISO [6] の標準規格。	Microsoft 社の比較的新しいフォーマットで、Office 2007 でリリースされた。Microsoft 社は、既存の ODF 国際規格を使用せず、独自の仕様 (OOXML) を開発した。このフォーマットは、人間可読な XML ファイルとその他のコンテンツを 1 つの ZIP ファイルにまとめたものである。	文書化されたオープンな標準形式である .xlsx は、理論的には長期保存に適しているが、埋め込まれたコンテンツは個別に保存する必要がある。最終的なファイルは基本的に圧縮されたアーカイブであり、ファイル自体は非圧縮形式で保存すべきである。すべてのスプレッドシートデータと同様に、最大限のアクセシビリティを確保するために、プレーンテキストフォーマットへの移行を考慮する必要がある。
.sxc	OpenOffice XML スイートフォーマットの一部で、Open Office 1.0 Calc で使用されている (現在は .ods に置き換えられている)。	OpenOffice XML sxc 形式 (現在は Open Document .ods 形式に置き換わっている) は、その後継形式と同様に XML ベースのフォーマットです。OpenOffice を含む多くのアプリケーションでサポートされている。このフォーマット自体は、圧縮された XML ファイルで構成されており、実際のスプレッドシートデータはファイルに、画像やその他のコンテンツは別のディレクトリに保存されている。	オープンでテキストベースのフォーマットではあるが、.sxc よりも新しい .ods を使用することが推奨されている。
.ods	XML ベースのオフィス文書フォーマットの ISO 標準 (ISO/IEC 26300:2006 [7]) である OpenDocument スイートの一部。主に、OpenOffice Calc (バージョン 2.0 以降) で使用されている。	OpenDocument Text (.odt) フォーマットと同様、.ods ファイルは基本的に、個別のスタイル、テキスト (XML として)、および埋め込みコンテンツ (画像など) ファイルを含む圧縮された zip ファイル。	XML ベースのオープンフォーマットである .ods は、長期保存に適しているが、理想的にはファイルは圧縮されていない状態で保存されるべきである。また、ドキュメントに画像やその他のコンテンツが含まれている場合、それらは適切な保存形式で個別に保存することが理想的である。ODS は、csv ではファイルの重要な特性をすべて適切に保存できない場合の保存用フォーマットとして使用できる。どの形式が最も効果的かを確認するために、xlsx へのファイル変換も試す価値があるかもしれない。Calc では、csv にエクスポートする際に、文字セットのオプションが用意されている。これは、UTF-8 エンコーディングで xls から csv への変換が必要な場合、良いオプションである。

フォーマット	プロパティ・技術	説明	推奨
.123 および .wk* (.wk4、.wks など)	Lotus 1-2-3 スプレッドシートアプリケーションで作成されたファイル。バイナリ形式と独自形式がある。	Lotus 1-2-3は、1980年代と1990年代に人気のあったアプリケーションだが、2002年以降は開発が中止されている。これらのファイルフォーマットは、MS Excel 2000、またはOpenOffice Calcのいずれかに読み込んで、より適切なフォーマットに変換することができる。	このフォーマットは、長期保存には推奨されない。また、他のオフィスソフトウェアを使用してファイルを変換すると、問題が発生する可能性がある(例:関数や数式が異なる方法で計算されたり、使用できない場合がある)。そのため、重要な追加機能を説明した完全なドキュメントを提供することが特に重要である。
.wq* /.qpw	Quattro Pro(現在はWordPerfect Office Suiteの一部)で作成されたファイル。	表計算アプリケーションQuattro Proで使用される独自のフォーマット。古いファイルはMicrosoft Excel 2000でもサポートされているが(Quattro Pro for Windowsのバージョン1~5ではQuattro Pro Converterのインストールが必要)、これらのファイルをExcelにインポートすると問題が発生する可能性がある(Quattro Pro for Windowsのグラフ、埋め込みオブジェクト、マクロはExcelにインポートすると失われる)。現在のバージョンのQuattro ProはOOXML形式をサポートしている。	長期保存用のフォーマットとしては推奨できないので、ファイルはXMLまたはプレーンテキストベースのフォーマットでエクスポートすること。

- [1] <http://office.microsoft.com/en-us/access-help/which-file-format-should-i-use-in-access-2010-HA010342120.aspx>
- [2] <http://office.microsoft.com/en-us/access-help/convert-a-database-to-the-accdb-file-format-HA010341552.aspx>
- [3] <http://hsqldb.org/>
- [4] <http://www.clicketyclick.dk/databases/xbase/format/index.html>
- [5] ECMA-376
- [6] ISO/IEC 29500-1:2008
- [7] [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=43485](http://www.iso.org/iso/catalogue_detail.htm?csnumber=43485)

### セクション 3. データベースとスプレッドシートのアーカイブ

#### 3.1 アーカイブ化するファイルの決定

スプレッドシートやデータベースは、テキスト文書と同様、殆どの場合作成期間中は単一のフォーマットです。また、これらのファイルは自己完結型であるため、ファイル内に画像やその他のメディアを取り込むことはほとんどありません。もし取り込んでいる場合には、埋め込まれたコンテンツは別個に保存され、それぞれ保存経路をたどることができるようになることが推奨されます。

#### 3.2 どのようにアーカイブするかを決める

多くのアーカイブの関心は、スプレッドシートとデータベースの中核となるデータテーブル/ワークシートそのものであり、テーブルやシートの内容や関係を記述した文書やメタデータにあります。列や行の順序やレイアウトも重要な意味を持つ場合があります(特にレイアウトが重要なスプレッドシートの場合)、フォーム、レポート、クエリ、マクロはコアデータとは見なされないため、保存されないことが多いでしょう。

重要なプロパティとは、保存・管理すべきファイルの最も基本的な要素のことで、その概要は以下のとおりです。

#### 重要なプロパティ

重要なプロパティとは、保存・管理すべきファイルの最も基本的な要素のことで、その概要は以下のとおりです。

- **値**：スプレッドシートやデータベース内の実際のデータ(データやテーブルのシートが複数ある場合もあります)で、セルの見出しや入力値そのものを含みます。
- **グラフィック**：スプレッドシート内のすべての図、グラフ、チャート。ほとんどの場合データベースでは、オブジェクトを埋め込むことはできず、外部ファイルへのリンクとなっています。しかし、新しいデータベース(Accessの.accdbファイルなど)では埋め込み機能が実装されているので、データベース内で使用されている可能性があることを認識しておくことが重要です。
- **レイアウト**：特にスプレッドシートでは、フォーマットや色、特定のスタイルの使用がデータに意味を与えている場合があるので、そのような事例ではレイアウトは重要な特性となり、何らかの方法で保存する必要があります。これは表形式のデータをレイアウトするためにスプレッドシートを使用する場合によく見られます(ワープロソフトを使用した方がよい場合もあります)。このような場合、プレーンテキストベースのフォーマットにエクスポートすると、通常、レイアウトの書式は失われるため、データを保存するために代替フォーマット(PDFなど)を使用する必要があります。
- **リレーショナル**：特にデータベースでは、表やシート間の関係を文書化し、理解できるようにしておくことが重要です。

## 一般的な確認事項

ファイルの変換時にファイルの重要なプロパティが保持されることを確認するだけでなく、変換を実行する前に確認した方が良い項目があります。これは変換プロセス中にスプレッドシートやデータベースの重要なプロパティが失われないようにするためのものです。

- ・ **レイアウトとフォーマット**：前述のとおり、特にスプレッドシートのレイアウトと書式は、入力値などに追加の意味を与えるために使用されることがあり、データを区切り文字付テキストに移行する際にも問題となります。複数の行や列にまたがる見出し（セルの結合）や、色、境界線、フォントなどの書式によって示されている情報をチェックする必要があります。ファイルフォーマットの性質によっては、データを移行する前に、意味が失われないように手作業で編集するか（例：スプレッドシートの場合、結合されたセルを分割し、その中のテキストを各セル内にコピーする必要があります）、または別の交換形式を見つける必要があります。
- ・ **テーブルとシート**：データベースやスプレッドシートは全体を移行することを前提としていますが、どのテーブルやシートを移行すべきかを確定するために、各ファイルの評価を行う必要があります。スプレッドシートには、作成時にデフォルトの数の空白のワークシートが含まれていることがあり、ユーザーは、保存を目的としないデータを一時的に保存するために、追加のシート（データベースではテーブル）を作成することがよくあります（訳注：これらを保存対象にするかチェックすべきです）。
- ・ **数式、クエリ、マクロ**：ファイルに複雑な数式やクエリが含まれており、それ自体を保存する必要があるかどうかを確認すべきです。移行したバージョンのスプレッドシートやデータベースでは、関数によって計算された結果の値のみが保存され、関数自体は保存されない可能性があります。複雑な数式（ワークシートをリンクするものなど）やクエリは、個別にテキストファイルに保存し、後日スプレッドシートの機能を再現できるようにする必要があります。
- ・ **コメントやメモ**：マクロや数式と同様に、移行プロセスでは、ファイルに追加されたコメントやテキストのメモが保存されない場合があります。移行する前に、コメントを別のテキストファイルに保存し、どのファイルのどのセルに関連しているかを明確に示す必要があります。
- ・ **非表示または保護されたデータ**：スプレッドシートには、非表示または保護されたセルが含まれている場合があります。移行時にそのようなデータが含まれているかどうかを確認する必要があります。
- ・ **特殊文字と区切り文字（デリミタ：delimiter）**：データベースやスプレッドシートには、データセット内に特殊文字や一般的な区切り文字（タブや|：パイプまたはパーティカルバーなど）が含まれていることがあります。“|”、“、”（カンマ）、タブなどの区切り文字は、ファイルを区切り文字付テキストに移行する際に問題となる場合があります。このような文字は、移行手続き時に考慮されるよう、早い段階で特定しておく必要があります。区切り文字に加えて、“&”（アンパサンド）、引用符（“”、“”など：スマートクォート）、emダッシュ（-：ダッシュ）などの特殊文字（約物）や外来文字も、データのエ

クスポートやその後の表示に支障をきたす可能性があります。また、データベースやスプレッドシートには外国語の文字が含まれている場合、文字エンコード（UTF-8など）が指定されていなければ、テキストファイルにエクスポートできない（訳注：文字化けを起こす）ことがよくあります。

- ・ **リンク**：特にデータベースでは、テーブル間の関係を確認、文書化し（下記の3.3を参照）、修正することが重要です（重複や孤立したレコードが存在しないかどうかをチェックする）。スプレッドシート内のワークシートも、セルが他のシートにリンクして値を参照していることがあります。さらに、データベースもスプレッドシートも、外部に保存されている画像などのファイルにリンク（またはそのファイル名を保持）していることがあります。このような外部ファイルがプロジェクトのデータセットの一部に含まれている場合、ファイル名が正しく保存されているかどうかを確認する必要があります。

## ファイル名の付け方

データの保存方法によっては、データベースのテーブルやスプレッドシートのワークシートの名前を変更する必要があります。ファイルを区切り文字付テキストとして保存する場合、データベースとスプレッドシートのいずれでも、最終的なデータセットはテーブルやワークシートごとのテキストファイルとして構成されるでしょう。可能であれば、ファイルは元のファイルと同じ名前で保存されるべきです（ただし、ファイルの拡張子は異なってよい）。複数のワークシートやデータベーステーブルをエクスポートする場合、ファイル名には、元のスプレッドシート/データベースの名前と、データの元となったワークシートやテーブルの名前の両方を反映させる必要があります（例：[データベース名] - [テーブル名]）。

- ・ データベース名] - [テーブル名] .txt  
例：Findsdatabase-lithics.txt
- ・ スプレッドシート名] - [ワークシート名] .txt  
例：siteregister-photos.txt

画像、クエリ、メモ、数式などを保存するために別のファイルを作成する場合は、[スプレッドシート名]-[ワークシート名] - [グラフ名] .tif など、どこから入手したのかを正確に追跡できるような論理的な方法で名前を付ける必要があります。

場合によっては、テーブル名を短くしたり、ファイル名に使用できない句読点を削除したりして、テーブル名を変更する必要があるかもしれません。このような場合には、できるだけ元の名前に近づけることが重要です。

## ファイル形式

多くのデータベースやスプレッドシートでは、データを保存する際の形式として区切り文字付テキストが好まれます。しかし、前述のように、スプレッドシートやデータベースには、特定のフォーマットでなければ維持できないスタイルまたは機能的な要素が含まれている場合があります。このような場合には、互換性のあるオープンなXMLベースのフォーマット（例：.odsまたは.xlsx）を使用するか、データをテキストとしてエクスポートするとともにファイルの要素を文書化して保存することが推奨されます。

保存形式	要件
区切り文字付テキスト (例: .txt, .tab, .csv)	区切り文字付テキストは、一般的にスプレッドシートやデータベースファイルの保存に適したフォーマットで、多くのアプリケーションから利用できる一般的な出力形式である。このようなファイルでは、セルは区切り文字(デリミタ)と(オプションで)修飾子(例: 要素を二重引用符で囲みカンマで区切った .csv ファイル)で区切られ、行は改行で表される。一般的な区切り文字としては、カンマ(csv ファイル)、タブ、  (パイプ)などがある。区切り文字付テキストは広く使用され、多くの形式(例: .csv)は MS Excel などの一般的なアプリケーションで直接開くことができるという利点もある。ただし、このフォーマットは単にデータを保存するだけで、追加要素(画像、数式など)は別途アーカイブする必要がある(以下のセクションを参照)。
.ods	表計算ファイルの長期保存に適しているが、埋め込まれたグラフやチャートは別個に保存する必要がある。
.xlsx	表計算ファイルの長期保存に適しているが、埋め込まれたグラフやチャートも別個に保存する必要がある。

### 移行後の確認事項

ファイルを新しいフォーマットに移行する際には、その過程でデータが失われたり破損したりしていないか、いくつかのチェックを行うことが重要です。このようなチェックには次のようなものがあります。

- ・ エクスポート後の行数のチェック
- ・ テキストフィールドの長さを確認。(特にデータベースの場合) データ長がフィールドの長さとは一致している場合、値が切り捨てられている可能性がある。
- ・ すべてのシートやテーブルがエクスポートされているか確認
- ・ 特殊文字が保存されているか確認

### その他のフォーマット

スプレッドシートデータの公開には、Portable Document Format (できれば .pdfa。「ドキュメントとデジタルテキスト」の章を参照) が使用されることがあります。これは、csv ファイルでは適切に再現できず、データを XML ベースのフォーマット (ods または xlsx) に移行すると失われてしまうような書式やレイアウトの形で重要な情報が示されている場合にのみ使用してください。複雑なスプレッドシートの場合、ユーザーがスプレッドシートの外観を確認しながらデータを操作できるように、csv 版と一緒に PDF 版を提供すべきである場合もあります。

表計算ソフトには XML も広く使われていますが、今後普及する可能性があるアプリケーションとして、GNOME デスクトップの一部である Gnumeric [1] があります。Gnumeric は、自由に利用できるオープンソースで、OpenOffice Calc や Excel と同様に、圧縮された XML をネイティブファイル形式としています。

XML は、データベースデータを保存するための信頼性の高いフォーマットでもあります。スキーマ (または DTD) は、データ辞書を保持していますが、これは区切り文字付テキストファイルでは簡単には保存できず、便利なものではありません。さまざまなデータベース形式に対応したコンバータが用意されています。また、テーブルの定義やデータを SQL の DDL や DML 文として保存することもできます(ただし、SQL の方言の違いには注意が必要です)。

また、データベースの選択肢としては、SIARD [2] データベース・アーカイブ・ツール (Software Independent Archiving of Relational Databases) があり、Planets のツールキットの一部組み込まれています。SIARD Suite は、XML、SQL:1999、UNICODE などの国際標準に基づいており、現在、Oracle、Microsoft SQL Server、Microsoft Access など、多くのデータベースをサポートしています。スイス連邦公文書館では、SIARD Suite を無料で配布しています (ライセンス契約の条件があります)。

### 3.3 メタデータと文書化

データベースやスプレッドシートを確実に保存・再利用するためには、さまざまなレベルのメタデータや文書化が必要です。以下の要素は、データセットと一緒に記録・保存されるべき事項です。

要素	説明
プロジェクト名称	-
データベース/スプレッドシートのファイル名	-

スプレッドシート/データベース内の各ワークシート/テーブルについて、以下の項目を繰り返しメタデータとして文書化してください。

要素	説明
ワークシート/テーブル名	-
ワークシート/テーブルの目的	-
行数	-
主キー(データベースの場合のみ)	-
外部キー(データベースの場合のみ)	-

スプレッドシート/データベース内の各列/フィールドについて、以下の項目を繰り返しメタデータとして文書化してください。

要素	説明
フィールド名	データベースのフィールドまたはスプレッドシートの列名
フィールドの説明	使用されているフィールドやコード、用語の完全な説明。データセット内で使用されているコードを別の文書として提供することもできる
データの種類とフィールド長 (データベースの場合のみ)	—

以上の要素は、基本的にデータベースに紐づけられた「データ辞書」を構成します。また、特にデータベースでは、テーブル間の関係を言葉で表現したり、エンティティ・リレーションシップ・ダイアグラムを作成したりすることも必要です。

ドキュメントには、数式、クエリ、マクロ、コメントなど、保存が必要なスプレッドシートやデータベースの追加機能も含まれます。これらは通常、テキストファイルとしてデータと一緒に保存することができます。

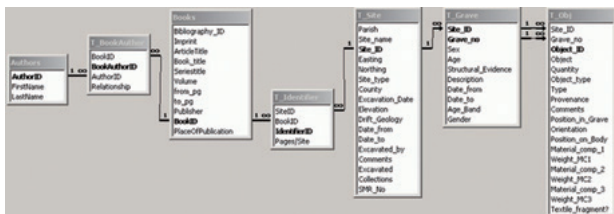


図1：エンティティ・リレーションシップ・ダイアグラムの例

## ACE のケーススタディ：INRAP ArchéoDB 調査登録システムのためのアーカイブの準備

Emmanuelle Bryas and Carine Carpentier (フランス国立事前考古学研究所: French National Institute for Preventive Archaeological Research)

このケーススタディは、Archaeology in Contemporary Europe (ACE) モビリティ奨学金スキーム (mobility bursary scheme) の助成を受けて、2012年4月にADSで行われた2週間のワークショップの一環として作成されました。

### 3.4 データの構造化

ファイル間の関係はファイル名を見れば一目瞭然ですが、エクスポートした表やワークシートを1つのディレクトリにまとめておくのが便利です。エクスポートされた画像やドキュメントは、この構造の中のサブディレクトリに格納することができます。

[1] <http://www.gnome.org/projects/gnumeric/>

[2] <http://www.bar.admin.ch/dienstleistungen/00823/00825/index.html?lang=en>

#### 引用文献

OASIS (2005) *Open Document Format for Office Applications (OpenDocument) v1.0*. <http://www.oasis-open.org/committees/download.php/12572/>

OpenDocument-v1.0-os.pdf

## ArchéoDB

INRAP (フランス国立事前考古学研究所: Institut National de Recherches Archeologiques Preventives) は過去4年間にわたり、タブレット PC を使用して、発掘調査現場段階から直接データを記録、収集されたすべての情報をリレーショナル・データベースによって一元化し、発掘後の段階で調査チームが収集したデータを共有サーバー (NAS) に保存できるようにするという実験を行ってきました [1]。Nicolas Holzem (INRAP Centre) が開発した初期のデータベースは DataDiag と呼ばれ、2010年の夏からさまざまな評価テストが行われました。このデータベースは、Lassay-sur-Croisne や Neuvy-Pailloux などのいくつかの発掘調査で使用された後、Etrechet の最初の2つの発掘調査「Croc au Loup」と「Le Four à Chaux」での新たな評価テストにより、特定方向性の低いシステムへと進化しました。

ACE スキーム実施の間、ADS チームにいくつかの INRAP データベースシステムが紹介された後、ArchéoDB データベースのケーススタディに焦点を当てることが合意されました。このケーススタディでは、データベースとそれに関連するドキュメントファイル (写真、図面、GIS ファイル、イベントリ) のバックアップなど、アーカイブのさまざまな異なる側面にアプローチすることができます。

このデータベースは、ここで取り上げている Etrechet “Fets de Renier” の第3次発掘調査に導入するために開発されたバージョン 1.3.20 です。このデータベースには、1122 の構構とその層序単位の記録、写真と撮影時刻の記録 (サムネイルで表示) が含まれています。このデータベースには、今後も他のデータ (日付、特定の研究結果) を追加していく予定です。GIS の利用はまだ始まったばかりで、これから進化していくでしょう (図2)。つまりこれは、アーカイブすることが決定されたデータベースの中間バージョンです。実際の手続きとしては、登録者が、より完全なデータベースの登録を要請し、最終的に第2版 (内容的には同じ) がアーカイブされることを想定しています。

### 保存と公開の観点から採用された選択肢

保存すべき情報と使用すべきフォーマットの選択は、ADS ウェブサイトで利用可能なオンラインガイド「Guides to Good Practice」の推奨事項に沿って行われました [2]。しかし ArchéoDB の特徴にもとづき、メイン画面のコピーと TIFF ファイル形式の入力フォームが追加で (人間工学で行われた作業の記録とデータ取得の整理のため) 保存することになりました。

また、データベースで作成されフォーマットされたインベントリ (レポート) は、センター所在地の地域考古学サービスの現在の要件と一致している場合、アーカイブと公開のために PDF/A-1b (PDF/A-1a として保存するには互換性がない) で保持されました。

データベースは、永久保存に最も適した TXT ファイル形式で、「|」 (パイプ) による区切り文字付テキストにより各テーブルを保存することにしました。

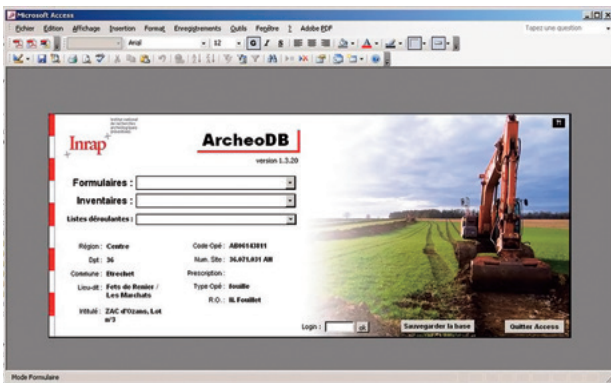


図1：ArcheoDBのメイン画面

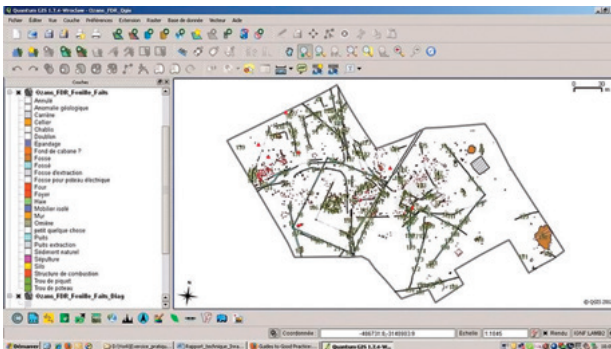


図2：DBに付随するGISプロジェクト

ADSの指示により、入力フォームのドロップダウンメニューのために作成された参照テーブル（データベースのプレフィックスは“lst\_”）や、クエリによって生成されたテーブルは保存されていません。しかし、データベースのテーブルを記述したメタデータファイルには、フィールドを編集する際に参照テーブルや値のリストを使用したことが一貫して記述されています。

ラスター画像（写真と撮影時刻）は、元々 JPEG ファイル形式で作成されていましたが、再登録の際の損傷を避けるため、TIFF ファイル形式へ変換しました。JPEG ファイル形式は、参照用、およびオンライン公開のために残されています。

データベースには、1枚のベクター図（minute field template）が含まれています。これは元々 Adobe Illustrator で作成されたものですが、保存と公開のために SVG ファイル形式で保存されています。

GIS ファイルについては、シェープファイルを構成する3種類のファイルのみを保存・配布しています。SHP (Shape Format)、SHX (Shape Index Format)、DBF (Attribute Format in dBase) の3種類です。ADSの勧告に従い、QGISアプリケーションのプロジェクトファイルをはじめ、LYR (レイヤーシンボロジー)、PRJ (投影フォーマット)、QML (スタイルダイパー)、SBN、SBX (フィーチャの空間インデックス) は保存用に残していません。

## データベースのファイルの準備[4]

### データベースのテーブルを.txt形式で出力する

テキスト形式は、データの保存とオンラインでの公開の両方に有効です。ArcheoDBの主要なデータベーステーブルに対応してテキスト形式で生成されたファイルは、対応する

アーカイブフォルダの2つのセクション（「保存」および「公開」）に複製されます。

以下に、TXTファイル形式のテーブルをエクスポートする手順を、順を追って説明します。

- ・エクスポートする表を開きます（図3）
- ・「デザインビュー」で、「一般」ボックス [5] にあるフィールド名とそのキャプション（存在する場合）のアクセント記号、特殊文字、スペースを削除して、フィールド名がきれいになっていることを確認します（図4）
- ・「ファイル」→「エクスポート」を選択します
- ・テキストファイル形式を選択し、ファイル名を「データベース名-テーブル名」（スペース、アクセント、特殊文字なし）とし、「エクスポート」をクリックします（図5）
- ・エクスポートフォーマットとして「区切り文字付」を選択し、「次へ」をクリックします（図6）
- ・区切り文字の種類として「その他」を選択し、値として「|」（パイプ）を入力し、「最初の行にフィールド名を含める」にチェックを入れます（図7）
- ・「次へ」をクリックし、「完了」をクリックします

## リレーションのバックアップ

リレーショナル・スキーマの（画面表示の）コピーは、データベースを記録するために TIFF ファイル形式で保存しましたが、オンライン公開のために JPEG ファイル形式でも保存することにしました。

リレーショナル・スキーマから画像を生成する手順を以下に示します（Microsoft Access 2003 で直接画像フォーマットに出力する機能はありません）。

- ・メイン画面のフォームを終了し、テーブルのリストにアクセスする。「リレーション」をクリックする（図8）
- ・すべてのツリー関係が展開され、画面に表示されていることを確認し、「ファイル」→「リレーションの印刷」を選択する（図9）
- ・「ファイル」→「ページ設定」をクリックする（図10）
- ・すべてのテーブルのレイアウトを調整し、フィールドやリンクがよく見えるようにする（例では、A3横長のディスプレイフォーマットを選択し、「特定のプリンタを使用」を選択してプリンタをクリックする（図11）
- ・「ファイル」→「印刷」を選択します（図12）
- ・プリンタを「Adobe PDF」として選択する（図13）
- ・「保存」を選択します（図14）
- ・このPDFプリントから、TIFFファイル形式（持続可能な保存のために唯一有効な形式）のアーカイブがまだ生成されていません [6]。これを行うには、生成されたPDFを Adobe Acrobat Pro で開き、「ファイル」→「名前を付けて保存」をクリックします（図15）。
- ・TIFFファイル形式を選択し、「データベース名-relations」としてファイル名をつけて、「保存」をクリックします（図16）
- ・結果として得られた画像を JPEG ファイル形式で再保存し、オンラインで公開します

### 入力フォームのスクリーンショット

データベースの記録のために、入力フォームのスクリーンショットを保存することにしました。このバックアップは、オープンソースのアプリケーションGIMPを使って、以下の手順でTIFFファイル形式で行います。

- ・キーボードの「Print Screen」キーを使ってスクリーンショットを作成する（図17）
- ・写真処理ソフトを開き、「ファイル」→「新規画像」（Adobe Photoshopでは「新規」）を選択する（図18）
- ・ファイルの属性（名前、幅、高さ、解像度）を設定する（図19）
- ・印刷画面を貼り付ける（図20）
- ・保存する画像領域を選択し、カットオプション（Adobe Photoshopのクローズツールに相当）を使って選択範囲を切り抜く（図21）

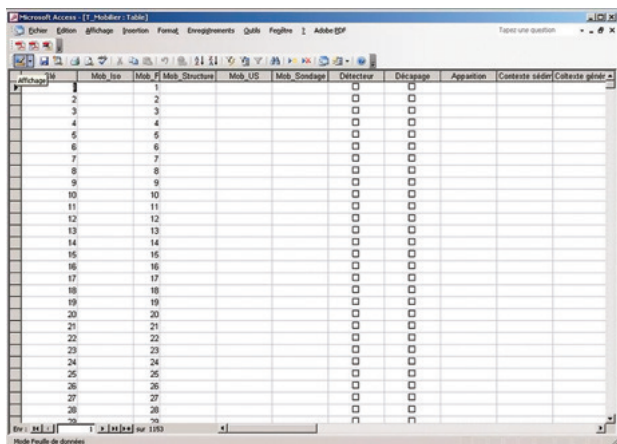


図3：エクスポートするテーブルを開く

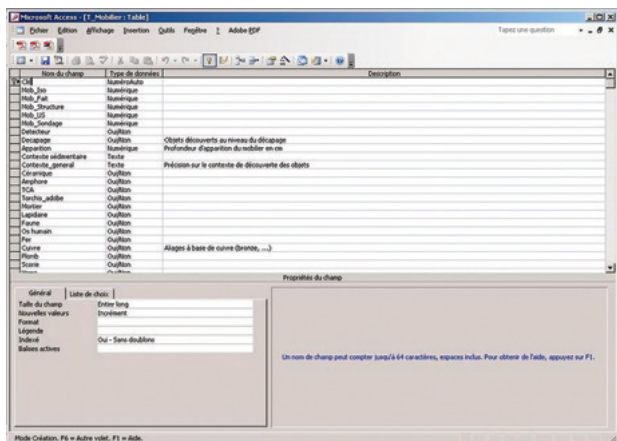


図4：テーブルのフィールド名の編集（デザインビュー）

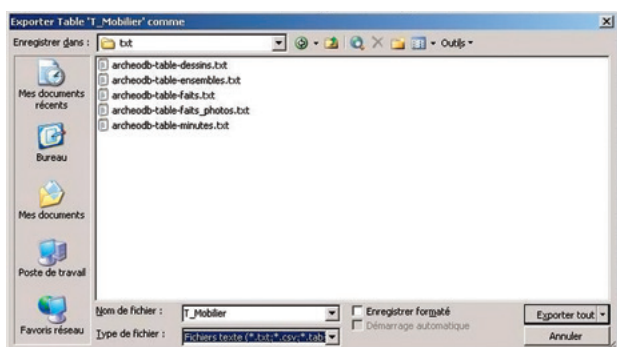


図5：テキスト形式でテーブルに名前を付けて保存

・「ファイル」→「名前を付けて保存」を選択し、TIFF形式で保存する（図22、23）

### データベースに関連したファイルの作成 写真と撮影時刻（ラスター画像）の扱いについて

デジタル写真と撮影時刻は、元々はJPEGファイルの形式で、アーカイブの公開セクションに同じ形式で保存されています。

アーカイブする前に、オリジナルの画像は、まずXnViewというアプリケーションを使ってセットとして名前を変更す

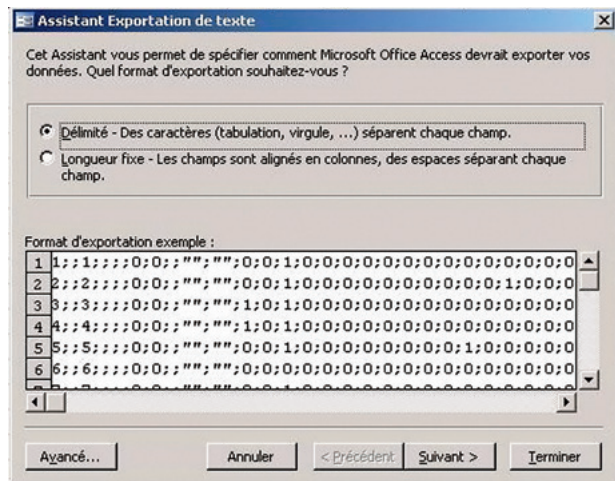


図6：エクスポート形式として「区切り文字付」を選択

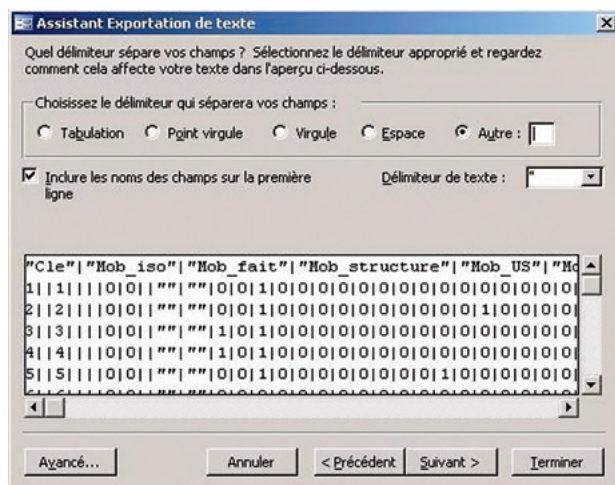


図7：区切り文字と「先頭行をフィールド名として保存」を選択

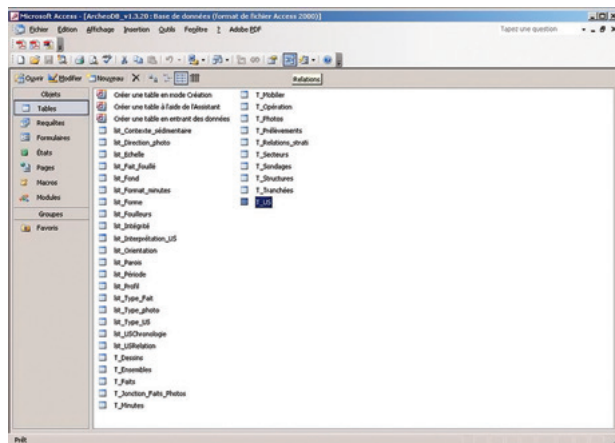


図8：表示するリレーションの選択



する必要があります。名前は以下のように構成されています。「データベース名-関連するテーブル名-オリジナルの画像名」(スペース、アクセント、特殊文字は使用しない)。

手順は以下の通りです。

- すべての画像を選択し、「編集」→「名前の変更」を選択します。
- ファイル名の前に、データベース名と、画像が添付されているテーブルの名前を追加します。ファイル名の前にダッシュとアスタリスク(\*)を付けると、元の名前が自動的に

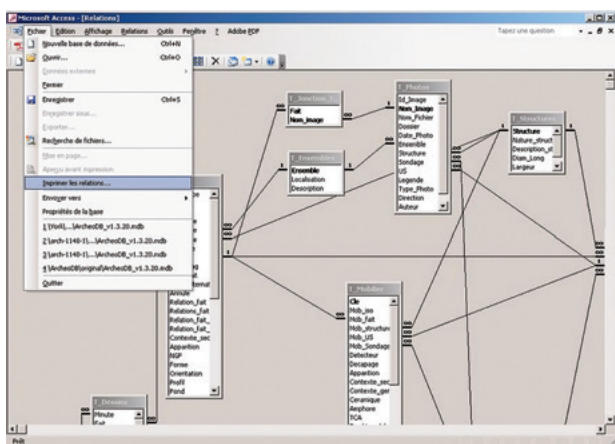


図9：メニューより「リレーションの印刷」を選択

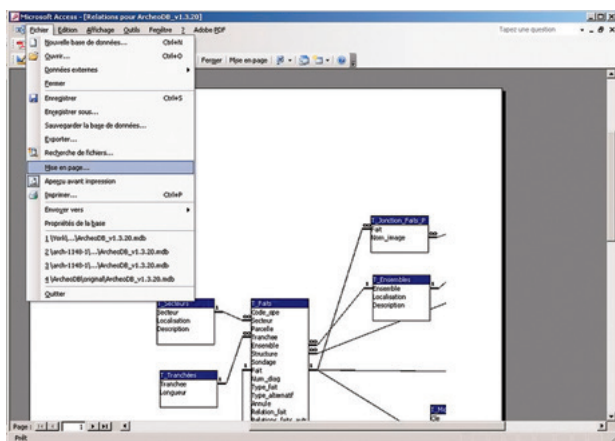


図10：リレーションの状況をレイアウト

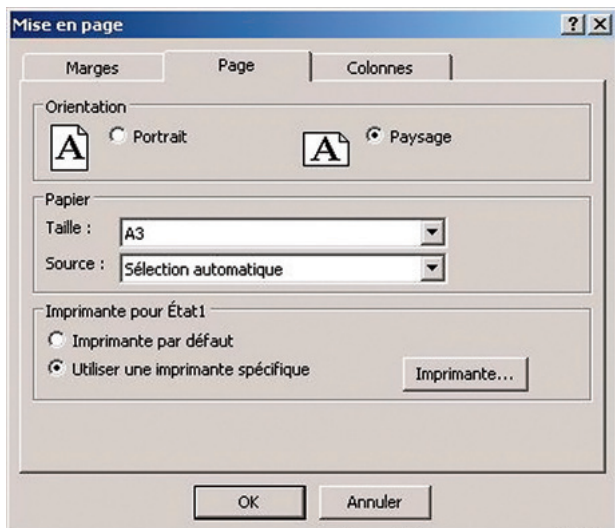


図11：印刷ページ設定

動的に挿入されます。ファイルには拡張子を付けなくてください(図24)。

恒久的な保存のために、これらの画像を Adobe Photoshop を使って TIFF ファイル形式に変換します。手順は以下の通りです。

- アプリケーションを開き、「ファイル」→「スクリプト」→「イメージプロセッサ」をクリックする。
- 修正する画像のあるフォルダを選択し、処理後の画像の保存先を指定する。
- TIFF ファイル形式を選択した後、実行する(図25)。

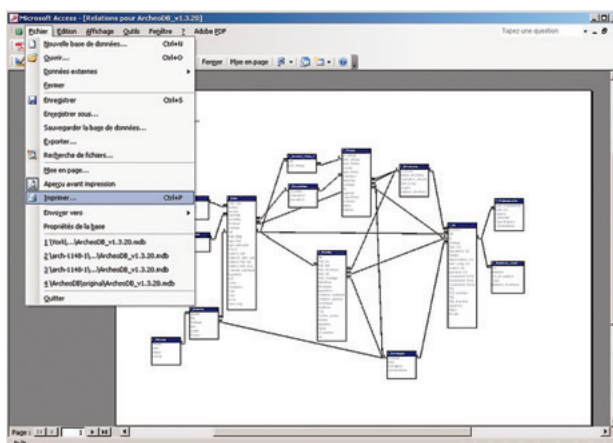


図12：リレーションを印刷

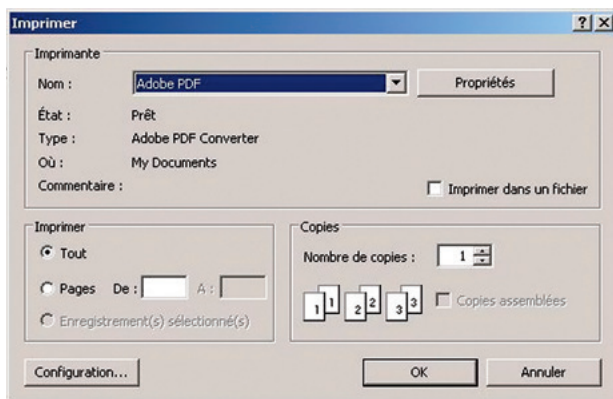


図13：「Adobe PDF」を出力先(プリンター)に指定

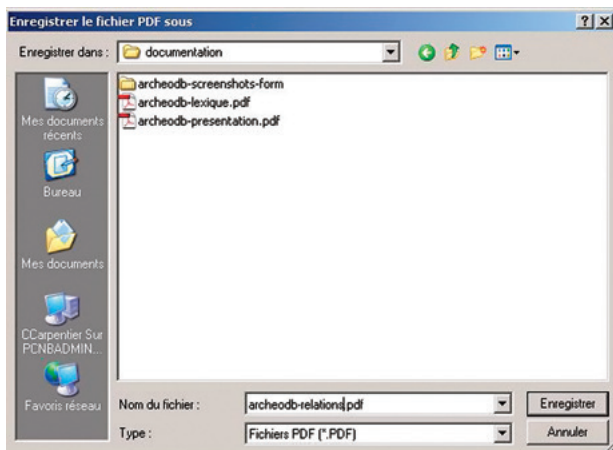


図14：データベース名のプレフィックスを付けてPDFに書き出しを実行

## 図面の扱い(ベクター画像)

ArcheoDB データベースには、Adobe Illustrator で作成された 1 枚の図面のみが関連付けられます。これは、現地調査の基礎として使用することを意図した空白のテンプレートです。そのため、ラスター画像ファイルはなにひとつ関連付けられていません。

図面は、保存と公開のために SVG ファイル形式に変換しました。その手順は以下の通りです。

- ・ Adobe Illustrator のファイルメニューの「名前を付けて保存」または「スクリプト」から「SVG 形式で保存」を

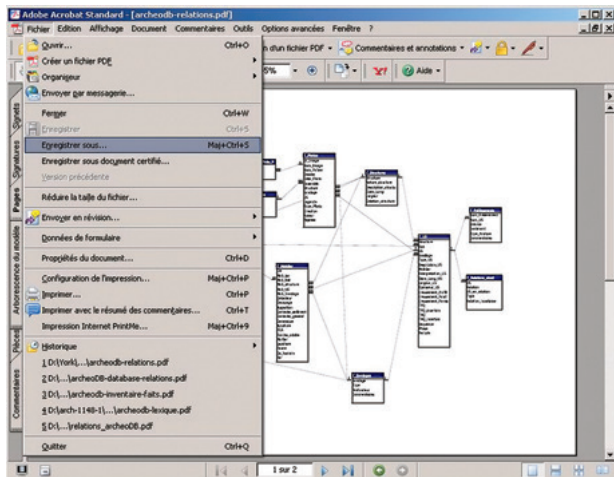


図 15 : PDF に保存されたリレーションを開き、再保存

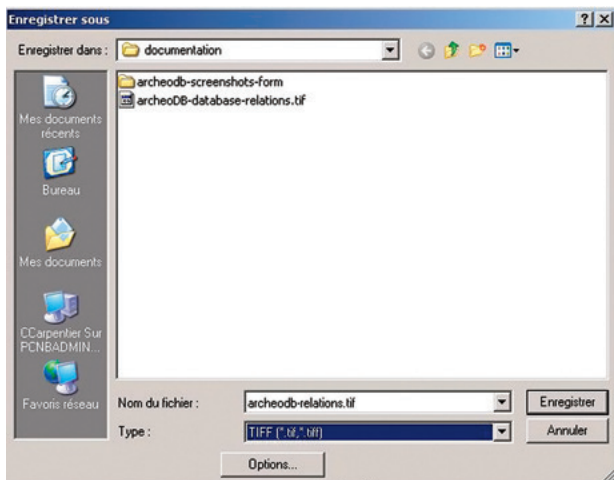


図 16 : TIFF を選択して保存

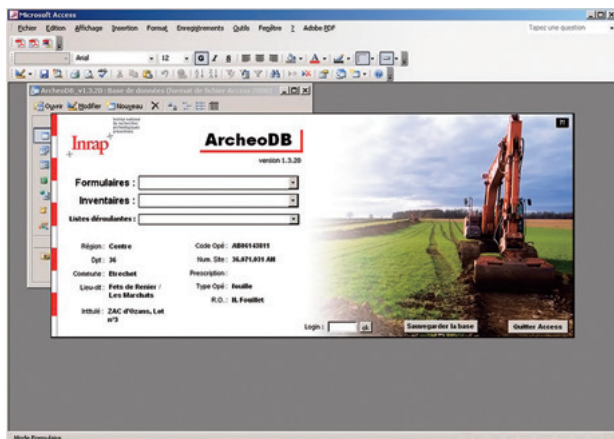


図 17 : ArcheoDB のメイン画面のスクリーンショット

選択します (図 26)。

- ・ ファイルの名前を以下のように指定します: 「データベース名 - 関連テーブル名 - オリジナルファイル名」 (スペース、アクセント記号、特殊文字を含まない) [7]。

## 一覧表の扱い(書式を整えられたテキスト)

通常、テキスト文書は、その書式やレイアウトを保存したい場合を除き、TXT ファイル形式でアーカイブされます。書式やレイアウトを保存したい場合には、PDF/A の使用が推奨されます。PDF/A-1 仕様は、ISO (19005) [8] によって発行され、電子文書の普及と交換の安全性と信頼性を確保する

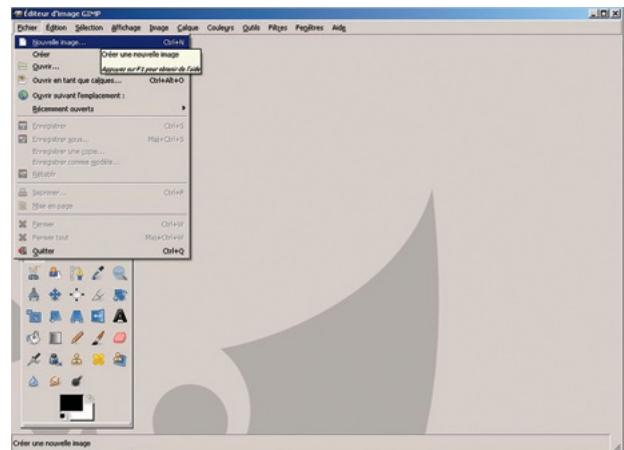


図 18 : 画像処理ソフトで新規画像を作成

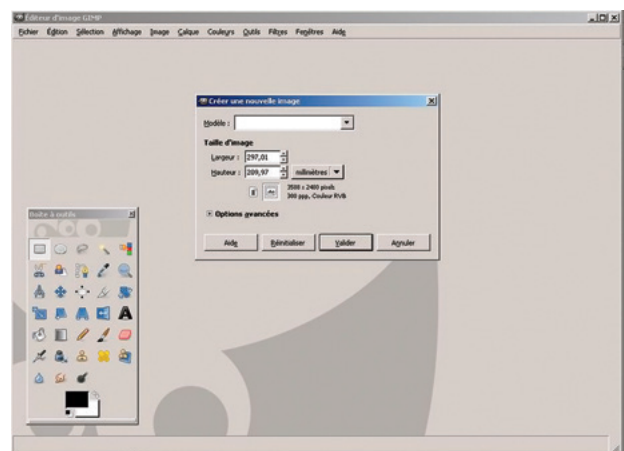


図 19 : ファイルの属性を設定

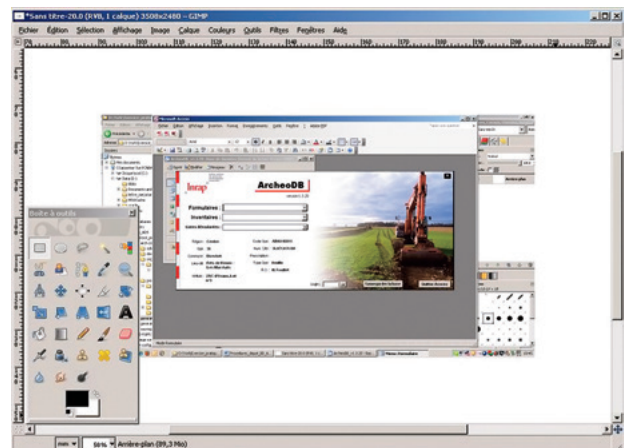


図 20 : スクリーンショットを (画像処理ソフトに) 貼り付け

ために、世界中の標準化団体で使用されています。

PDF/A-1には、2つのバリエーションがあります。

- ・ PDF/A-1 : ISOのフル規格
- ・ PDF/A-1b : ISO企画の簡易版(このバージョンでは、文書の可読性と、良好な画面表示、印刷性能を保持している)。

元のファイルがPDF/A-1aに変換可能な場合は、それを優先します。

以下は、Microsoft Accessから直接一覧表を変換するため

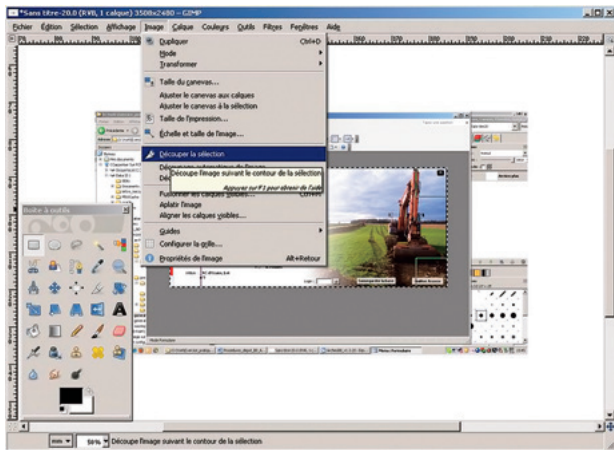


図21: 画像の切り抜き

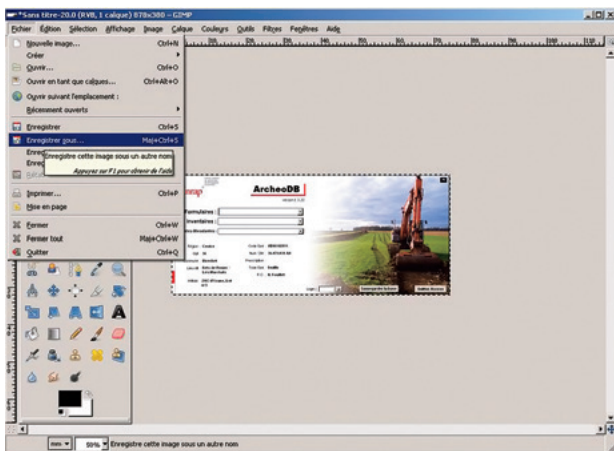


図22: 保存オプションの選択

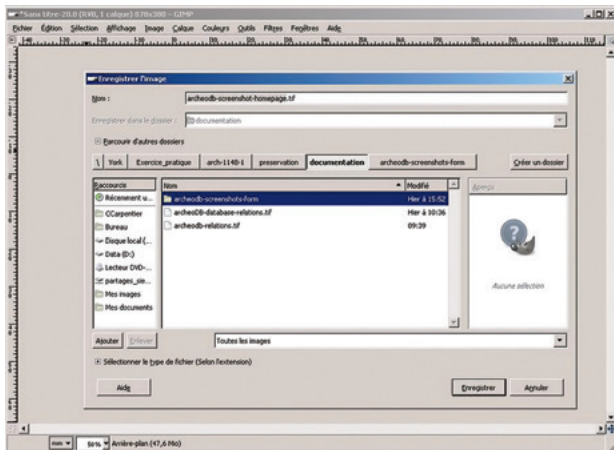


図23: TIFF形式を選択して保存

の手順です [9]。(訳注: アクセスのレポートをPDFとして出力している)

- ・ 変換したい一覧表を Access で開き、「Acrobat」→「設定」を選択します (図27)。
- ・ PDF A 準拠ファイル (可能ならば A-1a 形式、それ以外ならば A-1b 形式) で保存 [10] のオプションをチェックし、OK をクリックします (図28)。
- ・ 「PDFの作成」を選択します (図29)。
- ・ ファイル名を「データベース名---一覧表---一覧表のオブジェクト」で指定し、保存します。

## GISファイルの処理

GISからのファイルには何の処理も必要ありません。シェープファイル(SHP、SHX、DBF)を構成するすべての利用可能なファイルの中から必要なものを選択するだけです。オリジナルのGISプロジェクトで表現されていた各レイヤーについて、対応するファイルの三つ組が保持されていること

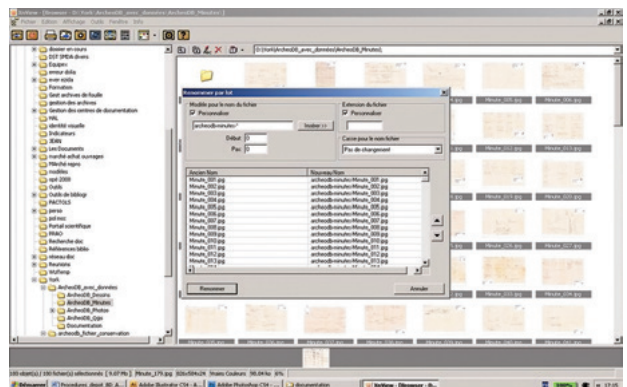


図24: 選択した画像ファイル名の一括変換

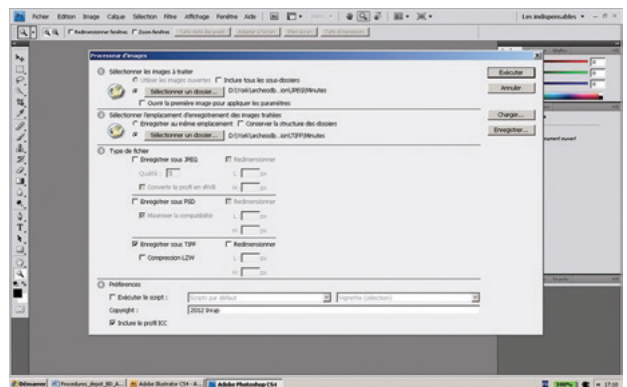


図25: 選択した画像のTIFF形式への一括変換

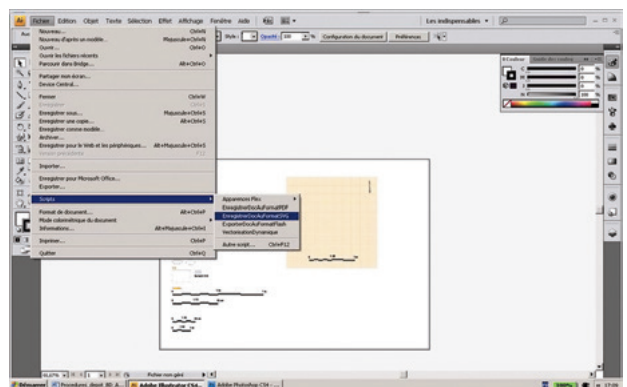


図26: IllustratorからのSVGファイル形式での保存

を確認します。

## プロジェクトの文書化:メタデータファイル プロジェクトに関連するメタデータ

最初に埋めなければならないメタデータファイルは、プロジェクトに関するものです。ADSのウェブサイトで公開されている登録者用ガイドライン(“Collection-level Metadata Template”)[11]の中から対応するモデルをダウンロードしました。

要求されたさまざまなメタデータ(タイトル、説明、主題、場所、著者、日付など)を記入しました。このファイルは“archeodb-metadata-project”と名付けられ、ODTファイルとして保存しました。

### ファイル単位のメタデータ

このメタデータの埋め込みは、オンラインの登録者用ガイドラインからダウンロードできるテンプレートではなく、技術的要件を満たすメタデータを自動的に生成できる無料のアプリケーション DROID [12] を使用しました。このツ

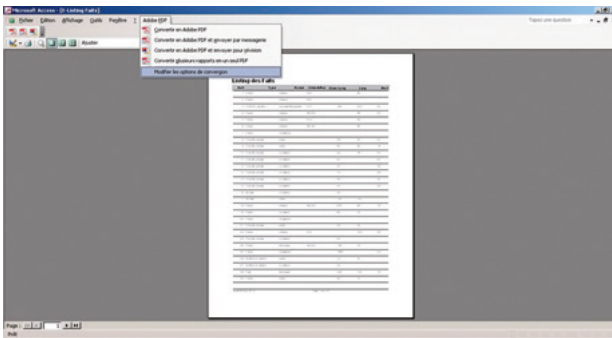


図27: PDF変換のオプションを選択

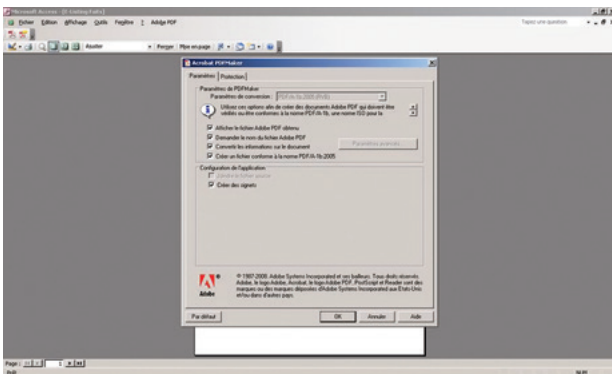


図28: PDFに変換するためのオプションの設定。

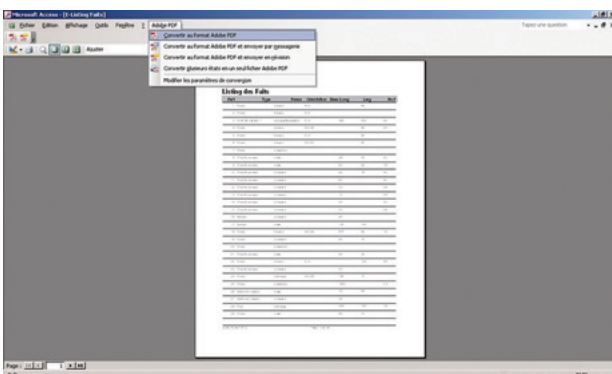


図29: アクセスレポートのPDFへの変換

ルで行った分析結果は、CSVファイル形式で“archeodb-metadata-files”というファイルに保存されます。

以下に、そのメタデータの生成手順を説明します。

- 無名の「プロファイル」を自動的に作成する DROID アプリケーションを起動し、「追加」を選択します。
- アーカイブ内のすべてのファイルを選択して「OK」をクリックします(図30)。
- 「開始」を選択してファイルの識別を開始し、スキャンが完了するのを待つ(図31)。
- 「保存」を選択してコンテンツ・プロファイルを保存し、データベースの名前を付けます(図32)。
- メタデータが表示されていることを確認し、「エクスポート」を選択します(図33)。
- 「行IDごとに1つのサイズ」のオプションをチェックし、「プロファイルのエクスポート」をクリックします。
- ファイル名を「データベース名-メタデータ-ファイル」とし、CSV形式を選択して「保存」をクリックします。

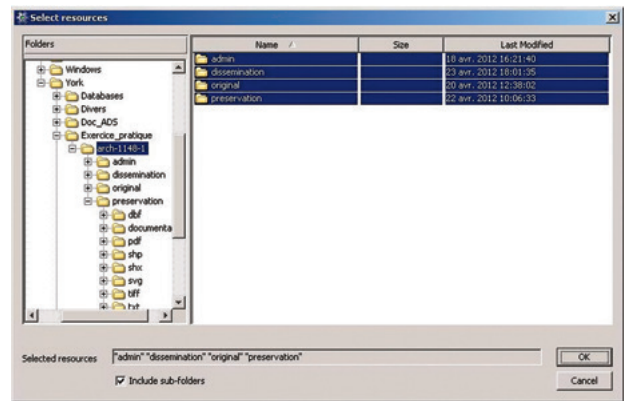


図30: メタデータ登録を行うファイルの選択

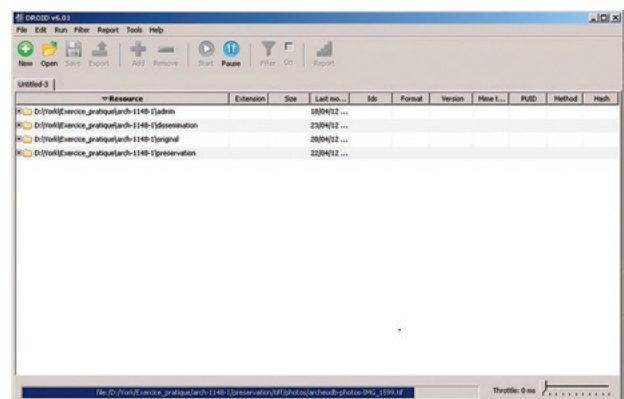


図31: 登録処理中

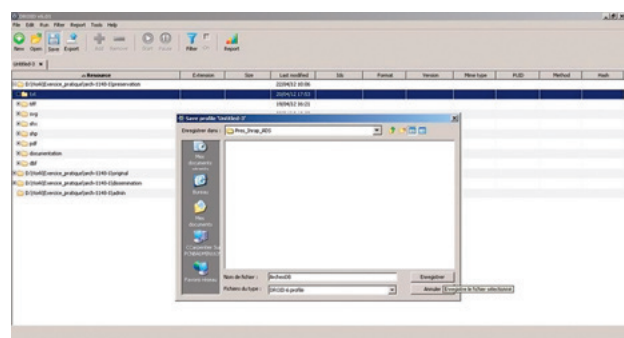


図32: コンテンツ・プロファイルの保存

CSV ファイルの内容を読みやすい表計算ソフトの形式で表示するには、次の手順を実行するだけです。

- ・ Microsoft Excel で CSV ファイルを開き、最初の列を選択して「変換」を選択します (図34)。
- ・ 区切り記号としてカンマを選択し、「次へ」をクリックします (図35)。
- ・ デフォルトのデータ形式である「標準カラム」を維持し、「終了」をクリックします。
- ・ メタデータが読みやすい形で表示されます (図36) [13]。

### ドキュメントに伴うメタデータ

異なる種類のドキュメントに伴うメタデータの作成には、Guides to Good Practice [14] に記載されているメタデータフィールドを使用しました。メタデータファイルはODTファイル形式で保存しました。

ドキュメントの種類ごとに1つずつ、5つのメタデータファイルを生成しました。メタデータ：archeodb-metadata-database、写真：archeodb photo-metadata、撮影時刻：archeodb-metadata-minutes、図面：archeodb-metadata-drawings、そしてGIS：archeodb-metadata-gisです。

### アーカイブ構成要素のツリー構造の作成

#### トップレベルフォルダの作成

アーカイブ全体を含むファイルは、以下のような名前：Arch - コレクション ID - バックアップのバージョン番号にする必要があります。ここで扱うコレクションの場合、第一階層のフォルダ名は“arch-1148-1”となります。

#### 第2レベルのフォルダの作成

##### 「管理(admin)」フォルダ

今回の実習では、このフォルダは空白のままにしてありま

すが、通常、以下が含まれます。

- ・ ADS の指示により CMS で作成されたコレクションに関連するメタデータのエクスポート (DC\_metadata.txt)
- ・ 登録者と ADS が共同で署名した「ライセンス」登録の最初のページのスキャン (licence.tif)。

#### 「オリジナル(original)フォルダ」

オリジナルコレクションを格納するフォルダです。このフォルダには、CMS によって生成された「受け入れ番号」(ここでは「2246」)を名付けられたサブフォルダが含まれています。その中には、さらに別のレベルのサブフォルダがあり、登録の日付を示しています (例：“2012-04-18”)。私たち自身のリポジトリのファイル版の中には、データベースの名前 (“archeodb”) で名付けられたトップレベルのフォルダがあります。これにより、arch-1148-1/2446/2012-04-18/archeodb というパスでファイルにアクセスすることができます。

私たちのフォルダ「archeodb」では、オリジナルのファイルは以下のように整理されています。

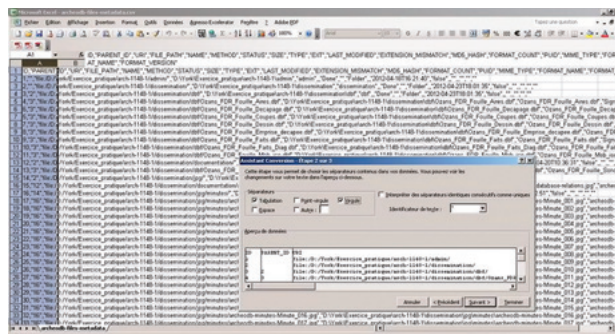


図35：区切り文字の選択

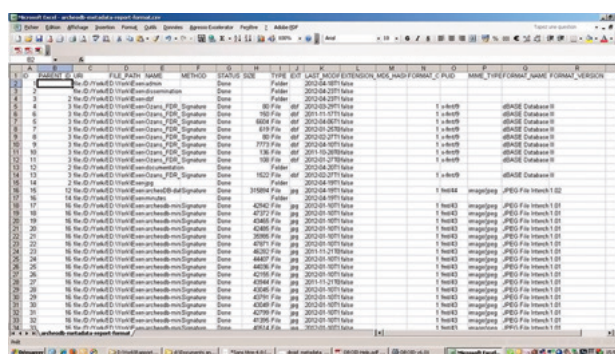


図36：メタデータを Excel の表 (スプレッドシート) で表示

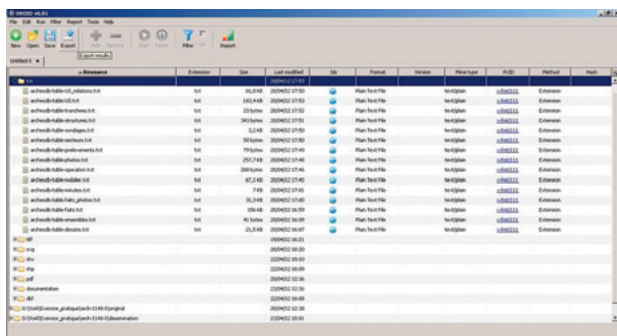


図33：取得したメタデータの保存

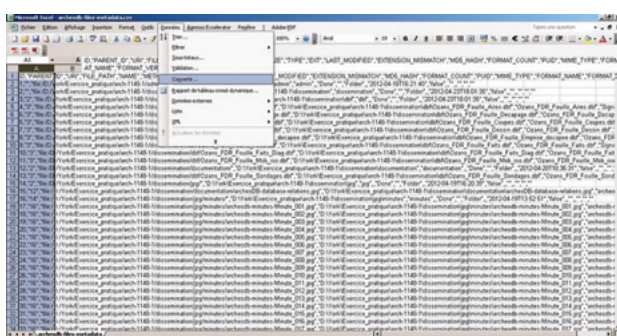


図34：メタデータを CSV 形式で開く

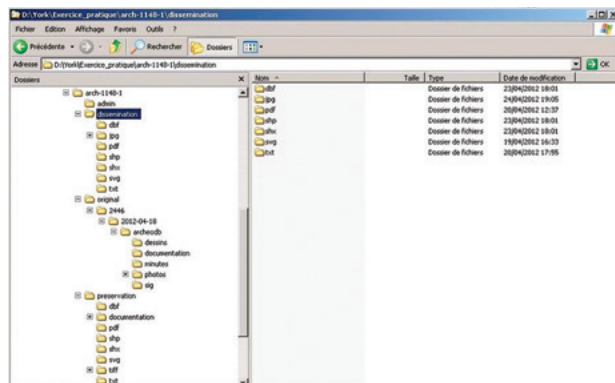


図37：アーカイブコレクションのツリー構造

- ・ 図面 = Adobe Illustrator による現地調査のモデル。
- ・ GIS = GIS プロジェクトに関連するすべてのファイル。
- ・ 記録時刻 = 調査現場での記録時刻をオリジナルのフォーマットでスキャンしたもの。
- ・ 文書化 = 作成者によるデータベースに関連した文書 (PDF ファイル形式)。
- ・ 写真 = 現場で撮影した写真のサムネイル (オリジナルフォーマット) [16]
- ・ ArcheoDB\_v1.3.20.mdb = Microsoft Access フォーマットの ArchéoDB データベース。

### 「保存 (preservation)」フォルダ

第二階層のこのフォルダは、保存用のファイルで構成され、ADS が管理するサーバーに保存されます。ファイル拡張子名をつけられた多くのサブフォルダが含まれています。またアーカイブには、ドキュメント専用のサブフォルダが必要です。我々の場合、以下の 8 つのサブフォルダを作成しました。

- ・ 文書化 (documentation) = データベースを文書化するもの (ODT ファイル形式のすべてのメタデータファイル、データベースのリレーションシップ・スキーマ、フォームのスクリーンショット、データベースの作者が与えた情報、すべて TIFF ファイル形式 [17])
- ・ シェープファイル (dbf, shp および shx) = シェープファイルを構成する 3 種類のファイルに対応する 3 つのサブフォルダ (データベースに伴う GIS)
- ・ PDF (pdf) = インベントリのバックアップ (データベースによって形作られ、生成されたレポート、PDF/A 1b に保存)。
- ・ SVG (svg) = ベクトルファイル (フィールド調査のテンプレート) のバックアップフォーマット。
- ・ TIFF (tiff) = 保存のために選ばれた画像 (写真、発掘調査現場での撮影時刻のスキャン) のバックアップフォーマット。
- ・ テキスト (txt) = データベースのテーブルのバックアップフォーマット。

### 「公開 (dissemination)」フォルダ

第二階層のこのフォルダは、ADS のウェブサイトでの情報発信のために保存されたファイルで構成されています。このフォルダは、ファイルの拡張子に対応した多くのサブフォルダで構成されています。

我々の例では、以下の 7 つのサブフォルダを作成しました。

- ・ dbf, shp et shx = シェープファイル (データベースに関連付けられた GIS) を構成する 3 種類のファイルに対応する 3 つのサブフォルダ。
- ・ jpg = イメージ (写真、フィールド議事録のスキャン、データベースのリレーションシップ・スキーマ) の普及のために選ばれたバックアップフォーマット。
- ・ pdf = インベントリのバックアップ (データベースによって形作られ、生成されたレポート、PDF/A 1b に保存)。
- ・ svg = ベクトルファイルのバックアップフォーマット (フィールドサーベイのテンプレート)。
- ・ txt = データベースのテーブルのバックアップフォーマット。

- [1] The recording was originally done on the field using Excel files to the “facts” and “stratigraphic units”.
- [2] <http://guides.archaeologydataservice.ac.uk/g2gp/Main>
- [3] The delimiter used preferentially by ADS is the comma. This colliding with decimal numbers present in the database (numeric fields are not distinguished as the text fields in quotation marks), it was decided to use that other delimiter also tolerated by ADS. Note that the problem of the use of the comma does not arise for English system because it is the point that acts as a separator in decimal numbers.
- [4] The treatment of the original files here and in the next section (processing of associated files), satisfies the constraints of long-term archiving but also of dissemination online. The formats used are those recommended in the various sections of Guides of Good Practice, available online at the following address: <http://guides.archaeologydataservice.ac.uk/>.
- [5] The captioned or named fields with special characters and accents are not taken into account when exporting to .txt format. About the “captions”, the simplest way is to remove them when they are given (useless information as part of export).
- [6] This operation can be performed only by those with Adobe Acrobat Pro. If the applicant does not have this application, the structure ensuring that archiving will do it.
- [7] If we had several drawings to treated, we could proceed like for the images, with a simultaneous renaming of files using the application XnView.
- [8] <http://www.iso.org/iso/fr/home.htm>
- [9] This method is usable only if Adobe Acrobat Pro (version 8 or higher) is installed on the computer. The conversion of an existing PDF to PDF/A can also be obtained from the application Adobe Acrobat but the procedure is more complicated and less effective. ADS preferably uses and recommends the application PDFTRON for the treatment PDF files: <http://www.pdftron.com/>.
- [10] Adobe Acrobat Pro automatically detects here with what version of PDF / A supports the current file (1a and / or 1b).
- [11] ADS Guidelines for Depositors : <http://archaeologydataservice.ac.uk/advice/guidelinesForDepositors>
- [12] The application DROID is currently in use by ADS to check and possibly complete the metadata files received. It is freely downloadable at the following address: <http://droid.sourceforge.net/>.
- [13] Metadata representing columns and archives files representing the lines (one line per file format).
- [14] As part of this exercise we used the metadata described in the sections “Documents and Texts”, “Databases and Spreadsheets”, “Raster Images”, “Vector Images” and “GIS” (<http://guides.archaeologydataservice.ac.uk/>).
- [15] The archive of a collection must be organized in a very precise tree and particular attention should be paid on folders and files names that compose it.
- [16] The author of the database did not sent us the photos in their original size for size problems, so we treated here the thumbnails (used to display the forms in the

database) as if s' were the original photos to archive.  
[17] Documentation describing the database provided by the author in PDF format has been converted in TIFF

file format because the original version of the file was not compatible for conversion to PDF/A.

## ラスター画像

### セクション1. ラスター画像概説

#### 1.1 ラスターイメージとは何か？

ラスター画像は、考古学アーカイブでは一般的なものであり、さまざまな異なるプロセスの産物や構成要素として存在しますが、基本的に構成は同じ、すなわち一定のサイズ／解像度を持つピクセルの行列からなる画像です。考古学において、ラスター画像は、デジタル写真、スキャン、図面などオリジナル取得データから、物理探査データのプロットやGISレイアウトからの画像などの出力、成果物に至るまで、さまざまなプロセスで生成されます。

このガイドでは、考古学調査の一環として作成される一般的なラスター画像の種類を網羅することを目的としており、それには以下のようなものが含まれます。

- ・デジタル写真
- ・航空写真
- ・文書のスキャン
- ・ベクター・アプリケーションのスクリーンショットや出力
- ・イラストやポスターなどのオリジナル画像

本ガイドでは、いくつかの技術的な説明を行いますが、解像度、色空間、ビット深度などラスター画像の詳細に関する説明は行いません。また、既存のガイド、特に JISC Digital Mediaの「静止画に関するアドバイス」[1] や、AHDSビットマップ（ラスター）画像保存ハンドブック [2] で十分に説明されています。同様に、ベクター画像については、「ベクター画像」の章で詳しく説明されています。ラスター画像、ベクター画像ともに、大規模なプロジェクトのワークフローに共通する要素であり、CAD、GIS、バーチャルリアリティなどの技術に特化したガイドにもリンクされています。

#### 1.2 現在の問題点と懸念事項

デジタル・ラスター画像の最も明白な問題の一つは、画像を作成・保存するためのフォーマットが多岐にわたることです。ラスター画像のフォーマットは、個々の機能や性能が大きく異なり、特定のソフトウェア用の商用（プロプライエタリ）フォーマットから、オープンな標準形式まで、あらゆる種類のフォーマットが存在します。また、圧縮率（可逆／不可逆）、色深度、透過性のサポート、メタデータの埋め込みなど、各ファイルフォーマットが持つ個々の機能の範囲も重要な要素であり、データの作成段階や長期保存の際には、作成する画像に適したファイルフォーマットを選択することが重要です。また、プロジェクトのワークフローによっては、使用方法に応じて、プロジェクトのさまざまな段階で画像の

フォーマットを変更することがあります。このような場合、各フォーマットがどのような機能やメタデータをサポートしているのか、またフォーマットの移行時に何が失われる可能性があるのかを認識しておくことも重要です。

- [1] <http://www.jiscdigitalmedia.ac.uk/stillimages/>
- [2] <http://www.ahds.ac.uk/preservation/ahds-preservation-documents.htm>

### セクション2. ラスター画像の作成

#### 2.1 一般的な検討事項

ラスターイメージは、ソースや用途によって大きく異なりますが、いくつかの特徴があるので、ファイルを作成および使用する際に考慮する必要があります。多くのファイルタイプと同様に、これらの要素の正確な設定を指定することは実際には不可能なので、プロジェクト全体の中で考慮され、目的に合ったレベルで使用すべきでしょう。

##### 解像度

解像度とは、基本的に画像の細密度をピクセル数で表したものです（例：1インチあたりのピクセル数（ppi）、1インチあたりのドット数やサンプル数（ddi/spi）など）[1]。解像度が高ければ高いほど、画像細部の特徴をより多く捉えることができますが、その結果、ファイルサイズが大きくなります。画像の解像度は、すべてのラスター画像にとって重要な検討項目であり、作業内容に応じて適切な解像度を選択する必要があります。解像度が高くなるとファイルサイズも大きくなるため、画像に求められる細密度と作成するファイルサイズのバランスをとることが重要です。

##### ビット深度

ビット深度（または色深度）は、各ピクセルで使用される色情報のレベルを意味します。画像のビット深度は、2ビット（白黒）から8ビット（通常グレースケール）、24ビット（標準カラー）まであります。解像度と同様に、作成する画像との関係でビット深度を検討し、必要な情報を取り込みつつ、ファイルサイズを最小限に抑えられるものを選ぶことが重要です。

##### 色空間

色空間とは、ビット深度に加えて、画像に使用されているカラーシステムまたはモデルを指します。一般的なモデルとしては、カラー画像ではRGBとCMYK、白黒やグレースケール画像では白黒二値とグレースケールがあります。画像に最

適なシステムを使用するだけでなく、主に画面表示用に使用される RGB システムと、印刷用の CMYK システムには大きな違いがあります。RGB システムは、CMYK システムよりも多くの色の組み合わせを保持することができるため、印刷時には RGB ベースの画像が正確に印刷されない可能性があることを認識しておくことが重要です。

## 圧縮

ファイルフォーマットの圧縮については、「デジタルデータ作成計画」の序章で簡単に説明しました。画像の圧縮には、GIF や PNG のような可逆タイプのファイル形式と、JPG のように（圧縮によりデータの一部が）破棄される非可逆タイプのファイル形式があります。TIFF や PNG のように、圧縮を行わずにデータを保存できる形式もあります。データを作成する際には、どのような場合に圧縮が行われるのか（例：カメラで JPG 画像を撮影する場合）、またどの程度のレベルで圧縮が行われるのかを認識することが重要です。画像圧縮については、英国国立公文書館が作成した「Digital Preservation Guidance Note 5: Image Compression」(Brown 2009) で詳しく説明されています。

## 透過性

画像の透過性は、大部分のベクター形式でサポートされていますが、一部のラスター形式（TIFF、PNG、GIF など）ではサポートされていません。したがって、アルファチャンネルや透過色を使って画像を透過させた場合、その機能が他のフォーマットや、ワークフローの中の異なる作業段階ではサポートされていない可能性があることに注意する必要があります。

## 画像のレイヤー化

画像内の構成要素を重ねる機能は、多くの一般的な画像編集ソフト（Photoshop や Photo-Paint など）では一般的な機能ですが、ラスター画像形式ではサポートされていません（レイヤーは上から下に向かって結合されます）。ラスター形式で保存すると、レイヤーは失われてしまうことに注意する必要があります。

## 事例

上記の要素は、作成されるラスター画像の種類によって意味が異なります。例えば、デジタル写真の場合、最近のカメラ

ラでは、さまざまな解像度の RAW（非圧縮）または JPG 形式で撮影することができます。そのため、被写体に合わせて最適なフォーマットを選択する必要があります。また、多くのカメラにはモノクロやグレースケールのモードが用意されており、撮影時にこれらを使用するか、後からソフトウェアでこれらのモードに変換するかを選択する必要があります。また、デジタルカメラは一般的に RGB の色空間で画像を作成するため、後に出版物に組み込む場合、ファイルを確実に CMYK 形式に変換しなければならない可能性があります。同様に、ドキュメント・スキャナーでは、生成されるデータセットに適した解像度、ビット深度、カラー空間を選択することが重要です。スキャンされたファイルの多くは、PDF ファイルへなど、後で変換が行われますが、その際にデータがダウンサンプルされる（つまり、解像度が下がる）ことがよくあります。このような場合、元のファイルがダウンサンプリングに適した解像度であること、また、重要と思われる場合は、元のファイルを最終版として維持することが重要です。

ベクター編集ソフトからエクスポートされたデータ（GIS や CAD のスクリーンショットなど）や、Adobe Photoshop や Illustrator などのアプリケーションで作成されたオリジナルデータの場合、ベクターベースであるがゆえに固有の解像度を持たない（または、ラスターファイルが重ねられている場合には解像度が変化する）ことがよくあります。このような場合、ラスターファイルを作成すると、設定された解像度で画像が固定され、結果としてファイル内の個別のエンティティやレイヤーが失われてしまいます。したがって、画像の用途に応じて適切な解像度を選択してファイルを出力するとともに、さまざまな機能を保存する必要があると考えられる場合には、可能であればオリジナルファイルを別の経路で保存することが重要になります。

また、画像を作成する際には、非可逆圧縮された画像を編集・保存（再圧縮）することを繰り返すことで、徐々に画質が劣化していく「世代間損失」（「デジタルデータ作成の計画」参照）に留意する必要があります。

## 2.2 ファイル形式

以下の表は、現在使用されている一般的なラスター画像フォーマットの概要です。

フォーマット	プロパティ／技術	説明	推奨
.tif / .tiff	非圧縮ベースライン TIFF v.6	事実上の標準であり、画像のアーカイブ版を保存するために非圧縮形式で広く使用されている。TIFF フォーマットは柔軟性があり、LZW [2] 圧縮、複数ページ、メタデータの埋め込みなど、多くのオプションに対応しているため、データ作成者は自分が作成するタイプを明確にしておく必要がある。TIFF フォーマットは、EXIF メタデータの埋め込みに加えて、ジオリファレンスのための GeoTIFF メタデータの埋め込みにも対応している。	一般的には、非圧縮のベースライン v.6 が保存目的のための唯一の TIFF であると考えられているが、GeoTIFF や TIFF/EP（いずれも特殊なメタデータを含む）などの拡張機能も同様に受け入れられ、この規格に基づいている。先に述べた圧縮の問題に加えて、TIFF で使用されている LZW アルゴリズムはプロプライエタリなソフトウェアに基づいており、これを実装したファイルは長期保存には適さないと考えられている。



フォーマット	プロパティ/技術	説明	推奨
.png	Portable Network Graphics, ISO 規格であり、W3C がサポートしている [3]。	GIF フォーマットに代わるものとして設計された PNG は、32 ビットの色深度、可逆圧縮、アルファチャンネル (透明度) のサポート、その他多くの機能を備えている。しかし、PNG は標準的な EXIF メタデータをサポートしておらず、また、主にインターネット用に設計されたフォーマットであるため、RGB 以外の色空間をサポートしていない。	PNG は可逆圧縮フォーマットの選択肢となっている。画像の長期保存には TIFF が適しているが、PNG フォーマットには多くの特徴があり、可逆圧縮が必要な場合など、可能な限り GIF フォーマットよりも優先して使用されるべきである。このフォーマットには、JPG に比べて目立たないなどの利点があるが、デジタル写真での使用は推奨されない。
.jpg / .jpeg	Joint Photographic Expert Group によって開発された ISO 標準形式	JPG フォーマットは、主に写真や絵画のように滑らかに変化する階調の画像用に設計されている。このフォーマットは、32 ビットの色深度と、非常に効率的な非可逆圧縮アルゴリズムを備えている。また、JPEG では EXIF や IPTC のメタデータをファイルに埋め込むことができる。	JPG ファイルは、通常、TIF や PNG に比べて大幅にサイズが小さくなるが、非可逆圧縮を使用しているため、長期保存には適していない。圧縮が必要な場合は、データ作成者は可逆圧縮の JPEG2000 フォーマットを使用することを推奨する。
.jp2 / .jpx	JPEG2000 は、ISO 標準規格であり、JPG フォーマットの代わりとなるもの。	2000 年 12 月にリリースされて以来、JPEG2000 の普及は遅れており、ブラウザでのサポートも不十分である。このフォーマットは、ファイルサイズと圧縮による変化・毀損の最小化を目的とした、高パフォーマンスな可逆圧縮を提供する。このフォーマットは標準的な JPG とは異なり、XML (JPX フォーマット) を使用してファイル内にメタデータを格納し、IPTC、GML、ダブリンコアのメタデータ要素を含むことができる [4]。	このフォーマットは、可逆圧縮の使用と拡張性により保存のための有効な形式として広く検討されている。Wellcome Library (Buckley 2009)、University of Connecticut (Lowe & Bennett 2009)、Digital Preservation Coalition (Buckley 2008) などのさまざまな機関や組織がここ数年で報告書を作成し、JPEG2000 が保存用フォーマットとしてますます普及することを示唆している。
.gif	Graphics Interchange Format, CompuServe によって開発された商用 (プロプライエタリ) 形式	Web 上で静止画と動画の両方に使用されている GIF フォーマットは、可逆圧縮が可能だが、パレットが限られており (8 ビット / 256 色)、メタデータを埋め込むオプションも限られている。	現在では、より最新のフォーマット (例: PNG) に取って代わられているが、GIF フォーマットはまだ広く使用されている。圧縮が必要な場合は PNG フォーマットを、長期保存の場合は TIFF を使用することを推奨する。
.bmp	Bit-Mapped Graphics Format、Microsoft が開発したもの。	多くの古い Windows アプリケーションで使用されている BMP フォーマットは、GIF に似ており、シンプルなグラフィックに最適なフォーマットである。BMP 形式はオプションで圧縮機能を備えているが、GIF ほど効率的ではない。BMP は、メタデータを埋め込むためのオプションが限られている。	作業用フォーマットや長期保存用フォーマットとしては推奨できない。
.psd	Adobe Photoshop のドキュメントファイルで、Adobe 社が所有する商用 (プロプライエタリ) 形式。主に画像の作成や編集に使用される。	PSD フォーマットは非常に柔軟性が高く、「業界標準」の画像ソリューションとして注目されている。また、EXIF、IPTC、XMP などのメタデータにも対応している。PSD は圧縮機能が限られているため、ファイルが大きくなる傾向がある。	PSD フォーマットは、ファイルを作成・編集するのに適したフォーマットだが、商用 (プロプライエタリ) でクローズドな性質を持つため、サードパーティによるサポートが限られており、長期保存には適していない。ファイルのコピーを PSD 形式で保存しておけば、将来的に編集することができるが、長期的なアクセスのためには、画像を圧縮されていないオープンな形式 (TIFF など) に移行することを推奨する。

フォーマット	プロパティ/技術	説明	推奨
.cpt	Corel Photo-Paint イメージ、Corel 社が所有する商用(プロプライエタリ)形式です。主に画像の作成や編集に使用される。	CPT は、Corel 社の Photo-Paint ソフトウェアのネイティブフォーマットであり、Adobe Photoshop の主な競合製品として、同様に幅広い機能を提供している。	PSD フォーマットと同様に、CPT は画像の作成および編集に最適です。将来の編集のためにこのフォーマットでコピーを保存することもできるが、Corel 社のソフトウェアに非常に特化しているため、画像のコピーは TIFF などの非圧縮フォーマットで保存する必要がある。
.dng	Adobe Digital Negative format Adobe 社が開発したオープンフォーマット。	TIFF/EP フォーマットをベースにして、それに準拠した DNG は、デジタルカメラで生成された RAW ファイルを保存するためのオープンなドキュメント・アーカイブ形式である。Adobe 社は、複数のカメラモデルやメーカーの RAW ファイルを扱う際に、より効率的なワークフローを可能にする単一の RAW 処理ソリューションとなることを目指して DNG を開発した(既存の RAW 画像フォーマットを DNG に変換するコンバーターは、Adobe 社の Web サイトから無料でダウンロードできる [5])。DNG フォーマットは、オリジナルの RAW ファイルからすべての画像タグ(EXIF および IPTC)を読み取り、DNG 画像に保存することができ、さらに、XMP を介してその他のメタデータを埋め込むこともサポートしている [6]。	DNG は、画像データの長期保存に適している。
raw (各種拡張子)	Raw のビットマップ・ファイルは、通常、ベンダー(デジタルカメラ製造企業)固有のものであり、商用(プロプライエタリ)形式。	Raw ファイルは、デジタルカメラ(場合によってはデジタルスキャナー)で直接作成された、さまざまな形式の未処理のビットマップ・ファイルである。標準化されていないため、Raw ファイルとして全体をまとめて考えることはほとんど不可能である。多くのファイルは非圧縮ですが、非可逆圧縮または可逆圧縮の両方を使用している(ユーザーが選択できる)ものもある。また、標準化されていないため、多くのファイルを開くには特定のソフトウェアアプリケーションが必要となる。	Raw 形式は長期保存には不向きであり、データ作成者は可能な限り TIFF や DNG などの標準形式に変換することを推奨する。

- [1] JISC Digital Media Glossary entry for 'Resolution'  
<http://www.jiscdigitalmedia.ac.uk/glossary/#r>
- [2] The Lempel-Ziv-Welch lossless data compression algorithm.
- [3] <http://www.libpng.org/pub/png/>
- [4] <http://www.jpeg.org/jpeg2000/metadata.html>
- [5] <http://www.adobe.com/products/dng/index.html>
- [6] <http://www.adobe.com/products/xmp/index.html>

### セクション3. ラスターイメージのアーカイブ

#### 3.1 アーカイブする内容の決定

何をアーカイブするかは、画像が作成・利用される個々のワークフローによって異なります。しかし一般的には、可能

ならばオリジナルデータを非圧縮のオープンフォーマットでアーカイブすることが望ましいとされています。

- デジタルカメラは、保存できるファイルの選択肢は限られていることが多いでしょう。デジタル写真画像は、オリジナル(未加工)のJPG(できれば非圧縮のTIFF)、またはDNG形式のRAWファイル、TIFF形式のいずれかで保存することになります。
- ドキュメント・スキャナーからの画像の場合、通常、画像をスキャンした後に保存するフォーマットを選択する必要があります。スキャンした画像は最終的なフォーマットに関わらず、処理を行う前に、TIFFなどの非圧縮(または可逆圧縮)のオープンフォーマットで保存することをお勧めします。
- 画像処理ソフトで作成された画像は、ユーザーが保存形

式を選択する必要がありますが、ほとんどのソフトウェアは幅広いフォーマットに対応しています。GISやCADなどのソフトウェアから出力されたファイルの場合、ファイル形式の選択肢はより限られたものになるでしょう。いずれの場合も、非圧縮形式が望ましいですが、それができない場合は、他のパッケージでファイルを適切なアーカイブ形式に変換することができます。

前節のフォーマット一覧表で強調したように、多くのファイル、特にデジタル写真には、ファイルのドキュメンテーションや再利用に役立つメタデータがファイルフォーマット内にパッケージされています。メタデータがある時、ほとんどの場合でそれを保持し、アーカイブすることは有効です。

### 3.2 どのようにアーカイブするかを決定

多くの場合、上述したように、デジタル画像をアーカイブするには安定した保存形式に直接保存するか、変換する必要があります（望ましいフォーマットは下の表に記載されています）。画像編集ソフト（PhotoshopやPhoto-Paintなど）で編集画像を作成した場合、最終バージョンのファイルを「結合」または「フラット化」した編集画像として保存するだけでなく、個々の構成要素を個別のアーカイブファイルとして保存することにも価値があります。このような戦略は、GIFなどの単純なアニメーション画像ファイルにも適用され、各フレームを保存用フォーマットで個別のファイルにエクスポートします。

#### 重要なプロパティ

ラスター画像の重要なプロパティについては、InSPECT Significant Properties Testing Report on Raster Images (Montague 2009) で詳細に説明されていますが、ここでは以下に概要を示します（前のセクションで詳細に説明されて

います）。

- ・ **画像サイズと解像度**：ファイル形式を変換する際には、保存用のファイル形式との間で解像度と画像サイズを一致させる必要があります。またファイルを新しいフォーマットに変換する際には、画像に非可逆圧縮を適用しないことが重要です。
- ・ **ビット深度と色空間**：ファイル形式を変換する際には、元の画像のビット深度と色空間が保存用のファイル形式でもサポートされていることと、変換時に画像が劣化しないことを確認する必要があります。

これらのプロパティはすべての画像形式の構成要素ですが、ファイルを保存用形式に変換する際には、プロパティが同一であること、または同じ値を維持することの確認が重要です。

さらに、EXIFやIPTCなどの埋め込まれたメタデータも、場合によっては画像の重要なプロパティとみなすことができるので、それらのメタデータがある場合には、ファイルと一緒に保存するか、別のプレーンテキストまたは区切りテキストまたはXMLファイルにエクスポートして、画像と一緒に保存する必要があります。JPEGのEXIFをTIFFのタグ構造の中に保存することは可能ですが、後の移行時における消失や破損のリスクを回避し、メタデータに簡単にアクセスできるようにするためには、別のファイルに保存することが望ましいでしょう。EXIFフィールドの抽出は比較的簡単で、多くのフリーツールが提供されています。

#### ファイルフォーマット

デジタル・ラスター画像の長期保存には、下表のフォーマットを推奨します。

保存形式	推奨
.tif / .tiff	TIFF 6.0は、デジタル・ラスター画像の保存形式として広く受け入れられており、EXIFメタデータにも対応している。また、TIFF/EPやGeoTIFFなどの拡張フォーマットも同様に有効な保存形式である。
.dng	EXIF、IPTC、XMPメタデータをサポートするTIFF/EP規格のオープンエクステンションであるAdobe DNGフォーマットは、生の画像データ（主にデジタル写真）を保存するための規格として急速に受け入れられている。

代替ファイル、主にアーカイブサイズを縮小するための圧縮を用いたファイルが最近提案され、調査されています（Gillesse et al. 2008 参照）。PNGやJPEG2000などのファイルをアーカイブ形式として使用する余地があるかもしれませんが、これらは圧縮を使用しているため、本ガイドでは推奨していません。

### 3.3 メタデータと文書化

ラスター画像のメタデータについては、“Guidelines for Handling Image Metadata” (Metadata Working Group 2009) や、JISC Digital Media のページにある“Metadata and Digital Images” [1] で詳しく述べられています。JISC Digital Mediaでも紹介されているように、画像のメタデータには、構造的なもの、記述的なもの、管理的なもの、技術的なものなどさまざまな種類があり、画像ファイルの中に埋め込まれているもの（EXIF、IPTC、XMPなど）と、別のファ

イル（データベース、スプレッドシート、テキスト文書）として保存されているものがあります。

#### 技術的メタデータ

一般的にEXIF形式で埋め込まれた技術的メタデータは、一般的に画像ファイルの作成に関連しており、画像の撮影に使用されたさまざまなカメラの設定を記録しています。EXIFのような自動生成されたメタデータを重要なプロパティとみなすべきかどうかについて多くの議論があります。このようなメタデータが、データ作成者によって意識的に使用され、本質的に重要な意味を持つ可能性は当然ありますが、一般的にはそれほどでもないでしょう。また特定のメタデータ要素（日付など）は、意図的に使用・設定されていない場合や、誤った情報を含んでいる可能性があります。しかし画像にメタデータを埋め込むことは一般的になってきており、画像を変換・複製する際には、このメタデータを保持すべきかどうかを判

断する必要があります。一般的には、埋め込まれたメタデータに価値があると考えられる場合は、それを保存すべきです。

その他の技術的要素は、埋め込まれたメタデータセットに含まれることもあります。データ作成者やアーカイブの要求が異なるため、より一般的には別個のファイル（テキストファイルやスプレッドシートなど）として記録されます。一般的な技術的メタデータの要素としては、撮影・取り込み機器の情報、ソフトウェア、編集履歴などが挙げられます。

### その他のメタデータ

技術的なメタデータ以外にも、IPTC や XML などのフォーマットに埋め込まれていたり、別のファイルに記録されていたりする場合もあります。このようなメタデータは、主に説明的・管理的なデータを記録することを目的としており、画像の被写体、制作者、コンテンツを記録します。IPTC Photo

Metadata set (IPTC 2010) は、国際新聞電気通信評議会 (International Press Telecommunications Council) がニュースやストックフォトのために作成したもので、画像の内容、関連する場所、被写体、権利などをカバーする要素が含まれています。さらにデータ作成者やアーカイブは、ダブリンコア [2]、MIX [3]、PREMIS [4]、ANSI/NISO Z39.87 [5] などの既存のメタデータ規格を参考にして、作成・保存する画像に最適なメタデータ要素を指定することができます。

以下は、提案されたメタデータ要素であり、上述の情報源（すなわち、AHDS Bitmap handbook、Metadata Working Group guidelines、ダブリンコア）から集められたものです。これらの要素の多くは、コレクション全体（例：写真のセット）に適用することができ、ファイルごとに繰り返す必要がない（訳注：個別ファイルについて文書化しなくても良い）ことに注意することが重要です。

要素	説明
識別子	画像ファイル名 例:survey01.tif
タイトル／キャプション	画像のタイトルまたは適切なキャプション
説明	画像の説明
作成	画像作成者の名前
日付	画像が作成された日付
権利	著作権またはその他の権利の詳細および所有者の詳細
キーワード	キーワード:時期、遺跡、遺構などの名称。適切なシソーラスがある場合はそれを使用する
位置情報	画像の位置情報。可能であれば、標準化されたフォーマット（緯度／経度など）や、適切なシソーラス (Getty Thesaurus of Geographic Names [6] など) のキーワードを使用してください。

### 追加の基本的な技術メタデータ

要素	説明
ファイルフォーマットとバージョン	例:TIFF 6.0
ファイルサイズ	ファイルのサイズ(単位:バイト)
空間解像度	画像の解像度をインチ単位ピクセル数(ppi)で表したものの
サイズ	画像サイズをピクセルで表したもの 例:400 × 700px
色空間	画像に使用されている色空間 例:RGB やグレースケール
ビット数	例:24 ビットまたは8ビット
取得機器	カメラのメーカーとモデル、またはスキャナーの詳細
ソフトウェア	画像のキャプチャーに使用したソフトウェア(一般的にはスキャン画像に使用) 例: Adobe Photoshop CS3

埋め込みメタデータを使用する際には、それがプロジェクトレベルのメタデータで使用されているかどうかを明示すること、埋め込みメタデータが各ファイルでどのようにサポートされているか、データの移行がどのように影響するかを認識することが重要です。例えば、EXIF メタデータは JPEG や TIFF ではサポートされていますが、JPEG2000 (LaBarca 2010) や PNG フォーマットでは (少なくとも同じ方法では) サポートされていません。また IPTC メタデータは、現状の Photo Metadata セットと従来の IIM セットの 2 種類が存在し、XMP フォーマットでもサポートされています。

### 3.4 アーカイブの構造化

画像は論理的に構造化し、個々のプロジェクトに適した方法で作成する必要があります。ファイル構造は、特定の遺跡や遺構、あるいはフィールドワークや写真測定の個別事例を反映するでしょう。

埋め込みメタデータを保存する場合は、シンプルなテキストまたは XML 構造で抽出して保存し、画像ごとにメタデータのテキストファイルを作成することをお勧めします。標準的な命名方法としては、#image\_filename#.meta が提案され

ており、これらのファイルは、保存された画像と一緒に文書化フォルダに保存されます。IPTCやXMPなど、他の埋め込みメタデータもXML構造に抽出できるかもしれませんが、また、使用されている要素が少ない場合や重要と思われる場合には、メタデータをスプレッドシートやデータベースに簡単に保存することもできます。

#### 参考文献

[1] <http://www.jiscdigitalmedia.ac.uk/stillimages/>

- advice/metadata-and-digital-images/  
[2] <http://dublincore.org/>  
[3] <http://www.loc.gov/standards/mix/>  
[4] <http://www.loc.gov/standards/premis/>  
[5] ANSI/NISO Z39.87 Data Dictionary - Technical Metadata for Digital Still Images [http://www.niso.org/kst/reports/standards?step=2&gid=None&project\\_key=b897b0cf3e2ee526252d9f830207b3cc9f3b6c2c](http://www.niso.org/kst/reports/standards?step=2&gid=None&project_key=b897b0cf3e2ee526252d9f830207b3cc9f3b6c2c)  
[6] [http://www.getty.edu/research/conducting\\_research/vocabularies/tgn/](http://www.getty.edu/research/conducting_research/vocabularies/tgn/)

## ベクター画像

### セクション1. ベクター画像の紹介

#### 1.1 考古学におけるベクター画像

##### ベクター画像とは？

ベクター画像は、ラスター画像とは対照的に、オブジェクトを配列されたピクセルのグリッドとしてではなく、幾何学的な実体、つまりベクトルを有するオブジェクトとして表現します。ベクトルを有するオブジェクトには、直線、円、長方形、曲線などがあり、これらはすべてポイントとパスで結ばれています。これらのオブジェクトは、座標や数式によって定義されるため、品質を損なうことなく真の意味での拡大縮小が可能です。ベクター画像には2次元や3次元の形状が含まれており、多くのファイルにはベクターとラスターの両方のデータが含まれています。

ベクター画像は考古学では比較的一般的なもので、さまざまな場面で作成することがあります。最も一般的な例は2Dの画像やイラストで、通常は出版目的で作成され、CAD（例：考古学的特徴や建物のイラスト）やGIS（サイトプランや分布）のデータセットからよく作成されます。これらの画像は、発掘調査やプロジェクトの報告書の一部として、PDFファイルに組み込まれることが多くあります。

本ガイドでは、CADやGISなどの個別技術に属さない単純な2Dベクター画像を対象としており、これら特定のアプリケーションで作成されたベクターファイルについては、別の章で扱っています。ベクター画像は、AHDS Digital Images Archiving Study (Anderson et al 2006) や、以下のJISC Digital Mediaドキュメントにも詳細が記載されていますので、詳細が必要な場合は、これらを参照してください。

- ・ベクター画像フォーマットの紹介 [1]
- ・ベクター描画ソフトウェア [2]
- ・ベクター画像イラストレーション用語集 [3]
- ・ベクター画像リソース [4]

- [1] <http://www.jiscdigitalmedia.ac.uk/stillimages/advice/introduction-to-the-vector-image-format/>  
[2] <http://www.jiscdigitalmedia.ac.uk/stillimages/advice/vector-drawing-software/>  
[3] <http://www.jiscdigitalmedia.ac.uk/stillimages/advice/vector-graphics-illustrated-glossary/>  
[4] <http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/vector-graphics-resources/>

### セクション2. ベクター画像の作成

#### 2.1 一般的な検討事項

前節で述べたように、写真などの一般的なラスター画像とは異なり、ベクター画像の多くは、CADやGISなどの他のアプリケーションで作成・保持されたデータに由来していません（さらに、物理探査やレーザースキャンなどのさまざまなデータ収集技術に由来している場合もあります）。画像がCADファイルなどの他のデータセットから派生したものである場合、派生画像よりオリジナルファイルの保存を優先することをお勧めします。多くの場合、派生したベクター画像はプロジェクト報告書などの文書で説明のために頻繁に使用されるため、プロジェクトのデータセットの他の場所にも存在することがあるでしょう。

画像がベクター画像編集ソフトでネイティブに作成されている場合や、派生ファイル（説明文などの機能が追加されたイラストなど）それ自体が保存する価値があると考えられる場合には、本章のガイダンスに従うべきです。以下の表は、ベクター画像の作成に使用される一般的なフォーマットと、それらがファイルの長期保存に適しているかどうかを示しています。

## 2.2 ファイルフォーマット

フォーマット	特性/技術	説明	推奨事項
Adobe Illustrator (.ai)	商用(プロプライエタリ)の、主に 2D のレイヤー・ベースのラスター/ベクターフォーマット。	フォーマット自体は時間とともに変化しており、初期のバージョンは PostScript をベースにしていました。3D 機能は Illustrator CS で導入された。人気があり市場をリードしているが、このファイル形式は他のアプリケーションではあまりサポートされていない。	保存には適していない。
CorelDraw (.cdr)	商用(プロプライエタリ)のレイヤー・ベースの 2D フォーマット。	Illustrator の .ai ファイルフォーマットと同様に、.cdr にはラスターデータとベクターデータの両方を格納できる。	保存には適していない。
PostScript (.ps)	Adobe 社が開発したプログラミング言語で、デバイスに依存しない方法でテキストや画像の外観を記述する。	このフォーマットには、EPS や PDF (下記参照) などの派生フォーマットがあり、とくに PDF は徐々に Postscript に取って代わっている。	保存に適している可能性はあるが、Postscript はもはや一般的には使用されておらず、PDF/A のような新しい代替フォーマットの方がより強固な保存フォーマットとなっている。
Encapsulated PostScript (.eps, .epsf, .epsi)	基本的に EPS は、他の文書に埋め込むことができる、制限された自己完結型の PostScript ファイルである。	EPS には、ベクトルまたはビットマップのデータとフォントが含まれており、通常はプレビュー画像が含まれている。EPS のヘッダーにはバージョン番号が含まれていることが多く、プログラムによってはバージョン番号がないとファイルを拒否することがある。	Postscript と同様に、E より適切なフォーマットがあるため、保存には適していないと考えられる。
Portable Document Format (.pdf, .pdf/a, .pdf/e, pdf/x)	PDF は、Adobe によって開発された Postscript のサブセットである。	純粋にページ(版面)単位に限定されているが、.pdf には、特定の目的(アーカイブやエンジニアリングなど)のために設計された、いくつかの特殊なサブセットが組み込まれている。PDF フォーマットは、主にプラットフォームに依存しない文書共有のために設計されている。PDF のコンテンツは ASCII テキストの場合もあるが、ファイルは圧縮されているのが一般的で、バイナリ形式である。	PDF/A プロファイルファイルは、保存に適していると考えられるが、大部分を自己完結させるために必要な制限があるため、すべてのタイプのコンテンツに適しているわけではない。
Scalable Vector Graphics (.svg)	W3C [1] によって開発された、2D ベクトルグラフィックスを記述するための XML ベースのオープンスタンダードフォーマット。	多くの場合、SVG はウェブ上で特定の機能を提供するために Javascript と一緒に使用される。このような場合、コンテナ(すなわちウェブページ)と SVG はリンクされたファイルとして扱われ、両方とも保存するか、一緒に提示されなければならない。	保存に適している。
Computer Graphics Metafile (.cgm)	年代物の 2D ラスター/ベクターフォーマットで、文書化された標形式(ISO/IEC 8632[2])。	このフォーマットのサブセットである WebCGM は、Web での使用に適したフォーマットのバージョンとして W3C によって近年開発されている[3]。	ドキュメント化されたオープンな標準として、WebCGM は保存用フォーマットとしての可能性を秘めているが、SVG にほとんど取って代わられている。
Microsoft Windows Metafile (.wmf, .emf)	GDI(Graphics Device Interface) コマンドを使用して画像をレンダリングする、独自のベクトルベースのファイルフォーマット。	EMF は、32 ビットアーキテクチャ用の WMF 規格の拡張版である。	保存には適していない。
WordPerfect Graphics Metafile (.wpg)	ビットマップ、ベクター画像、EPS データを格納できる 2D グラフィックスのメタフォーマット。	WordPerfect 5.0 以前のバージョンでは、ビットマップ(ラスター)またはベクターイメージ画像のいずれかを保存できますが、両方を同時に保存することはできない。	保存には適していない。

フォーマット	特性/技術	説明	推奨事項
Macintosh PICT (.pict, .pic, .pct)	1984年にApple社が開発した、ビットマップ画像とベクター画像の両方を保存できるメタフォーマット。	このファイルには、画像を描画するためのQuickDrawコマンドがすべて含まれている。1つのビットマップのみを含むPICTファイルは、WindowsではQuickTime for Windowsを使ってサポートされている。	保存には適していない。
Macromedia Flash (.swf, .fla, .swd, .flv, .swc, .swt, .flp)	Web配信用に最適化された2Dベクターベースのアニメーション・フォーマット。	フォーマットの仕様は自由に利用できるが、一部の要素はまだ非公開とされている。Flashファイルは、純粋にベクター画像だけではなく、スクリプトやその他のマルチメディアを組み込み多様な機能を提供している。	保存には適していない。
Macromedia Freehand (.af)	ビットマップとベクターのハイブリッド・ファイルタイプを使用する普及したイラストレーション・パッケージ。	ファイル形式が大幅に変更されたため、初期のバージョンから移行する際に問題が発生することがある。2007年現在、Freehandの開発とサポートは終了しているが、Adobe Illustratorではファイルがサポートされている。	保存には適していない。
Micrografx Designer	.drw, .dsf	ベクトル/ビットマップ・ベースのプログラムで、芸術的な描画よりも技術的な描画を目的としている。	保存には適していない。

### その他のフォーマット

上記のファイル以外にも、データ作成者が知っておくべきベクター形式がいくつかあります。

上の表から除外されている最も明白なグループは、CADおよびGISアプリケーションに関連するベクター形式です。DXF、DWG、SHPなどのフォーマットは、考古学プロジェクトで頻繁に使用されており、本ガイドのCADおよびGISの章で扱われています。

Drawアプリケーションを含むOpen Officeスイートの人気が高まっていることから、データ作成者がDrawのODGフォーマットで画像を作成・保存する可能性が出てきました。このフォーマットはXMLベースでOASIS Open Document仕様の一部ですが、他のアプリケーションではあまりサポートされていません。しかし、DrawはSVGや他の多くのフォーマットへのエクスポートをサポートしており、現在のところ、ファイルの長期保存のためにはこれらのフォーマットでの保存が推奨されています。

- [1] <http://www.w3.org/Graphics/SVG/>
- [2] <http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html>
- [3] <http://www.w3.org/TR/2010/REC-webcgm21-20100301/>

## セクション3. ベクター画像のアーカイブ

### 3.1 アーカイブする内容の決定

セクション1で述べたように、考古学において、ベクター

画像はデータ処理アプリケーションからの出力として、あるいは高品質のイラストを作成するために、より大きなプロジェクトのワークフローの中の一段階として作成されることが多いでしょう。いずれの場合でも、これらのファイルは「最終成果物」のイメージを表しており、通常はプロジェクト報告書に含められ、たいていは印刷されたページのフォーマットになっています。さらに、ベクター画像をラスター化してJPGやTIFFファイルにしてからドキュメントに収録することもあります。したがってデータ作成者は、ワークフローにおけるベクター画像の重要性を評価し、対象となるファイルが特定のフォーマットで再利用される価値があるかどうかを決定することが重要です。画像がソース（CADなど）と派生（ラスター画像など）の両方の形式で存在する場合、ベクター画像は保持せず、アーカイブしないことを決定することもあり得ます。

多くのベクター画像アプリケーションでは、画像そのものに加えて、ファイル内にスクリプト（IllustratorではJavaScript、AppleScript、VBScript [1] など）を組み込むことができます。このようなファイル内のスクリプトは、作成または編集段階での作業を自動化または簡略化することを目的としており、ベクター画像の核となる構成要素とは考えられていません。このようなスクリプトを保存する価値があると判断した場合は、画像とは別に保存し、文書化する必要があります。

### 3.2 アーカイブ方法の決定

#### 重要なプロパティ

他のファイルタイプと同様に、ベクター画像をアーカイブする際には、ファイルの核となる重要なプロパティを特定し、選択したフォーマット内で維持することが不可欠です。JISCレポート「The Significant Properties of Vector Images」

(Coyne et al 2007) には、ベクター画像をアーカイブする際に識別・維持すべきプロパティがしっかりと記載されていますので、ここでは簡潔に紹介します。

ベクター画像のプロパティの意義は、画像のジオメトリ、ベクター・オブジェクトと相互の関係、ファイル内で使用されている装飾（色、線種、太さ）、およびファイルの移行やアーカイブにおいて表示の見た目を維持することにあります。ベクターファイルのアーカイブでは、以下の要素を確認する必要があります。

- ・ポイント、パス（オープンまたはクローズ）、オブジェクト（例：テキスト、参照または埋め込みオブジェクト）
- ・基本図形／プリミティブ（例：矩形、円、楕円、など）
- ・画像構造（例：オブジェクトのグループ化とレイヤー化）
- ・クリッピングとマスキング
- ・変換と座標系（存在する場合。これらは主にGISやCADのベクターファイルに含まれます）

これらの特定のベクター要素が同じフォーマットでファイル内に物理的に存在することを確かめるだけでなく、適用可能ないくつかのプロパティがあります。標準化された慣例が適用されている場合（特定の意味が与えられた線色や網掛けの種類）、これらのプロパティやフォーマットがアーカイブファイルで維持されることが不可欠になります。このようなプロパティには次のようなものがあります。

- ・線幅、線端、結合タイプ
- ・線の接合箇所の表示（マイターリミット：面取りか尖端処理か）
- ・網掛け（ダッシュパターンまたはハッチング）とオフセット

### ファイルフォーマット

保存形式	要件
.svg	xml ベースのオープン標準形式である SVG は、ベクター画像の長期保存に適したフォーマットである。
.pdf/a	.pdf/a ベクター画像の保存には適していますが、「最後の手段」として推奨される。このフォーマットはファイルのベクトル性を保持するが、再利用の可能性を単純な「表示のみ」に減らしてしまう。

### 3.3 メタデータと文書化

ベクター画像のメタデータについては、The Significant Properties of Vector Images (Coyne et al 2007) で詳しく説明されており、そこでは下表の要素が引用されています。また、JISC Digital Media のページでは、「Metadata and Digital Images」[2] について説明されています。以下の要素は、主

- ・色
- ・不透明度
- ・レンダリング（描画方法）
- ・内部の定義
- ・グラデーションとグラデーションの滑らかさ
- ・パターン
- ・フォントサイズ、ウェイト（太さ）、スタイルなどのテキスト属性

興味深いことにJISCのレポートでは、重要なプロパティの潜在的な階層が提案されているので、ファイルが保存フォーマットに簡単に移行できない場合、どのプロパティを優先すべきかを評価するために使用することができます (Coyne et al 2007, 41)。

残念ながら、すべての要素が正しい方法で描画（レンダリング）されていることを視覚的に評価する以外に、オリジナルのファイルと移行されたバージョンを比較するシンプルで客観的な方法はありません。ファイルを新しいフォーマットに移行する際には、目に見える要素の他に、目に見えないオブジェクトやレイヤーなどの隠れた情報をチェックし、それらがアーカイブに適しているかどうかを評価することも重要です。また、埋め込まれたラスターファイルや外部フォントなどのリンクデータが存在している場合がありますが、その場合、データの外観はファイルの移行時にも維持されなければなりません（訳注：リンクデータを埋め込むなどの対応）。ベクターファイルの保存がうまくいくかどうかは、オリジナルデータの性質、複雑さ、純粋なベクターデータだけが含まれているか、ラスターデータも含まれているか、などに大きく依存します。混合ファイルは、データタイプを適切な保存フォーマットに分離する必要があるでしょう。

にダブリンコア [3] の要素であり、画像の作成者が生成するのが最も効果的でしょう。さらに、いくつかの技術的な要素が提案されており（『The Significant Properties of Vector Images』に提示されている要素の一部に過ぎません）、データ作成者が任意に記入することができます。

要素	説明
識別子	画像の一意の識別子で、通常はファイル名
タイトル／キャプション	画像のタイトルまたは適切なキャプション
説明／目的	画像の説明
作成者	画像の作成者の名前



要素	説明
日付	作成日
権利	著作権またはその他の権利の詳細および所有者の詳細
キーワード	キーワード: 時期、遺跡、遺構名称など。適切なシソーラスが存在する場合はそれを使用する
場所/他のドキュメントとの関係	画像がどこで作成されたか、また(もしあるならば)他のファイルとどのように関連しているかの説明
装飾	文書に含まれていない場合、使用されているすべての装飾(色、レイヤー、線種、線の太さ)の説明

#### 追加の基本的な技術メタデータ

要素	説明
ファイルフォーマットとバージョン	例: SVG1.1
ファイルサイズ	ファイルのサイズ(単位: バイト)
ソフトウェア	画像作成に使用されたソフト(例: Adobe Illustrator)

### 3.4 アーカイブの構造化

ベクター画像は論理的に構造化し、個々のプロジェクトに適した方法で作成する必要があります。ファイル構造は、特定の遺跡や遺構、あるいはフィールドワークや写真測量の個別事例を反映するでしょう。メタデータを保存する場合は、簡単なテキスト(プレーンまたは区切り)またはXML構造で抽出し、保存画像とともに「ドキュメント」フォルダに保存することをお勧めします。

[1] <http://www.adobe.com/devnet/illustrator/scripting.html>

[2] <http://www.jiscdigitalmedia.ac.uk/stillimages/advice/metadata-and-digital-images/>

[3] <http://dublincore.org/>

#### 参考文献

- Anderson, S., Pringle, M., Eadie, M., Austin, T., Wilson, A., Polfreman, M. (2006) *Digital Images Archiving Study*. AHDS. [http://www.jisc.ac.uk/uploaded\\_documents/FinaldraftImagesArchivingStudy.pdf](http://www.jisc.ac.uk/uploaded_documents/FinaldraftImagesArchivingStudy.pdf)
- Coyne, M., Duce, D., Hopgood, B., Mallen, G., Stapleton, M. (2007) *The Significant Properties of Vector Images*. JISC. [http://www.jisc.ac.uk/media/documents/programmes/preservation/vector\\_images.pdf](http://www.jisc.ac.uk/media/documents/programmes/preservation/vector_images.pdf)

## デジタルビデオ

### セクション1. デジタルビデオ概説

#### 1.1 デジタルビデオ

デジタルビデオは、考古学と、測量、手続き、インタビューなど考古学調査に関連するイベントを記録する手段としてますます普及しています。1997年にInternet ArchaeologyがEVA会議で発表した論文[1]によると、以前はデジタルビデオを利用する考古学者はほとんどいませんでしたが、今日、状況は急速に進歩し、デジタルビデオを記録する機器が安価で簡便になっただけでなく、スチルカメラや携帯電話などの他の機器にも動画撮影機能が搭載されるようになりました。また、動画編集・配信アプリケーションも急速に普及し、特にYouTubeやFacebookなどのウェブ上での配信が盛んに行われています。その結果、デジタルビデオは、計画された成果であるかどうかにかかわらず、考古学プロジェクトの一般的な構成要素となっています。

#### どのように使われるのか?

デジタルビデオは、他のデータ収集技術に付随して、記録、補足するツールとして調査現場で使用されることが多いでしょう。特に、水中考古学者の間では、陸上に比べて遺跡へのアクセスが容易ではないため、デジタルビデオは調査プロジェクトの一般的な要素となっています。さらに、デジタルビデオは、3DモデリングやVRなど、さまざまなデータ収集・分析技術を用いたプロジェクトのアウトプットにもなります。

データの作成に関しては、キャプチャしたデジタルビデオをデジタルビデオ・テープやDVDに直接記録することが多いのですが、前述のように、スチルカメラや携帯電話などの他のデバイスで撮影したり、3Dモデルなど他のデータソースから取得することも可能です。次項で説明するように、作成方法によって、映像ファイルの種類、ファイルサイズ、品質が決まります。デジタルビデオ・テープやDVDを使用するカメラなどでは、特定の動画ファイルは作成されないことが

多く、その作成のためにはコンピュータでさらに処理する必要があります。また、スチルカメラなどでは、作成するファイルの種類に大きな制限がある場合もあるでしょう。

### アーカイブする理由

考古学プロジェクトで作成されたデジタルビデオをアーカイブする理由は、その目的によって大きく異なり、以下に述べる課題と照らし合わせて判断する必要があります。デジタルビデオを保存する理由は、ユニークなオリジナルデータ（写真のような他のフォーマットで記録されていないもの）が含まれている場合や、他のデータセットにとって貴重なサポートやドキュメントを提供する場合です。Wessex Wrecks on the Seabed プロジェクト [2] の場合、水中考古学関連のデジタルビデオは、ダイバーやROVが水中遺跡で観察したものを記録する重要なものであり、特にトラック・ログと関連している場合には重要です（VENUS Guide の Navigational and Positional data のセクションを参照）。将来のユーザーの多くは、編集されていないダイビングのフル動画を見たいとは思わないでしょうが、それを保存することは重要だと考えられます。デジタルビデオは、沈船遺跡の状態を評価し、時間経過に伴う損傷をモニタリングするツールとして活用できます。つまり、従来の紙の調査日誌のように、最新のビデオ日誌はプロジェクトの構成要素をまとめることができます。こうしたビデオは、歴史学の資料にもなります。

3Dモデリングやインタラクティブな3D環境の場合、環境内でのユーザーの様子や、環境やモデル自体のフライスルーを撮影したビデオは、他のフォーマットでは簡単に保存できない貴重なドキュメントとなります。さらに、ビデオによるフライスルーは、大規模な3Dデータセットを普及させ、ユーザーがその内容を手軽に評価するための手段となります。

## 1.2 現状の課題と検討事項

デジタルビデオの一般的な問題は、「生」の高品質フォーマットのデータ自体が非常に大きいことです。例えば、Wrecks on the Seabed プロジェクトでは、約40時間のデジタルビデオが作成されました。例えば、潜水調査のハイライトを見せるなど、ビデオ映像を「よりきれいに」編集したバージョンが作成された場合、どのバージョンをどのような品質のフォーマットで保存すべきか、十分な情報に基づいた決定を下すことが重要です。この決定は、そのファイルの目的（唯一の記録なのか、それとも他のデータセットの補助に過ぎないのか）と照らし合わせて判断する必要があります。またデジタルビデオ・テープの寿命はますます短くなっており、理想的にはディスクベースのストレージに移行すべきです。DVDの寿命も限られており、ハードドライブ（内蔵または外付け）が最も安全な保存媒体となるでしょう。

[1] <http://intarch.ac.uk/news/eva97.html>

[2] [http://www.wessexarch.co.uk/projects/marine/alsf/wrecks\\_seabed/index.html](http://www.wessexarch.co.uk/projects/marine/alsf/wrecks_seabed/index.html)

## セクション2. デジタルビデオの作成

### 2.1 一般的な検討事項

デジタルビデオの作成と保存を網羅した詳細な「デジタル

ビデオ入門」が JISC Digital Media [1] によって作成されています。本ガイドは、この文書やその他の文書の要点を紹介することを目的としているので、詳細については、JISC Digital Media のガイドを参照してください。一般的に、本ガイドはデスクトップ・ビデオファイル（キャプチャデバイスまたはコンピュータのいずれかにファイルとして存在するビデオ）の概要を説明することを目的としており、アナログ素材のデジタル化や、デジタルテープまたはディスク/DVDベースのビデオのデスクトップ・フォーマットへの変換を対象としていません。本ガイドで扱う他のデータタイプと同様に、データはDVDやテープなどの物理的なメディアではなく、ネットワークベースのストレージに保存するのが最適ですと想定されています。

前のセクションで述べたように、データ、デジタルビデオのソースによって、ビデオファイルのフォーマット、品質、最終的な物理的サイズが決まります。アナログソースからデジタル化する場合、最終的なファイルの品質とサイズを決定するいくつかの検討事項に加えて、クリエイターにはさまざまなフォーマットが用意されています。アナログ素材のデジタル化については、JISC Digital Media のドキュメント「Selecting a video digitisation system」[2] および「Equipping a video digitisation system」[3] に詳細が記載されています。他のデータセットからビデオファイルを作成する場合（例：3Dモデルの「フライスルー」）、デジタル化の場合と同様の検討事項がありますが、最終的なファイル形式と品質は使用するソフトウェアアプリケーションによって制限されるでしょう。

### コンテナとコーデック

デジタル画像やデジタル音声ファイルと同様に、デジタルビデオファイルには、ビデオファイルの品質とサイズを決定するさまざまな重要なプロパティが含まれています（次項で説明）。これらのプロパティの多くは、デジタルビデオに使用されるさまざまなファイルフォーマットの機能に直接関連しています。しかし、他のファイルフォーマットとは異なり、広く使われているデジタルビデオ・フォーマットの多くは、実際には、ビデオとオーディオの別々のストリームを単にカプセル化する「ラッパー」または「コンテナ」フォーマットです。そのため、データ作成者は、使用されているコーデック、その機能、使用目的を正確に把握することが非常に重要です。また、ラッパーのフォーマット自体も、MPEGなどの特定の形式では、各ストリーム内で使用されるコーデックの種類が制限されるなど、機能面で異なる場合があります。

一般的に、データ作成者は以下の点に注意する必要があります。

- ビデオとオーディオの両方のストリームにおける圧縮。他のさまざまなフォーマットと同様に、非可逆圧縮ではデータが失われてしまいます。オリジナルのビデオファイルには、圧縮を行わない（可能であれば可逆圧縮）ことをお勧めします。
- フレームサイズ/1フレームあたりのピクセル数は、ビデオ画像の物理的な幅と長さを表しており、画像の解像度と同様に、ファイルに取り込まれる詳細のレベルを決定します。
- フレームレートとは、1秒間に撮影・表示されるフレー

ム数のことです。フレームレートが高いほど映像は滑らかになりますが、ファイルサイズも大きくなります。

- ・ビットレート（オーディオとビデオ）は、フレームサイズ、フレームあたりのビット数、フレームレートを組み合わせたもので、多くのデジタルビデオアプリケーションでプリセットオプションとして表示されています。多くのアプリケーションでは、一定の標準的なプリセット

に加えて、フレームの複雑さに応じてレートを調整できる可変ビットレートのオプションが提供されています。

以下の表は、デジタルビデオファイルに使用される一般的なコーデックおよびコンテナフォーマットの概要です。これらの機能のより詳細な比較は、Wikipedia のコーデック [4] およびコンテナフォーマット [5] の比較表に記載されています。

## 2.2 ファイルフォーマット

フォーマット	プロパティ/技術	説明	推奨事項
MPEG 1 (.mpg,.mpeg)	公開されているオープンスタンダード [6] のビデオ・オーディオ用バイナリフォーマット。	ビデオ CD (VCD および SVCD) や、一般的ではないが DVD-Video のために、Moving Picture Experts Group (MPEG) によって開発された国際 ISO/IEC 規格 (11172) です。VHS テープと同等の品質のオーディオ / ビデオ再生が可能。MPEG-1 Audio Layer III は、MP3 オーディオに相当する。このフォーマットを扱うために、オープンソースのものを含む多くのツールが存在する。	保存や公開に適している。
MPEG 2 (.mpg,.mpeg)	公開されているオープン規格。 [7]	MPEG-1 と同様に ISO/IEC (13818) 規格だが、DVD とデジタルテレビ向けの規格である。MPEG-1 と同様に ISO/IEC (13818) 規格だが、DVD やデジタルテレビ用の規格であり、いくつかの異なる「プロファイル」(シンプル・プロファイル、メイン・プロファイル、4:2:2 プロファイル(デジタルテレビ用))があり、それぞれに表示サイズやデータレートなどの標準的な仕様が異なっている。MPEG-2 ビデオは、低ビットレート (1Mbit/s 以下) には最適化されていないが、3Mbit/s 以上のレートでは MPEG-1 と比較して優れた品質を提供する。	特に保存や公開に適している。
MPEG 4 (.mp4)	公開されたオープンスタンダード。ISO/IEC 14496-14:2003	MPEG ISO/IEC (14496) 規格の中では最も新しく、ウェブ(ストリーミングメディア)、会話(テレビ電話)、テレビ放送など、AV ストリームを圧縮することでメリットを得ることができる規格である。Apple 社の QuickTime .mov フォーマットをベースにした MPEG4 は、オーディオとビデオを中心に、3D オブジェクト、テキスト、スプライトなどのメディアタイプもサポートしており、インタラクティブな要素を盛り込むことができる。MPEG2 と同様、MPEG4 にも 2 つの主要バージョンと、それぞれの目的に合わせて最適化された膨大な数の「プロファイル」がある。このフォーマットは着実に普及しており、NDIIPP のフォーマットページ [8] にその概要が記載されている。MPEG4 はデータの保存や普及に適しているが、必要に応じてより高品質の MPEG フォーマットを使用する必要がある。	MPEG4 は、データの保存や公開に適しているが、必要に応じてより高品質の MPEG フォーマットを使用する必要がある。

フォーマット	プロパティ/技術	説明	推奨事項
DivX (.divx, .avi)	商用(プロプライエタリ)の一般的なビデオコーデック	DivX 形式は、MPEG-4 の別の実装であり、インターネット上での映画の配信に広く使用されている。バージョン 4 以降の形式 (OpenDivX としても知られている) はライセンスに基づいて使用されているが、当初は Microsoft の MPEG4 実装である .wmv の違法な「ハッキング」バージョンをベースにしていた (以前の違法バージョンは、しばしば「DivX ;-) (スマイリー付き)」と呼ばれている)。DivX コーデックは無料でダウンロードできるが、エンコードにはライセンスが必要である。	公開には適しているが、保存には適していない。
Xvid (.avi, .xvid)	MPEG-4 をベースにした GNU GPL ライセンスのコーデックで、OpenDivX をさらに発展させたもの	Xvid は DivX と同様に MPEG-4 Part 2 Advanced Simple Profile (ASP) をベースにしている。	公開には適しているが、保存には適していない。
Motion JPEG 2000 (.mj2, .mjp2)	ISO/IEC 15444-3 で標準化された、JPG2000 エンコーディングを使用したコンテナフォーマット [9]	不可逆圧縮と可逆圧縮の両方のビデオを保存できる Motion JPEG 2000 フォーマットは、ビデオを一連の個別画像として保存するため、デジタル保存の分野、特にフィルムのデジタル化に関して関心を集めている。	普及に適しており、特に保存に適している。
Matroska (.mkv)	商用(プロプライエタリ)だがオープンソースのコンテナフォーマット。 [10]	柔軟性の高いコンテナフォーマットである Matroska は、「AVI、ASF、MOV、RM、MP4、MPG ES などの既存のコンテナに代わるオープンソースのコンテナ」を目指している [11]。その結果、このフォーマットは、事実上、制限のない多数のコーデックをサポートしている。	使用するコーデックによっては、公開や保存に適している。
Flash Video (.flv)	人気のある独自のコンテナフォーマット。(例: YouTube)	FLV フォーマットは、初め Macromedia、後に Adobe によって開発された、圧縮されたビデオをウェブで配信するための人気のあるフォーマットである。このフォーマットには、限られたコーデックのサポートが含まれている。	公開に適している。
Audio Video Interleave (.avi)	Microsoft 社が開発した商用(プロプライエタリ)のコンテナ。	AVI は、オリジナルの仕様が古いため可変ビットレートへの対応など、くつかの制限があるが、幅広いコーデックに対応しているコンテナフォーマットである。	公開に適している。
Quicktime (.mov)	Apple 社が開発した独自のコンテナフォーマット。	前述の通り、MPEG-4 規格のベースとなったのは Quicktime フォーマットである。Quicktime フォーマットは、場合によっては MPEG-4 と同様 (あるいはそれ以上) の機能を提供するが、可能な限り、確立された規格である MPEG-4 の使用が推奨される。	公開に適している。

## 今後の方向性について

MPEG や DivX フォーマットの開発が進んでいますが、SMPTE (Society of Motion Picture and Television Engineers) が開発したオープンスタンダードである Material Exchange Format (.mxf) フォーマットも、近い将来、動画やビデオの保存用コンテナフォーマットとして好ま

れるようになるかもしれません (Wright 2011, 13 参照)。

- [1] <http://www.jiscdigitalmedia.ac.uk/movingimages/advice/introduction-to-digital-video/>
- [2] <http://www.jiscdigitalmedia.ac.uk/movingimages/advice/selecting-a-video-digitisation-system/>

- [3] <http://www.jiscdigitalmedia.ac.uk/movingimages/advice/equipping-a-video-digitisation-system/>
- [4] [http://en.wikipedia.org/wiki/Comparison\\_of\\_video\\_codecs](http://en.wikipedia.org/wiki/Comparison_of_video_codecs)
- [5] [http://en.wikipedia.org/wiki/Comparison\\_of\\_container\\_formats](http://en.wikipedia.org/wiki/Comparison_of_container_formats)
- [6] <http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=25371>
- [7] <http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=37679&ICS1=35&ICS2=40&ICS3=>
- [8] <http://www.digitalpreservation.gov/formats/fdd/fdd000155.shtml>
- [9] <http://www.digitalpreservation.gov/formats/fdd/fdd000127.shtml>
- [10] <http://www.matroska.org/>
- [11] <http://www.matroska.org/technical/whatis/index.html>

### セクション3. デジタルビデオのアーカイブ

#### 3.1 アーカイブする内容の決定

何をアーカイブするかは、映像がどのように作られたかによって大きく異なります。冒頭で述べたように、多くのデジタル機器は、作成できるビデオの質や種類が限られているため、機器の性能をプロジェクトの要件と照らし合わせて評価することが不可欠です。理想的には、デジタルビデオをアーカイブに適した品質で撮影し、オリジナルを保存し、その「マスター」から低品質のファイルを作成することが望ましいでしょう。完全に非圧縮の映像を作成して保存することは、多くの人にとって可能な選択肢ではないかもしれませんが、一度ファイルをフォーマット間で変換(トランスコード)してしまうと、低品質のフォーマットからデータを復元したり改善することはできないため、適切なフォーマットを選択することが重要です。以下に示す重要なプロパティは、データをフォーマット間で移動する際に維持すべき点を示しています。

フォーマット	要件
MPEG 1 (.mpg,.mpeg)	公開されたオープンな標準形式である ISO/IEC 規格(11172)で、保存と公開に適している。
MPEG 2 (.mpg,.mpeg)	MPEG 2 (.mpg,.mpeg) MPEG-1 と同様に、公開されたオープンスタンダード(ISO/IEC (13818))で、特に保存に適している。
MPEG 4 (.mp4)	MPEG 4 (.mp4) 公開されたオープンスタンダードである ISO/IEC 14496-14:2003 は、データの保存と公開に適しているが、必要に応じてより高品質の MPEG フォーマットを使用する必要がある。

また、Pearson & Gill (2005) が詳しく述べている Motion JPEG 2000 も、特にデジタル化プロジェクトの場合には、デジタルビデオの有効な保存形式可能性があります。その適合性については、現在も研究が進められているところです。

#### 3.3 メタデータと文書化

他のデータフォーマットと同様に、デジタルビデオのメタデータは、そのデータがどのように作成されたかについての

#### 3.2 アーカイブする方法の決定

##### 重要なプロパティ

ビデオファイルの重要なプロパティについては、JISC レポート「The Significant Properties of Moving Images」(Coyne & Stapleton 2008) や AHDS preservation handbook for moving images (Knight & McHugh 2005, 7) で詳しく説明されています。要約すると、データを保存する際に変更してはいけないデジタルビデオファイルのプロパティは以下の通りです。

- ・ファイルの長さ(例: 5分31秒/150MB)
- ・フレームレート、1秒あたりのフレーム数(例: PAL は 25、NTSC は 30)
- ・フレームサイズ/ビデオ解像度(例: 720 × 576ピクセル)
- ・ビットレート(単位: kbps)
- ・オーディオビットレート(kbps)
- ・オーディオ周波数(kHz)
- ・使用されているオーディオチャンネル(例: ステレオ)
- ・関連するメタデータとドキュメント、ファイルサイズ

変換の前に、ビデオファイルを、提供されたドキュメントと照合し、ファイルが指定された長さ(分、秒)であることを、およびその重要なプロパティが正しく記録されていることを確認する必要があります。これにより、ファイルが完全なものであり、ファイル変換中に予期せぬ劣化が起こらないようにすることができます。

デジタルビデオのコーデック変換(あるデジタルフォーマットから別のデジタルフォーマットへのデータ移行)の実際のプロセスについては、JISC Digital Media のサイトで詳しく説明されています [1]。

##### ファイルフォーマット

デジタルビデオの長期保存には、下表のフォーマットを推奨します。

重要な情報を提供します。一般的なプロジェクトレベルのメタデータ(「プロジェクトのメタデータ」の項で説明)に加えて、ファイルの性質を理解できるように、デジタルビデオには特殊なメタデータを記録する必要があります。その場合、ここで紹介する特殊なメタデータの核となるものが、上述の重要なプロパティとひもづけられていることが重要です。

## 特別なメタデータ

要素	説明
ソフトウェア、バージョン、プラットフォーム	映像を作成するために使用したソフトウェア(デバイスから直接取得した場合はハードウェア)
ビデオコーデック	ビデオコーデックの名称とバージョン(必要に応じて)
ビデオのサイズ	ビデオのサイズ(ピクセル単位)
フレームレート	1秒あたりのフレームレート(fps)
ビットレート	ビデオのビットレート
オーディオコーデック	オーディオコーデックの名前とバージョン
音声サンプル周波数	-
オーディオのビットレート	-
オーディオチャンネル	使用チャンネル 例:ステレオ
長さ	ファイルの長さ(時、分、秒)
ファイルサイズ	ファイルのサイズ(単位:MB)

[1] <http://www.jiscdigitalmedia.ac.uk/movingimages/advice/transcoding-digital-video/>

### 参考文献

Coyne, M. & Stapleton, M. (2008) *The Significant Properties of Moving Images*. JISC. [http://www.jisc.ac.uk/media/documents/programmes/preservation/spmovimages\\_report.pdf](http://www.jisc.ac.uk/media/documents/programmes/preservation/spmovimages_report.pdf)

Knight, G. & McHugh, J. (2005) *Preservation Handbook: Moving Image*. AHDS Preservation Handbook. [http://](http://www.ahds.ac.uk/preservation/video-preservation-handbook.pdf)

[www.ahds.ac.uk/preservation/video-preservation-handbook.pdf](http://www.ahds.ac.uk/preservation/video-preservation-handbook.pdf)

Pearson, G. & Gill, M. (2005) 'An Evaluation of Motion JPEG 2000 for Video Archiving', *Proc. Archiving 2005* (April 26-29, Washington, D.C.), IS & T ([www.imaging.org](http://www.imaging.org)), pp. 237-243. [http://archive.nlm.nih.gov/pubs/pearson/MJ2\\_video\\_archiving.pdf](http://archive.nlm.nih.gov/pubs/pearson/MJ2_video_archiving.pdf)

Wright, R. (2011) *Audiovisual Digital Preservation Status Report 2*. PrestoPrime Deliverable D7.1.4. [https://prestoprime.ina.fr/public/deliverables/PP\\_WP7\\_D7.1.4\\_Annual\\_AV\\_Status\\_2\\_R0\\_v1.00.pdf](https://prestoprime.ina.fr/public/deliverables/PP_WP7_D7.1.4_Annual_AV_Status_2_R0_v1.00.pdf)

## デジタルオーディオ

### セクション1. デジタルオーディオの紹介

#### 1.1 デジタルオーディオとは何か

デジタルビデオと同様に、デジタルオーディオファイルは、この10年間ではるかに簡単に作成できるようになりました。デジタルビデオは、考古学でより多くの用途を見出されていますが、デジタルオーディオファイルは、オーラルヒストリーの記録や、考古学的な楽器を現代的に再構築したり、物理的または仮想的に再構築された教会、山小屋、劇場などの考古学的なコンテキストの中で音を録音するような、「考古学的な音」の再現を目的としたプロジェクトの一部として作成されることがあります。後者の例は、オンラインのInternet Archaeology ジャーナルに多数掲載されています(例: Thomas 2011)。

デジタルビデオガイドと同様に、このガイドは「ポーンデジタルな」オーディオファイルの作成と保存に関わる問題を扱うことを目的としており、アナログによるオリジナル音源からのファイル作成は対象としていません(ただし、ここで

議論されている課題の多くは同様に適用されます)。アナログオーディオファイルのデジタル化については、JISC Digital Media ガイド「Audio.Digitising analogue media」[1]で詳しく説明されています。このガイドは、JISC Digital Media のサイトにある豊富な資料に加えて、デジタルオーディオファイルの保存に関する他の重要なガイド、すなわち AHDS の「Preservation Handbook: Digital Audio」(Knight & McHH)を参考にしています。このガイドは、AHDS の「Preservation Handbook: Digital Audio」(Knight & McHugh 2005)、JISC の「Significant Properties Testing Report: また、IASA 技術委員会の「Guidelines on the Production and Preservation of Digital Audio Objects」(第2版)(Bradley 2009)もあります。また、デジタルオーディオファイルの技術的な側面については、ガイド「An Introduction to Digital Audio」[2]に詳しく紹介されています。

#### 1.2 現状の課題と検討事項

デジタルビデオデータと同様に、デジタルオーディオファイルは、非圧縮形式で作成/保存するとサイズが大きくなる可能性があり、低品質のファイルをいつ、どのように作成す

るかを決定する際に、十分な情報 JISC Digital Media を得た上で決定する必要があります。また、デジタルビデオと同様に、デジタルオーディオファイルにはさまざまなコンテナフォーマットやコーデックが混在しているため、オーディオファイルを識別して利用するためには、詳細な技術的メタデータが重要となります。またメタデータは、ファイルの作成プロセスやコンテンツ（インタビューの名前や日付、場所など）を文書化する上でも重要な役割を果たします。

- [1] <http://www.jiscdigitalmedia.ac.uk/audio/docs/category/digitising-analogue-media>  
 [2] <http://www.jiscdigitalmedia.ac.uk/audio/advice/an-introduction-to-digital-audio>

## セクション2. デジタルオーディオの作成

### 2.1 一般的な検討事項

デジタルオーディオファイルを作成する際のプロセスや注意点については、多くの JISC Digital Media 文書に詳しく説明されています。

- ・ Choosing a Digital Audio File Format] [1]
- ・ 'Basic Audio Editing' [2]
- ・ Choosing a Digital Audio File Format] [3]

セクション1で述べたように、デジタルオーディオファイルを作成する際にはファイルの品質が重要な検討事項となります。作成されるファイルが適切な品質であり、目的に適合していることを保証するために、データ作成者は以下の点に注意する必要があります。

- ・ オーディオコンテナまたは「ラッパー」フォーマットと、それに含まれるコーデックとの関係、およびそれぞれの機能。
- ・ コーデックの構成要素として、オーディオフォーマットには可逆圧縮と非可逆圧縮の両方の機能があること、また非圧縮のフォーマットもあること。データ作成者は、これらのオプションがデータ品質とファイルサイズの両方にどのような影響を与えるのか、ワークフローのどの段階でこれらのオプションが使用されるのか（もし使用されるのであれば）を認識する必要があります。
- ・ 多くのファイルはコンテナ形式でメタデータを埋め込むことができ、データ作成者はデータ作成プロセスの重要な側面を記録するためにこの機能を利用したいと思うかもしれません。このようなメタデータが存在する場合には、それを特定したり、フラグを立てたりして、フォーマットに関係なく維持できるような措置をとるべきです。
- ・ どのような権利が存在するか（例：著作権や著作者人格権）。これは特にインタビューの状況に関連しており、データ作成者は自分の作業が倫理的にどのような意味を持つかを認識する必要があります。どのような場合でも、記録に関わった人に許可を求めるべきです。これらの問題は、JISC Digital Media 文書「Copyright and Other Rights for Creating Time-based Media Resources」[4] で説明されています。

一般的な指針として、本ガイドでも繰り返し述べているとおり、データ作成者は、利用可能な限り高品質のオーディオデータを非圧縮形式で作成（およびアーカイブ）するべきです。このマスターデータセットから、フォーマットの移行（マイグレーション）やダウンサンプリングによって、ファイルサイズや品質の低いデータセットを得ることができます。

### 2.2 ファイルフォーマット

フォーマット	プロパティ/技術	説明	推奨事項
Waveform Audio (.wav)	Microsoft と IBM によって開発された、広く使用され、文書化されている、ただし商用（プロプライエタリ）のコンテナフォーマット。	.wav フォーマットは、一般的に非圧縮オーディオ (PCM) を保存するために使用されるが、MP3 のようなさまざまなロッキーコーデックのオーディオを格納することもできる。また、このフォーマットにはメタデータのタグ付けが可能で、XMP フォーマットでメタデータを埋め込むことができる。	使用するコーデックによっては保存に適している。
Audio Interchange File Format (.aif, .aiff)	Apple 社が開発した独自のフォーマットで、.wav フォーマットに似ている。	.wav と同様に、主に非圧縮の PCM オーディオファイルの保存に使用される。また、他の圧縮コーデック形式の音声と一緒にメタデータを保存することも可能である。	非圧縮での保存に適している。
Sun AU (.au)	Sun が Unix システム用に開発したフォーマット。	.wav や .aif と同様に、Sun AU のファイルは大きくて高品質な傾向があるが、UNIX コミュニティ以外ではあまりサポートされていない。主に PCM エンコーディングを使用しているが、.au ファイルは他の多くのコーデックをサポートしている。	UNIX コミュニティ以外では広くサポートされていないため、保存には適していない。

フォーマット	プロパティ/技術	説明	推奨事項
Advanced Audio Coding (.aac)	MPEG-2 および MPEG-4 フォーマットをベースにした ISO 標準フォーマット。	このフォーマットは、多くの一般的なデバイス (Wii、プレイステーション、iPod、iPhone など) で広くサポートされており、MP3 フォーマットの後継として設計されている。AAC フォーマットのデータは、MP4、3GP、ADIF、ADTS などのコンテナフォーマットでパッケージ化されている。	公開には適しているが、保存には適していない。
Broadcast Wave Format (BWF) (.bwf .wav)	欧州放送連合の仕様で、WAV フォーマットを拡張したもの [5]。	Broadcast Wave Format は、WAV フォーマットを拡張し、メタデータ用の「チャンク」を追加した非圧縮オーディオで構成されている。同じフォーマットに基づいているが、WAV ファイルと BWF ファイルの互換性には問題があり、特に WAV を BWF に移行する際に問題となることがある。	保存に適している。
Ogg Vorbis (.ogg)	Xiph.org [6] によって開発された、非独占的 (ノンプロプライエタリ) なオープンフォーマット。	Vorbis コーデックは、MP3 フォーマットに代わる完全にオープンなフォーマットとして開発され、主に配信フォーマットとして使用されている (Ogg ラッパーフォーマットには Vorbis 以外のコーデック (FLAC など) を含めることができる)。	公開用フォーマットとして適している。
MP3 (.mp3)	一般的な ISO 標準フォーマットで、MPEG-1 および 2 規格の一部である。	MP3 (.mp3) 一般的な ISO 標準フォーマットで、MPEG-1 および 2 規格の一部である。MP3 (または MPEG-1 Audio Layer 3) はロッシーフォーマットで、ファイルにメタデータ (ID3) を埋め込むことができる。	公開には適しているが、保存には適していない。
MPEG-4 (.m4a, .m4p)	ISO 標準のコンテナフォーマット。	MPEG-4 フォーマットは、多数の異なるオーディオコーデック (例: AAC, MP3) と埋め込まれたメタデータ (XMP を含む) をサポートしている。アップル社の iTunes ストアでも使用されているが、iTunes ストアで販売されているファイルは暗号化されたオーディオストリームを使用しており、.m4p という接尾辞で識別できる。	公開に適している。
RealAudio (.ra, .ram)	RealNetworks 社が開発した独自のオーディオコーデック。	RealAudio ファイルは、RealAudio Lossless Format (ralf) を含むいくつかのコーデックを使用できる。ストリーミングメディア用に開発されたため、一部のファイル (.ram など) は、それ自身がファイルではなく、オンラインファイルへのリンクに過ぎないことに注意が必要である。	保存には適していない。
Windows Media Audio (.wma, .asf)	Microsoft 社が開発した商用 (プロプライエタリ) のフォーマットで、いくつかのコーデックとコンテナフォーマットから構成されている。	WMA のコーデックセットには、標準フォーマットの可逆圧縮版、音声専用版、「プロフェッショナル」版がある。また、WMA のデータは独自の ASF コンテナでラップされていることが多い。	保存には適していない。



フォーマット	プロパティ/技術	説明	推奨事項
FLAC (.flac)	FLAC (Free Lossless Audio Codec)。Xiph.Org Foundation [6] によって開発されたオープンでフリーの可逆圧縮コーデックである。	FLAC はオープンソースのロスレスでより効率的な MP3 フォーマットの代替として開発され、人気が高まっている。FLAC でエンコードされたオーディオは、'Native Flac' や Ogg を含むさまざまな異なるコンテナフォーマットに埋め込むことができる。	公開に適しており、圧縮が必要な保存フォーマットとして使用することができる。
Speex (.ogg)	Xiph.Org Foundation によって開発されたオープンでフリーなコーデック [6]。	他の Xiph.Org Foundation フォーマット (例: Vorbis, FLAC) と同様に、Speex はフリーでオープンなコーデックである。Speex は高品質の圧縮された音声 / 音声ファイルを提供することを特に目標としている。	公開のために適しているが、保存には適していない。

- [1] <http://www.jiscdigitalmedia.ac.uk/audio/advice/choosing-a-digital-audio-file-format>
- [2] <http://www.jiscdigitalmedia.ac.uk/audio/advice/basic-audio-editing>
- [3] <http://www.jiscdigitalmedia.ac.uk/audio/advice/choosing-a-digital-audio-file-format>
- [4] <http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/copyright-and-other-rights-for-creating-time-based-media-resources>
- [5] [http://www.ebu.ch/fr/technical/publications/userguides/bwf\\_user\\_guide.php](http://www.ebu.ch/fr/technical/publications/userguides/bwf_user_guide.php)
- [6] <http://www.xiph.org>

### セクション 3. デジタルオーディオの保存

#### 3.1 アーカイブする内容の決定

本ガイドや Knight (2005) が述べているように、「作成の瞬間がデジタルオーディオ保存の最も重要な段階」であり、可能であれば、オリジナルファイルは、その重要なプロパティ (後述) を同等以上の価値で維持できる形式で保存されるべきです。また、ファイルの物理的プロパティに加えて、関連するメタデータや文書 (埋め込まれているか、別個に保存されているか) も保存・維持する必要があります。

#### 3.2 アーカイブする方法の決定

##### 重要なプロパティ

保存活動を通じて維持されるべきオーディオファイルの重要なプロパティについては、'Significant Properties Testing Report: Audio Recordings' (Knight 2010, 8-10) で詳しく説

明しています。以下にその概要を示します。

- ・継続時間：タイムコード文字フォーマット (TCF) でのオーディオファイルの長さ。ファイルが意図した長さと一緒にしているかどうかを確認する必要があります。
- ・ビット深度：1 サンプルあたりに格納されている情報のビット数を示し、オーディオ品質の指標となります (例：16 ビットまたは 24 ビット)。
- ・サンプルレート：1 秒あたりのサンプル数を表す。サンプルレートは通常、44.1kHz (一般的なサンプルレート) のようにヘルツで表され、ビット深度と同様にファイルの品質の指標となります。
- ・チャンネル：オーディオ・オブジェクト内の個別のストリームの数を示す記述的または数値的な値、またはその構成を示す記述的または数値的な値 (例：2 (ステレオ))
- ・チャンネルの割り当て：チャンネル番号とサウンドマップの位置。特定のチャンネルを特定の出力位置にマッピングし、出力音を設定するための 2 つの値があります。

##### ファイルフォーマット

以下に示すファイルフォーマットは、オーディオファイルの保存に推奨されるものです。JISC Digital Media 文書「Uncompressed Audio File Formats」 [1] で強調されているように、非圧縮オーディオフォーマットは「音の波を最も正確にデジタルで表現するが、デジタルオーディオを記録・保存する際に最もリソースを必要とする方法でもある」。このことから、オーディオファイルの保存には非圧縮フォーマット/コーデックが強く推奨されますが、それが不可能な場合は、高品質の可逆圧縮を使用することができます (例：FLAC)。

保存形式	要件
Waveform Audio (.wav)	非圧縮音声 (PCM) での保存を推奨する。また、このフォーマットはメタデータをタグ付けすることができ、XMP フォーマットでメタデータを埋め込むことができる。
Broadcast Wave Format (BWF) (.bwf .wav)	保存のために推奨されている。BWF は WAV フォーマットを拡張し、メタデータ用の「チャック」を追加しているが、WAV ファイルと BWF ファイルの互換性には問題があり、特に WAV から BWF への移行の際には注意が必要である。
Audio Interchange File Format (.aif, .aiff)	非圧縮の PCM オーディオファイルを保存するのに適したフォーマット。メタデータを埋め込むことも可能。

保存形式	要件
FLAC (.flac)	オープンでフリーな可逆圧縮コーデックで、非圧縮の音声が目ましくない場合の保存に適している。

### 3.3 メタデータと文書化

オーディオファイルにはさまざまなメタデータの仕組みがありますが、メタデータをファイル自体に埋め込むことができるフォーマットが多いことが特徴です。以下に示すメタ

データの要素は、主に技術的な側面をカバーしており、記録すべき最低限の内容を示しています。また、一般的なプロジェクトメタデータも記録されるべきでしょう。

要素	説明
ソフトウェア	ファイルの作成に使用されたソフトウェア(またはデバイス)
ビット深度	例:16ビットまたは24ビット
ビットレート	任意、kbpsで記録されることが多い
サンプルレート (KHz)	例:44.1kHz
使用コーデック	ファイルの作成に使用されたコーデック 例:FLACまたはAAC
録音の長さ	時間、分、秒(hh:mm:ss)で記録することが望ましい
著作権クリアランス	オーディオファイル、特にオーラルヒストリーやインタビューの場合、これらは非常に重要である。
インタビューの文字起こし	インタビューの文字起こしは、特に録音内容を明確にし、個人を識別するための重要な資料となる。

その他のメタデータスキームについては、IASA [2] が概要を示し、JISC Digital Media [3] が詳細を説明しています。

### 3.4 著作権について

第2章で述べたように、データ作成者は、人間を記録する際に生じるさまざまな権利を認識し、関連する権利や許可を確実に取得する必要があります。これらの問題は、JISC Digital Media文書「Copyright and Other Rights for Creating Time-based Media Resources」[4] で詳しく説明されています。

### 参考文献

- Bradley, K. (ed.) (2009) *Guidelines on the Production and Preservation of Digital Audio Objects*. Second edition. IASA Technical Committee <http://www.iasa-web.org/tc04/audio-preservation>
- Thomas, D. (2011) 'An Investigation of Aural Space inside Mousa Broch by Observation and Analysis of Sound and Light'. *Internet Archaeology* 30. [http://intarch.ac.uk/journal/issue30/thomas\\_index.html](http://intarch.ac.uk/journal/issue30/thomas_index.html)
- Knight, G. & McHugh, J. (2005) *Preservation Handbook: Digital Audio*. AHDS. <http://www.ahds.ac.uk/preservation/audio-preservation-handbook.pdf>
- Knight, G. (2010) *Significant Properties Testing Report: Audio Recordings*. JISC. <http://www.significantproperties.org.uk/testingreports.html>

- [1] <http://www.jiscdigitalmedia.ac.uk/audio/advice/uncompressed-audio-file-formats>
- [2] <http://www.iasa-web.org/tc04/basic-metadata>
- [3] <http://www.jiscdigitalmedia.ac.uk/audio/advice/metadata-and-audio-resources>
- [4] <http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/copyright-and-other-rights-for-creating-time-based-media-resources>