

プロジェクトのライフサイクル

デジタルデータ作成計画

プロジェクトが始まった瞬間から、プロジェクトのライフサイクルを通じて作成されるデジタル資料についての計画を慎重に考えなければなりません。最初の段階に基本的な計画を立てることで、デジタルアーカイブへのプロジェクト資料の登録とアーカイブの作成を大幅に簡潔にしましょう。最初の段階で計画を立てることは、プロジェクトのライフサイクルを通してデータのセキュリティと復旧を確実にすることにも役立ちます。新たにアーカイブを作成するにしても、既存のリポジトリに追加するにしても、前もって準備しておくことが大事です。考慮すべき具体的な問題としては、ファイルの命名、形式、バージョン、そしてストレージ、バックアップの戦略、文書化などがあります。この章では、これらの問題についてより詳細に論じます。

計画を立てる前に、素材のアーカイブの方法を確実に決定する：

- ・新しいデジタルリポジトリを作成する場合は、デジタルアーカイブに関する本ガイドの項目を確認し、提示されている課題にどのようにアプローチするかを検討します。アーカイブの作成者は、アーカイブへのアクセスと継続的な保存にも対応する必要があります。
- ・既存のリポジトリに資料を追加する場合、プロジェクトの計画段階でそのリポジトリに連絡し、この章で取り上げる課題について確認します。これによって、データに適したリポジトリを選択し、データの登録過程を簡素にすることができます。

計画の際に考慮すべき問題：

プロジェクトの設計は、しばしばプロジェクトと作成するデータ形式ごとに固有のものとなります（後に個別の手法の項目で説明します）、いくつかの一般的なプロジェクト計画の原則は、プロジェクトの種類に関係なく適用されます。

- ・データ作成のための計画（本章では一般的なものを取り上げています）
 - ・プロジェクト設計の準備：作成または取得するデジタルデータの種類を定義し、プロジェクトの全ての段階でファイルを作成・管理する範囲を定め、文書化する
 - ・ファイル形式の選択：データ作成および分析の段階で利用し、データの確実なアーカイブや普及に利用する形式にどのように関連づけるか。使用される形式はプロジェクトのライフサイクルを通じて同じかもしれないし、変更されることもあるかもしれません。
 - ・必要な文書化のレベルとメタデータの評価：プロジェクトとデータセットが理解可能かつ再利用可能であることを保証するために、ダブリンコアなレベル（プロジェクト、技術、ファイル）で行います。プロジェクトの文書化とメタデータに関する項目で扱います。
 - ・アーカイブ方法の選択
 - ・デジタルアーカイブの種類の評価：どのようなアーカイブを作成し、どのファイルを保存するかを決定します。データの選択と保存に関する項目で扱います。

- ・デジタルアーカイブ施設への連絡：どのガイドラインや基準に従うべきかを、ファイルを受け取ってもらう先に確認する。ガイドラインが（アーカイブ施設により）指定されない場合は、本ガイドラインに従うことを推奨します。

最も重要なことは、データ作成者は、プロジェクトを成功させるために必要なタスクを、開始時に計画に含める必要があるということです。プロジェクトの期間中、さまざまな活動の計画は、必要に応じて見直され、修正されるべきです。本章では、計画とデータ作成の段階において、一般的なプロジェクトレベルで考慮すべき事項を概説しています。

データ作成と取得

プロジェクトの初期段階では、はじめからデジタルで作成されたデータ（ボーンデジタル）と、非デジタルの原資料から取得されたデータ（デジタル化されたもの）を区別する必要があります。本ガイドのデータ形式または手法に関する項目では、データ作成段階で使用されるファイル形式とソフトウェア・アプリケーションについて、長期保存への適合性と他の形式への移行（マイグレーション）の可能性の点から論じています。安定して信頼性の高い形式を識別する指針として、アーカイブやリポジトリで一般的に引用されているいくつかの一般原則があります（Todd 2009参照）。

- ・オープンな形式と商用（プロプライエタリ）形式、標準化：

アーカイブとリポジトリでは、標準化されオープンな、可能であれば独自仕様の商用（プロプライエタリ）でない形式でのデータ保存が一般的に選択されます。そうした（選択される）形式は、ファイル形式のオープンさと利用しやすさ、そしてプラットフォームをまたがって、または複数のアプリケーションでの利用が可能であることから、持続可能性に優れていると考えられています。これについては、公開、外部依存性、特許の影響、適用の観点から、さまざまな機会に議論されています（Library of Congress [1] and Todd (2009)）。

- ・バイナリファイルとプレーンテキスト・ファイル：

ラスタ画像のような多くのデータセットでは、バイナリ・エンコーディングに基づくファイル形式が唯一の選択肢となります。しかし、スプレッドシートやデータベース、テキスト文書などの幅広いデータセットでは、一般的にテキスト・エンコーディングにもとづくファイル形式（ASCII プレーンテキストやXML）を使用するという選択肢があります（そしてアーカイブとして好まれてきました）。プレーンテキストの利点は、ファイル形式の透明性がより高く（人間が読みとり可能で直接的な分析に開かれている）、その結果、内容と関連するソフトウェアを特定しやすくなるということです。また、ファイルの内容に簡単にアクセスできることは、そのような形式は外部依存性が少なく、使用していたソフトウェアが利用できなくなった場

合でもデータを移行できる経路があるということを意味します。

・圧縮ファイル：

ファイルの圧縮は、圧縮済みアーカイブファイル（例えばZIPやRAR形式）の作成や、JPGやPNGのような特定のファイル形式において行われます。どちらの場合でも、可逆的（データの損失がない）、不可逆的な手法のいずれも使用することができます。単一の「アーカイブ」ファイルの場合、データ圧縮は、バイナリ・エンコーディングと同様にファイルの種類の識別や内容へのアクセスの潜在的な障壁となります。パスワード保護などのセキュリティ機能が組み合わさった場合、最終的にデータに永久にアクセスできなくなる可能性があります。JPG画像やMPEG動画などのファイル形式で使用される圧縮は、ファイルが継続的に再処理や再圧縮されると、さらなる問題や（「世代損失」として知られるような）データ品質の劣化を引き起こす可能性があります。圧縮が使用されているすべてのファイルに関して、不可逆技術を使用している場合にはデータの損失をもたらし、さらにはビット破損によりデータの損失や破損の影響を拡大させる可能性があります（Heydegger 2008）。

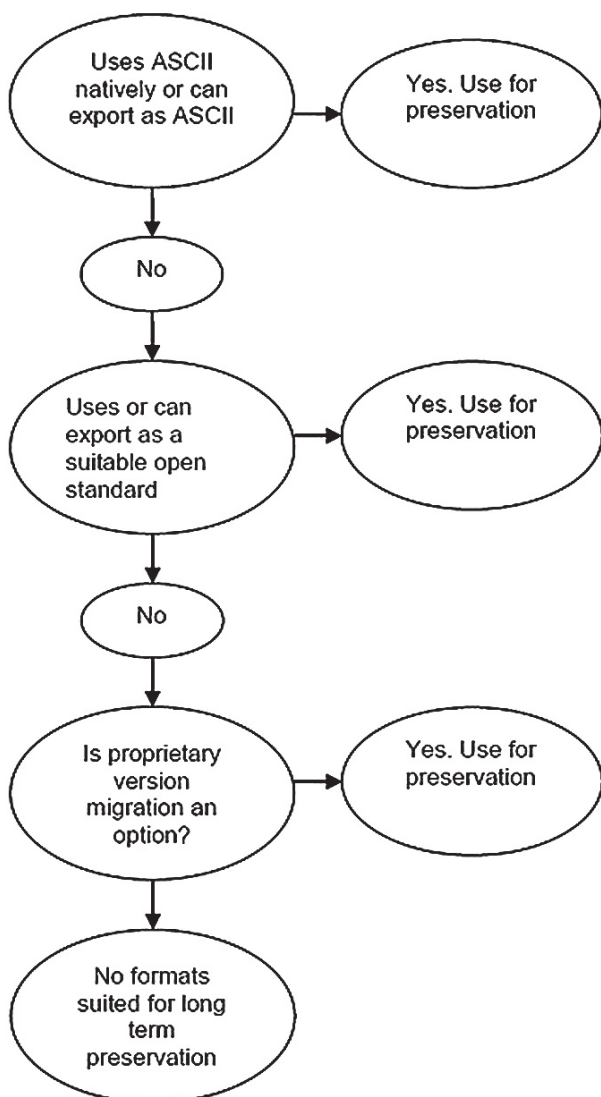


図1：保存に適したデータフォーマット選択の主な段階

ここでは、特定の形式は長期保存には適していないかもしれませんが、データの作成、開発、普及には最も適しているかもしれないということを確認しておきたいと思います（Brown 2008:5）。ここではプロジェクトのデータ作成段階でこうした検討課題を認識し、後で問題が生じる可能性のあるファイルを変換する必要がある場合には、予め適切な計画を含めておくことを推奨します。

適切なファイル形式を特定するための鍵となるもう一つの要素は、それを利用するコミュニティ内で特定のファイル形式がどの程度採用されているかということです。このことは、手法、データの種類、国によって異なることがあるので、本ガイドの後の章で詳しく説明します。

保存に適したデジタル形式を特定するための基準を詳細に記述した利用可能な資料が多数存在します。例としては、議会図書館 [2] やイギリス国立公文書館（Brown 2008）、デジタル保存連合（DPC） [3] が発行したものなどが挙げられます。

デジタル化されたデータ

本章は、主にポーンデジタルなデータを扱っており、デジタル化に関するアドバイスの提供を目的とはしていません。しかし、そのようなプロジェクトでは、内部でデジタル化を行う場合でも、作業を外部に委託する場合でも、結果として得られるデータセットに与えるファイル形式の影響を考慮する必要があります。プロジェクトは、アーカイブ品質のデータファイルを作成するために、ファイル形式にかかわらず最高品質のデータキャプチャを使用してオリジナル資料をデジタル化する計画を立てる必要があります。いくつかのファイルは、保存や共有のためにデータ品質の低下を招くことが多い手法を用いて圧縮されることがあるので、用途によって複数のバージョンのファイルを作成して保存することが望ましいかもしれません。デジタルデータを作成する際に、オリジナルの資料も有用な場合があるので、それを物理的に保存すべきかどうかを判断するために、文書アーカイブを検討する必要もあります。

デジタル化について、実践的なガイダンスを提供する多くの組織やガイドラインが存在しています。JISC Digital Media [4] は、既存の画像やアナログの音声・映像のデジタル化に関する幅広いアドバイスを提供しています。AHDS [5] や UKOLN [6] などの他のガイドラインでは、デジタル化プロジェクトの計画に関する短いガイダンスを提供しています。

ライセンス取得データと外部データ

ライセンス取得データと外部データは、著作権、ライセンス、またはソフトウェアが一部のアーカイブ作業を禁止している場合があるため、デジタルアーカイブに独特の課題を生じさせています。ファイルを元の形式からアーカイブにコピー、再生産、変換することが可能であるか、それらのファイルをアーカイブに登録することが可能かどうかなどが課題です。後者は特に困難な場合があり、例えば、基礎となる地図レイヤーがライセンス化されているGISファイルの場合、そのファイルを削除すると、後からそのファイルを解釈したり、読み込んだり、使用できる可能性が著しく低下します。

ファイル命名規則

ソースに関わらず、データを確実に把握する最初の（そし

て最も即時的な)方法の一つは、内容を反映した意味のあるファイル名を使用することです。データ作成者は、プロジェクトの開始から標準的なファイル命名規則とディレクトリ構造を使用するように計画し、可能であれば、すべてのプロジェクトで一貫した規則を使用すべきです。ディレクトリ構造とファイル名は後に特定のデータの種類に関連して論じますが、考慮すべき一般的な原則と規則は以下の通りです。

- ・PDF、DOCX、TIFなど、アプリケーション特有のコードには3～4文字のファイル拡張子を入れ、ファイル名の他の箇所にピリオド(.)を使用しない。
- ・一部のOSで問題が生じる可能性があるため、できる限りファイル名の中にスペースを使用しない。データ作成者はファイル名の中でスペースを表したい場合は、アンダーバー、つまり“_”を使うことを推奨する。
- ・ファイル名には、特有の参照番号やプロジェクト番号、プロジェクト名など、関連するアクティビティを識別する方法を含める。
- ・必要に応じて、ファイル名にバージョン番号の情報を含める。

DOSでは3文字のファイル拡張子と標準で8文字のファイル名を使用できますが、Windows環境では長いファイル名を使用できます。このように特定のOSで作成されたファイルには、特定の命名要件が生じる場合があります。

バージョン管理

ファイルを扱う際には、厳密なバージョン管理を維持していくことがとても重要です。一つのファイルを複数の人が作業する場合や、ファイルが複数の処理段階を経る場合には特に重要となります。例えば、ニューアム博物館考古学サービスのアーカイブ(詳細は「デジタルアーカイブとは何か?」を参照)にはどれが最新のものであるとは一切表示されずに、複数のバージョンの同一ファイルが含まれていました。アーカイブには、デジタルファイルに付随する文書がなかったため、ADSは日付とファイルサイズに基づいてファイルの状況の判断を下さざるを得なかったのです。

バージョン管理を提供するには、3つの一般的な戦略があります。

- ・ファイルの命名規則
- ・作成日とバージョン番号が記載された標準ヘッダー
- ・ファイルログ

実際には、どんな小さな変更であってもすべての変更を記録することが重要になります。不要になったバージョンは、適切なバックアップファイルが作成されていることを確認してから削除してください。

ファイル構造

デジタルファイルは理解しやすいディレクトリ構造に整理されるべきです。ディレクトリ構造からの検索を補助するために、標準的な命名規則を用いたフォルダーに関連するデータファイルを保存することが望ましいでしょう。特定の技術(例えば、物理探査やGIS)の結果として作成された一部のファイルは、標準化されたファイル構造により構造化され、保存されるべきです。このことは、本ガイドの後の関連する項目で論じます。

デジタルデータの保存—データ作成段階(およびそれ以降)の注意点

ほとんどのプロジェクトでは、作業期間中、デジタルデータは独立した(スタンドアロンの)PCのハードディスクや、ラップトップコンピュータ、ネットワークドライブ上に作成されます。さらにデータは、USBドライブやバックアップテープ、CD、DVD-ROM、またはその他の電子媒体に保存されることがあります。理想的には、それらがどのように作成され取得されたかどうかに関わらず、現在使用されているデジタルファイルは、良好な作業実践の一環として定期的にバックアップすべきです。

デジタル媒体を棚やデータ保管庫に放置しておくだけでは十分ではありません。デジタル媒体を安全に保管するためには、耐火性、耐磁性のある施設が非常に重要であり、バックアップのバージョンは、元の媒体から離して保管すべきです。データ作成者は、保管する前に、アーカイブが完全であること、文書が含まれていることを確認する必要があります。また、ファイルの保管場所や媒体のラベリング規則を記録した効果的なデータ管理システムを採用することも重要です。データ作成者が、プロジェクトの作成段階でデータの確実性を確保するために利用できる簡単な手順や戦略が多数あります。

安全なバックアップ

バックアップとは、一次的な場所とは別のところにデータの非常用コピーや一時保存(スナップショット)を確保しておくという作業です。小規模なプロジェクトの場合は、外付けディスクドライブやネットワーク上に保存された一つのファイルを意味する場合がありますが、大規模なプロジェクトやデータセットの場合は、耐火性のある戸棚やオフサイト(作業場所とは異なる場所に保存される)コピー、そして毎日、毎週、毎月の定期的なコピー作成など、厳密な災害対策の手順を意味する場合があります。これらはプロジェクトの期間中は重要ですが、プロジェクトが完了し、デジタルアーカイブが安全に保管されるとバックアップ作業は不要になるので、長期的なアーカイブとは異なります。

最も広く利用されているバックアップ戦略、いわゆる「三世代(祖父母-親-子)」戦略は、大きな機関ではデジタルテープを利用して実施されることが多いですが、他の記録媒体も利用可能です。このシステムでは、曜日や日ごとに、完全バックアップと部分バックアップのローテーション実施を採用しています。最新の完全バックアップである「親(Parent)」には、週の初め時点のネットワークやデータセット全体のスナップショットが入っています。「子(Children)」はより頻繁に、通常では毎日、その日に実行されたシステムへの変更のみを取り入れたバックアップです。これらのテープは、永久に保存する必要はありませんが、新しい親が作成される度に再生利用することができます。月に一度ぐらい、恒久的で完全な一時保存が行われ、これは永久に保存するもので通常は再生利用されません。この毎月のバックアップが「祖父母(Grandparent)」であり、もしもの時の復旧に役立てることがあります。週ごと月ごとのバックアップは、オフサイトで、なるべく耐火性、耐磁性のある安全な環境に保存するのが最善の方法です。もちろん、小規模、または比較的静的なデータセットには、定期的なコピーは過剰です。このバックアッ

プ・システムは、一つ一つの要求に合わせて調整でき、必要に応じて期間を拡大・縮小することができます。

また、すべてのファイルの形式と重要なデータが正確に保存されていることを確認するために、バックアップコピーを検証することも重要です。プロジェクトが完了、または活動休止した際、またはファイルサイズが大きくネットワーク上での処理が困難な場合には、バックアップを作成してください。各バックアップには明示的なラベルを付け、その所在を記録しておく必要があります。

ウイルスや他の問題の定期的なチェック

常時使われている（アクティブ）ものか、他所に保管されているものかに関わらず、デジタルデータセットから無作為で選んだサンプルを定期的にチェックする必要があります。適切なチェックには、ウイルスの検知やほとんどのコンピュータの OS に入っている定期的なスクリーニング手順が含まれています。このような定期的なチェックは、すべてのファイルに対して常に行われる厳格なウイルス検知に加えて実施されるべきです。

安全なバックアップを作成しても、保存されている媒体の劣化や、ハードウェアやソフトウェアの変化によりデータが復元できないものになってしまうこともあり、長期的にデジタルデータを十分に保護することはできません。デジタルアー

カイブを成功させるには、さらに追加のステップが必要です。

- [1] <http://www.digitalpreservation.gov/formats/sustain/sustain.shtml>
- [2] <http://www.digitalpreservation.gov/formats/sustain/sustain.shtml>
- [3] [http://www.dpconline.org/advice/preservationhandbook/media-and-formats\(*\)](http://www.dpconline.org/advice/preservationhandbook/media-and-formats(*))
- [4] [http://www.jiscdigitalmedia.ac.uk/\(*\)](http://www.jiscdigitalmedia.ac.uk/(*))
- [5] <http://www.ahds.ac.uk/creating/information-papers/checklist/index.htm>
- [6] <http://www.ukoln.ac.uk/interop-focus/gpg/DigitisationProcess/>

参考文献

- Brown, A. (2008) *Selecting File Formats for Long-Term Preservation*. The National Archives.
<http://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf>
- Heydegger, V. (2008) 'Analysing the Impact of File Formats on Data Integrity' in *Archiving* 2008, Volume 5.
[http://www.imaging.org/IST/store/epub.cfm?abstrid=38884\(*\)](http://www.imaging.org/IST/store/epub.cfm?abstrid=38884(*))
- Todd, M. (2009) *File Formats for Preservation*. DPC Technology Watch Series Report 09-02.
<http://www.dpconline.org/publications/technology-watch-reports>

プロジェクトの文書化

考古学者が「文書化」を考えると、一般的にはプロジェクトの経過や調査の背景、結果に関する情報を提示する報告書を思い浮かべるでしょう。これらは完全なデジタルアーカイブの重要な構成要素ですが、デジタルデータの再利用を真の意味で可能にするためには、より専門的な形式の技術的な文書化（またはメタデータ）も必要となります。このように構造化されたメタデータ（すなわちデータに関するデータ）は、プロジェクト名や場所、期間などのプロジェクト全体に広く適用できる情報から、アーカイブ内のファイル形式やそれを作成する際に使用されたアプリケーションなどの特定のデータセットやファイルレベルのメタデータまで、多岐にわたるものです。文書化とメタデータの簡潔な要約は英国データアーカイブのウェブサイト [1] に掲載されています。

文書化：

「データがどのように作成され、デジタル化されたのか、データは何を表しているのか、その内容と構造は何か、そして行われた可能性のある操作は何かということを説明するもの」

メタデータ：

「コアデータの文書化の一部で、データ収集に関する目的、出处、時間基準、地理的位置、作成者、利用状況を説明する標準化・構造化された情報を提供するもの」

文書化とメタデータ作成は、プロジェクトのライフサイク

ル全体を通して継続的に行うべきであり、単にプロジェクトの最後や、アーカイブへの登録の時点だけで行うことではないことを強調しておきます。継続的に行われた文書化とメタデータは、データ作成者にとって有用なものであり、プロジェクトが終了した後にデータを保存したりアクセスしたりする利用者（すなわち、アーカイブやコミュニティ）だけを対象としたものではありません。

本章では、多くの考古学プロジェクトの中核をなすべき文書化の種類とレベルについて考察します。以下、考古学データセットの発見、保存、再利用を支援するために追加で必要とされる特定のスキーマや標準などのメタデータの種類に焦点を当てていきます。

文書化

文書化は、アーカイブの基本の一つであり、関連データの管理を容易にするために、プロジェクト内の適切な場所で行われなければなりません。文書化は、遑って行うことが困難であることが多く、プロジェクトの開始時から積極的に実施すべきです。情報がファイル自体（プロジェクト報告書など）に潜在的に含まれていることがあるため、文書化の妥当性がしばしば疑問視されますが、しかしそのような潜在的な情報は資料の探索とデータ管理に役立つものではなく、このことはデータの再利用を成功させる鍵となります。

プロジェクトの背景に関するドキュメント

プロジェクトの背景に関する情報は、すべての考古学プロジェクトにおいて優れた専門的実践の一環として期待されるべきです。デジタルデータの再利用の可能性があるとき、この情報を明確に報告することが特に重要となります。資金やその他の機密情報を省略した上で、プロジェクト計画のデジタルコピーが、背景情報を供給するために提供されることがあります。

方法論

多くのプロジェクトのアーカイブでは、作成されるデータの種類として暗黙的に示されているとしても、使用する手法と記録システムを完全に文書化しておく必要があります。検討すべき項目は、本ガイドの中の別の項目で扱います。発掘調査プロジェクトでは、現場記録のためのガイドやマニュアルのデジタルコピー、一般公開されアクセス可能な出版物の書誌情報、特定のプロジェクト内での記録システムの使用方法などの詳細をアーカイブに添付する必要があります。層序単位や階層単位がどのように定義されているかも示されるべきです。また、遺跡コードや地区番号の詳細、遺跡の特定地区のグリッド番号の割り当て、未使用のグリッド番号などを含む、調査区の状況、層、個別の番号付けシステムも含めるべきです。文書化では、使用したすべてのコードを定義し、使用された標準システムまたは制御された語彙を参照して拡張されなければなりません。

人工物あるいは環境データの再利用を可能にするために特に必要となるので、遺跡の発掘調査現場、または発掘後に行われるサンプリングの手順は完全に文書化されるべきです。例えば、サンプリングされた層、湿式または乾式ふるいにかけられた層は、採用したサンプリング方法の正確な詳細とともに文書化しなければなりません。標準的な取り上げレベル（人工層位）が採用されている場合も、それを文書化するべきです。小規模および大量の出土品を定義するための手順も文書化する必要があります。

実測図記録を作成するための手順には、図面の縮尺の詳細や、遺跡で採用されている単一、または複数の平面配置状況などを含む文書化が必須です。遺跡グリッドの詳細や国家座標系との関係、そして遺跡の基準面と海拔標高の関係も記録するべきです。加えて、地形測量の場合は計測点は斜面の上部か下部いずれか、また通常のグリッドの場合はグリッドサイズなど、調査手順の詳細を記録し、文書化する必要があります。

遺跡の説明、歴史的／考古学的コンテキスト

これは、考古学的な調査結果とその重要性の全体像を潜在的な利用者に提示するために必要となる背景情報です。適切な発掘調査またはフィールドワークの報告書の書誌情報を提供するだけで十分な場合も多いですが、少なくとも短い要約がデジタルアーカイブの一部として提供されていることが重要です。発掘プロジェクトの場合、ほとんどのプロジェクトのアーカイブでは、遺跡全体についての説明がデジタル形式で保管されることが想定されています。また、デジタルアーカイブの長所や短所が要約の一部に強調されていたり、あるいは別のファイルに記載されているならば、デジタルデータの再利用可能性に対して意味があるでしょう（例：「全体的なCAD範囲」、「GIS分析の見込みのあるすべての出土品の完全なジオリファレンス」）。

追加のドキュメント

これには、データセットの保存と再利用を促すものが含まれています。例えば、出版された報告書、灰色文献の短い報告書、あるいはノートの数ページをスキャンしたものなどです。これらは、メタデータの記録に欠けている情報や、メタデータを補足する情報、あるいはメタデータの記録よりも詳細な情報を提供するかもしれません。

[1] [http://www.data-archive.ac.uk/create-manage/document/overview\(*\)](http://www.data-archive.ac.uk/create-manage/document/overview(*))

プロジェクトのメタデータ

メタデータとは、文書、データセット、画像、その他の資料について説明する記述的な情報のことです。ファイル内容を記述するための標準化されたフィールドと用語からなるメタデータの提供は、アーカイブに保存されたファイルの探索、アクセス性、有用性を向上させます。さらにメタデータは、データセットや画像のような、追加情報がなければ使用できない可能性のある非テキスト資料のコンテキストを提供します。一般的にメタデータとは、構造化された方法でデータアーカイブの異なる側面を多様なレベルで文書化するのに使用可能な標準化された情報のセット（スキームまたはスキーマ）を指します。

プロジェクト、リソース、ファイルごとに、メタデータは一般的なものと特殊なものがあります。メタデータは最低限、ファイルやプロジェクトに関連する「誰が、何を、いつ、ど

こで、どのように」という情報を提供しなければなりません。メタデータの情報は、他の人がそのファイルやプロジェクトが必要なものかどうかを評価できるように、十分に詳細でなければなりません。一般的なメタデータタイプとしては、以下のようなものがあります。

- ・ファイルまたはプロジェクトの内容に関する一般的な記述情報
- ・プロジェクトやファイルの所在情報
- ・登録、アクセス、権利などを追跡するための管理情報
- ・ファイルの作成方法、使用したプログラム、ファイルサイズ、その他の詳細情報を特定するためのファイルや技術的なメタデータ

メタデータ作成の過程と理由は、多くの既存のガイドライ

ンの中で、また数多くの組織やリポジトリ（例：NISO 2004, Day 2005, Ballegoie & Duff 2006）によって文書化されていますが、簡単に言えば、メタデータの目的は、記述的な情報と文脈に沿った情報を保存することでデジタルリソースを容易に識別、取得、利用可能にすることです。

メタデータのタイプ

本ガイドの目的に照らし合わせて、メタデータをいくつかのカテゴリに大まかにグループ分けしました（以下に説明）。ほとんどの場合、これらのメタデータの種類のレベル（例えば、プロジェクトや個々のリソース）で収集されますが、一般的なものから特定のファイルレベルに至るまでの特定の要素が記録されることがあるでしょう。さらに特定のメタデータ標準は、いくつかのレベルで機能するメタデータの要素を記録する場合があります（例えば「作成者」は、リソースの探索と管理情報の提供に役立つでしょう）。考古学プロジェクトに関連するカテゴリとして、プロジェクトレベルのメタデータ、リソースレベルのメタデータ、ファイルレベルのメタデータの3つがあります。

プロジェクトレベルのメタデータ（「記述」メタデータと「リソース探索」メタデータを合わせたもの）は、使用する手法に関わらず、プロジェクト・アーカイブ全体のレベルで記録されており、時代名称や年代、遺跡や出土品のキーワード、プロジェクトの詳細、遺跡コード、地理的位置などの要素をカバーしています。多くの場合、この情報の多くはアーカイブ内の文書（遺跡報告書など）に含まれています。

記述メタデータまたはリソース探索メタデータは、データセットや文書の包括的な記述と容易な検索とを可能にするように設計されており、プロジェクトとその結果に関するものとなっています。ダブリンコア [1] 規格は、記述的な要素と資料探索に焦点を当てた要素を多く取り入れたメタデータ標準の良い事例です。

リソースレベルのメタデータ：アメリカを拠点とする tDAR を含む多くのアーカイブでは、アーカイブプロジェクトを構成する個々のリソースに関連するメタデータ要素の中間層を設定しています。リソースは、コンテキストに応じて、1つまたは複数のファイルを含む場合があります。簡単な例として、PDF 文書、画像、アクセス・データセットなどが考えられます。より複雑な例としては、センサスキャン・データや GIS シェイプファイル、または1つの機能的な「リソース」が複数のファイルで構成されているデータベースなどがあります。

リソースレベルのメタデータの目的は、記述的なものやリソース探索に重点を置いたプロジェクトレベルのメタデータと大幅に重複しますが、規模の大きいプロジェクトの中ではさまざまな詳細部分に焦点を当てています。これには個別の引用書誌、日付、場所、資料の種類、またはプロジェクト内のすべてのリソースに横断的ではないキーワードリストなどの要素を含む場合があります。ある意味、これらのメタデータは、プロジェクト全体ではなく個別のリソースにおいて報告された結果であり、一義的には探索を補助するために文書化されています。

ファイルレベルのメタデータ（「技術」と「保存」メタデータ

を含む）は通常、非常に個別で、その名が示すように個々のファイルレベルに適用されます。ファイルレベルのメタデータには、チェックサムなどの検証方法とともに、ハードウェアやソフトウェアに関する情報が組み込まれています。データをデジタルアーカイブに登録する場合、このメタデータを作成するのはアーカイブそのものです。同時にデータ作成者、データ作成の過程についての多くの情報を提供する必要があります。その性質上、技術的なメタデータはデータの種類ごとにかなり特殊なので、本ガイドの関連する章で扱います。

上記の3つのカテゴリの中に、4つ目のカテゴリである**管理メタデータ**が存在し、作成、取得、改変、バージョン管理などの要素をカバーしています [2]。このメタデータには、知的財産権に関する情報が含まれています。このような情報は、全体的なレベルで記録することができます（例えば、データセット全体の所有権は、一人の個人や組織が所持している場合が多い）。しかし個別の技術や、担当者や著者と知的財産権保有者が異なるデータセットになどについても記録するべきでしょう。

構造化スキーマの利点に加えて、特定の標準化された単語リストやシンボラスを使用することで、あらゆるレベルのメタデータの精度をさらに高めることができます。これらは、English Heritage NMR Monument Type Thesaurus [3] や MDA Object Type Thesaurus [4] のようなリソースで、メタデータスキーマの要素（例えば、ダブリンコアの「主題 (subject)」要素）を統制された方法で完成させるためのものであり、メタデータの使いやすさと信頼性を向上させます。本ガイドでは、特定のメタデータスキーマが指定されているところでは、条件を満たす要素のための適切なリソースを示しています。

プロジェクトレベルおよびリソースレベルのメタデータ

一般的なプロジェクトレベルまたはリソースレベルのメタデータは、優れた図書館カタログのように、利用可能なリソースを迅速かつ容易に識別し、必要なものに接続することを可能にします。しかし、これが効果的に機能するためには、メタデータは正確に標準形式で実装されなければなりません。プロジェクトレベルのメタデータに一般的に（ADS でも）使用されている形式は、ダブリンコアです。「基本版」と「詳細版」の両方があるこの標準には地理的範囲、時間的日付、方法論、記念物・証拠の種類など、プロジェクト全体の詳細な概要を提供することができる15（適格版では18）の中核的な要素で構成されています。tDAR 内では、基本的なダブリンコアのメタデータスキーマが拡張され、考古学固有のフィールドだけでなく、メタデータオブジェクト記述スキーマ (MODS) [8] に基づくフィールドも含まれるようになりました。プロジェクトの初期段階では、メタデータを正確に編集できないことが多いですが、しかしプロジェクトに従事する方々は、一般的に必要なプロジェクトレベルのメタデータの種類に精通しておくべきです。データをアーカイブに登録する場合、寄託者は、関連分野に固有のテンプレート（例えば、地球物理探査）とともに、プロジェクトまたはリソースレベルのメタデータ記録テンプレートを完成させた状態にしておくことが求められる場合があります。

以下の例(表1)は、ADSのプロジェクトレベルのメタデータの基礎となるもので、基本的なダブリンコア要素で構成されています。tDARにおいて、これら一連のメタデータ要素は、プロジェクトレベルで収集された情報だけでなく、多くの場合、個々のリソースのために収集された情報の両方に対応しています(例えば、プロジェクトの名称は、リソースレベルでは個々の文書やデータセットの名称にゆるやかに対応しています)。また、tDARのメタデータ入力フォームは、このリストを拡張して、表2にグループ化されたダブリンコアな

詳細を獲得していることも注目に値します。tDARメタデータのより詳細な情報は、tDAR Data Dictionary [9]の中で確認できます。

ファイルレベルのメタデータ

ファイルレベルのメタデータは、基本的に、アーカイブがファイルを保存し公開するために必要な情報ですが、加えて、利用者はデータセット内のファイルの性質や再利用可能性を知ることができます。ファイルレベルのメタデータは、

表1：ADSプロジェクトメタデータ

要素	説明
プロジェクト名称	データセットの名称(およびそれに代わるもの)。
内容記述	データセットの内容についての簡単な要約記述と、データ収集を行った研究プロジェクト(または代替プロセス)の主なねらいや目的の簡単な要約。
主題	データセットの主題のキーワード(English Heritage NMR Monument Type ThesaurusやMDA Object Type Thesaurusなどを使用)。
時空間範囲	空間的範囲と時間的範囲の両方を含む。 空間的範囲には、データ収集を行った国、地域、郡、町、村の現在および同時代の名称を含み、可能なら Getty Thesaurus of Geographic Names [5]などの標準リファレンスを使用する。データの対象になった時期の名称や行政単位が現在と異なっていた場合は、別途記録する必要がある。地理座標は、さまざまな方法で、英国国家座標のコードとして入力可能である。例えば、ポイント(中心座標により小規模なプロジェクトエリアの記述に便利)、ライン(サイトの直線的限界を表す少なくとも2つの座標)、ポリゴン(より複雑なサイトエリアでは境界を表すのに3つ以上の座標を使う)。場合によっては郵便番号を含めることもできる。 時間的範囲については、データセットの対象となる日付/期間(可能なら RCHME Period Listのような既存のシソーラスを使って)を含める。
作成者	作成者、編集(編さん)者、資金提供機関、データ収集に関して知的に責任を負う団体や個人の詳細。情報には、姓、名、所属、住所、電話、ファックス、メールまたはURLが含まれる。
公開者	そのデータを公開した組織の詳細。
コントリビューター	リソースの形成に貢献した人や組織。
識別子	データセットを識別するために用いるプロジェクト番号または参照番号。
ソース	リソースの元となった重要な先行業績。
日付	データセットが作成された日付、考古学的プロジェクトが行われた日付、処理日、または電子化した日付などを必要に応じて示す。
著作権管理	データセットの著作権者名。コレクションが従業員の職務として作成された場合は、著作権者は通常、雇用主になる。資料が特別な著作権(イギリスの国王著作権など)で守られている場合は、これを表示する必要がある。
リレーション	データコレクションが、印刷物または機械可読なデジタルデータのいずれでも、他の公開または未公開ソースの全部または一部から派生している場合、リレーション要素には、オリジナルのソースの参照情報、ソースの保管場所と識別方法の詳細(受入番号など)を含めなければならない。コレクションが他のソースから派生している場合は、そのデータの転写/複製が全体なのか一部なのか、およびデジタル化に使用された方法を表示する。またデータコレクションそのもの、あるいはそれに基づく出版物への完全な参照も含める。
言語	データセットが書かれている言語(英語、フランス語、スペイン語など)。
リソースの種類	データセットが、一次(プライマリ)データ、処理済データ、解析結果、最終報告のいずれであるか。
記録形式	データの保存形式(WordPerfect 5.1、HTML、AutoCADなど)。

このタイプのスキーマの完成例はADS Guidelines for Depositors [6]にあります。

表2：tDAR一般メタデータ

要素	説明
基本情報	基本的なメタデータには、リソースと、それが含まれるより大規模なプロジェクトとの関連付け、ライフサイクルの中の状態（ドラフト、アクティブ、削除など）、言語、作成年、要約／説明、物理的保存場所が含まれる。
書誌メタデータ	ドキュメント固有、MODS (Metadata Object Description Schema、訳注：米国議会図書館ネットワーク開発・標準局が開発した XML ベースの書誌情報記述スキーマ) メタデータフィールドを使用してリソースを記述。
リソース作成者	リソースへの貢献者である個人や団体を正しく表示するために使われる。「役割」はリソースに対する個人の主要な役割を表す（作成者、編集者など）。
リソース固有あるいは機関の識別子	利用した機関識別子やプロジェクト識別子を記述する (Smithsonian Trinomial、AZ State Museum—ASM 番号など)。
調査の種類	リソースに関連する全ての調査の一覧を表示する（例えば、リサーチ・デザイン、遺跡のモニタリング、データの回復または発掘など）。
遺跡記述情報	遺跡名、遺跡タイプ（統制語彙による）、キーワードが含まれる。
資料材質の種類	リソースがカバーする遺物の種類（陶磁器、動物、金属、年代測定資料など）。
文化名称	考古学的文化（パレオインディアン文化、アルカイック文化など）と、ユーザーが作成した値（名称）を含む。
時間的範囲	時期区分名称（プエブロ IV 期など）、年代範囲、年代値の種類（例えば、暦年代、放射性炭素）、開始／終了年と詳細（例えば、較正值かどうか、誤差範囲など）を含む。
全般的キーワード	他のメタデータでカバーされていないリソースの側面を記述するためにユーザーが作成した値。
空間名称	地理的名称（デスヴァレーなど）、座標を含む。
リソース提供者	リソースのアーカイブおよび公開を承認した機関。
個人と組織の役割	リソースに関わった人物の名前、現在のメールアドレス、組織と役割。

表3：一般的な最小限のファイルタイプメタデータ

要素	説明
ファイル名	ファイルの名称 例：report.doc
ファイル形式	ファイルの形式 例：PDF/A や Open Office Document
ファイル作成に使用したソフトウェア	ファイル作成に使用したファイル 例：Microsoft Word 2007
ファイル作成に使用したハードウェア	ファイル作成に使用したハードウェア。これは、ファイルがレーザースキャナーや GPS デバイスなどの調査機器によって直接作成される場合、より重要になる。
ファイル作成に使用した OS	ファイルを作成した OS 例：Windows XP や Mac OS X 10.5。
作成／最終ファイル更新日	ファイルがいつ作成された、または更新されたか。
処理履歴やリネージュ	この要素は、ファイル間の関係、またそのファイルがソースファイルであるか他からの派生であることを示すために使用される。

記録されるデータの種類やファイルの種類に大きく依存するものですが、一般的にはファイルの作成に使用されたソフトウェア、ファイル名のリスト、ファイル間の関係などの要素に関する注記などで構成されています。

ADSの『発掘・フィールドワークからデジタルアーカイブへ Guide to Good Practice』の初版では、デジタルアーカイブ内の各ファイルはメタデータ記録を持つことが推奨されていますが、経験上、ファイルレベルの文書化はほとんど不要であることが実証されています。データの性質にもよりますが、同じ形式のファイルのグループや、特徴を共有するファイルグループは、1つのメタデータ記録で文書化することができます。ファイルレベルで文書を提供することは作成者に

大きな負担をかけ、アーカイブ構築の必要時間、さらにコストを大幅に増加させることが認識されていました。またデジタルデータの利用者は、考古学プロジェクト全体や特定のカテゴリーの情報を検索する可能性が高く、個々のファイルのメタデータの必要性は低いことが研究で示されています。

また、米 tDAR のメタデータスキーマは、考古学プロジェクトからアーカイブされる可能性のある多種類の記録やリソースに対して、メタデータ記録を維持するためにさらなる努力をする意義があると認めています。例えば、遺跡報告書などの文書や出土品の分析から生成されたデータセットは、個々のリソースに関連付けられたメタデータ記録によってより活用できます。しかし同時に、すべてのファイルに対して

表4：ファイルレベルの技術メタデータ

要素	説明
固有 ID	自動的に生成された固有 ID 例:1234567
ファイルの位置	ファイルパス すなわちディレクトリとファイル名 例:/adsdata/cottam_ba/jpg/fwking_plan.jpg
チェックサム・タイプ	使用するチェックサムアルゴリズム 例:MD5, SHA-1 など。
チェックサム値	アルゴリズムにより生成されたチェックサム値 例:578cbb18f73a885988426797bcab8770
プロジェクト ID	固有のプロジェクト ID 例:ADS-123
作成日	チェックサムが作成された日付 例:16-May-2006
作成者	チェックサムを作成した人物 例:Doe, J
最終検証日	ファイルが最後にチェックまたは検証された日付 例:16-May-2007

表5：ファイルレベルのプロセスメタデータ

要素	説明
プロセス ID	自動的に生成された固有 ID 例:1234567
プロジェクト ID	プロジェクトの固有 ID 例:ADS-123
ソースファイル形式	オリジナルファイルの形式 例:.txt
変換先フォーマット	宛先の形式 例:.shp
プロセス実行者(機関)	処理をした人 例: Doe, J
プロセスコメント	行われるプロセスに関連したコメント 例:「WGS84に参照」
プロセス開始日	プロセス開始の日付 例:17-May-2007
プロセス完了日	プロセス完了の日付 例:17-May-2007
プロセス記述	プロセスの記述 例:「分析目的と研究成果の普及のため XYZ データを ArcView へインポート」
プロセスガイドライン	プロセスと関連したガイドライン
プロセス使用ハードウェア	ファイル処理に使用したハードウェア 例:Viglen Genie Intel Pentium 4
プロセス使用ソフトウェア	ファイル処理に使用したソフトウェア 例:ESRI Arcview 9.1
プロセス入力	ソースファイルの完全なファイルパス 例:/adsdata/pro-453/xyz/file.xyz
プロセス出力	出力ファイルの完全なファイルパス 例:/adsdata/pro-453/shp/file.shp
プロセス結果	処理結果についてのコメント 例:「成功」
プロセスタイプ	実行したプロセスの記述 例:「変換—公開」

個別のメタデータを維持する必要はないでしょう。例えば、本質的に同じリソースである文書の 2 つのバージョン（1 つはアーカイブや研究目的のための完全なもので、もう 1 つは無許可の閲覧者から機密コンテンツを保護するために編集したもの）がデジタルアーカイブに保存されている場合などです。また画像のようなファイルタイプには、ADS が提唱しているものと同様に主にプロジェクトレベルのメタデータを用いることで、より効率的に処理することができます。ファイルレベルのメタデータは、データセットの中に何があるのか、そしてそれがどのように利用できるのかを正確に識別するために不可欠なものです。データ固有のガイドラインが本ガイドの他の箇所では示されていない場合には、以下の要素をファイルレベルで最低限記録することを推奨します。

外部アーカイブに保存された場合、この情報はアーカイブの内部ファイル管理システムに入力され、データの将来の移行と検証の手順を計画するために使用されます。この時点で、ファイルレベルのメタデータは、固定値やチェックサムなどの要素によって拡張されます。これらは「冗長性チェックの一つの形であり、データのエラーを検出し、データの整合性を保護する簡単な方法」[7] です。MD5（メッセージ・ダイジェスト・アルゴリズム 5）と SHA（セキュア・ハッシュ・アルゴリズム）は広く使われている暗号化ハッシュ関数であり、これらのアルゴリズムをファイルに適用すると、（ほぼ確実に）固有ハッシュ値またはチェックサム値が生成され、データの検証と確認のためのメカニズムを提供します。分離されたチェックサムは、それだけでは役に立ちませ

ん。チェックサムは構造化データとしてファイル、場所、プロジェクトに関連付けられていなければなりません。以下のスキーマは、このようなファイルレベルのメタデータを保持するために使用できる構造の一例です。

より詳細な固定情報に加えて、アーカイブ（および場合によってはデータ作成者）は、アーカイブ実践の一環としてプロセス履歴を維持しようと考えられるでしょう。例として、XYZ座標データをGISにインポートすることが挙げられます。これも、単純な構造化データとして記録することができます。同じ構造であれば、ファイルレベルとバッチ処理の両方の情報を保持することができます。表5は、AHDS (Art and Humanities Data Service) の実践に基づいたものです。

- [1] <http://dublincore.org/>
- [2] <http://www.dcc.ac.uk/resources/briefing-papers/standards-watch-papers/what-are-metadata-standards>
- [3] [http://thesaurus.english-heritage.org.uk/thesaurus.asp?thes_no=1\(*\)](http://thesaurus.english-heritage.org.uk/thesaurus.asp?thes_no=1(*))
- [4] [http://thesaurus.english-heritage.org.uk/thesaurus.asp?thes_no=144\(*\)](http://thesaurus.english-heritage.org.uk/thesaurus.asp?thes_no=144(*))

asp?thes_no=144(*)

- [5] <http://www.getty.edu/research/tools/vocabularies/tgn/>
- [6] Collection-level Metadata <http://archaeologydataservice.ac.uk/advice/guidelinesForDepositors>
- [7] <http://en.wikipedia.org/wiki/Checksums>
- [8] <http://www.loc.gov/standards/mods/>
- [9] <http://dev.tdar.org/confluence/display/TDAR/Data+Dictionary>

引用・参考文献リスト

- van Ballegooye, M. and Duff, W. (2006) 'Archival Metadata' in S. Ross & M. Day (eds.) **DCC Digital Curation Manual**. <http://www.dcc.ac.uk/resources/curation-reference-manual/completed-chapters/archival-metadata>
- Day, M (2005) 'Metadata' in S. Ross & M. Day (eds.) **DCC Digital Curation Manual**. <http://www.dcc.ac.uk/resources/curation-reference-manual/completed-chapters/metadata>
- NISO (2004) *Understanding Metadata*. NISO Press. <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>

データの選択: データ変換のポイント

データの選択は複雑なテーマです。特定のファイルのみを選択してアーカイブすることは費用対効果が高いかもしれませんが、未来の研究者にとって選択されなかったファイルやバージョンの方が重要になるかもしれないですし、新しいツールが既存のデータのよりいい分析を可能にするかもしれません。センサースキャンや録画動画などの複雑なデータタイプの場合、生データが最終製品に変換されるまでに大きな変化を受ける可能性があり、何をアーカイブするのかという選択の問題は非常に重要になります。

「データ変換のポイント」という概念は、データが大きく変更される機会を明確にします。取得、フィルター、変換、または普及などのタイミングが含まれるでしょう。データ変換のポイントは、データのための「証拠の連鎖」を提供するもので、理想的には、繰り返しの変換を保証するソース、結果、手順または方法を取得します。データ変換のポイントは、複雑なデータタイプに適用可能であることは明らかですが、画像のように生データを処理することで最終的な状態が大きく変化する可能性のあるソースにとっても重要です。

ここまでの各章では、プロジェクト完了後のデータに関連するアーカイブ戦略のための最低限の要件を概説しました。強調した通り、その戦略には以下が含まれるべきです。

- ・プロジェクト計画とデータ作成段階において、保存と再利用のための明確な移行パスを有する適切なファイル形式の識別と管理。
- ・保存・再利用を促すための適切な文書化やメタデータの

作成、およびプロジェクト期間中の管理・運営のサポート。

ある過程のデジタルでの成果物を長期保存することが望ましいと識別されたら、それはプロジェクトの初期段階から取りかかるべきタスクとなります。一方で、どのようなファイルを、プロジェクトのライフサイクルのどの段階において保存するのかを識別することは、プロジェクトの設計における重要な検討事項であり、データを作成、取得する実行段階の構成要素でもあるべきです。

データ変換のポイント (PIP)

複雑なデータ取得や分析作業の場合、各過程における重要なステップの後にデータのコピーをアーカイブすべきです。データ選択という概念は、データが1つまたは複数の手法によって取得され、最終的な「製品」が作成されるまでにいくつかの段階を経て統合または再処理される複雑な考古学プロジェクトにおいて特に注意する必要があります。プロジェクトのライフサイクルには、減算（デシメーション）、集計（アグリゲーション）、実行後再変換（リキャストリング）、注釈（アノテーション）、データ形式の移行（マイグレーション）など、データの変換をともなう一連の段階があるので、それに対応する複数のデータ変換のポイント（PIP）が存在し得るでしょう。

この過程の一例を以下に図示します（図1、VENUS プロジェクト [1] からの引用）。データの流れは、図中で左から右への処理に沿って、調査現場で様々な技術（ハードウェア

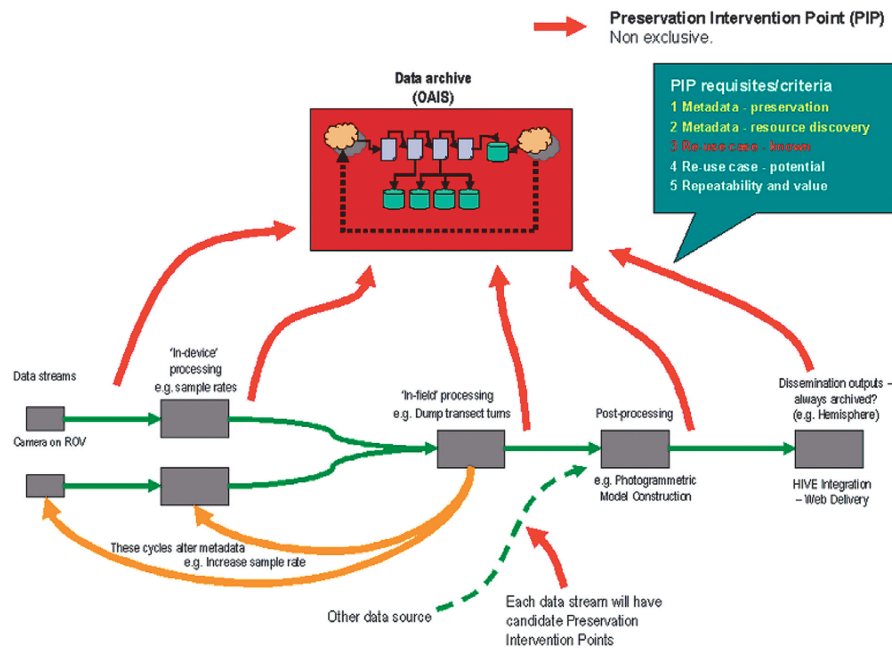


図1：データストリームとデータ変換のポイント。

アベースのもの）によって生成され、プロジェクトの公開版の成果物が作成されるまでの間に一連の変化を経ていることがわかります。図に示されている段階の例は、(VENUS プロジェクトの場合でさえ) 包括的なものでも決定的なものでもありませんが、以下のようなものがあります。

- ・ **データストリーム生成**—例えば、ROV カメラによる画像取り込みや水中超音波装置による水深調査、DGPS や電波三角測量による機材固有の位置情報など。
- ・ **装置内処理**—例えば、撮影時のサンプルレートや照明条件は変更され得るでしょう。可変（調整可能）であるため、これは処理とみなすことができ、機材によっては、取得された情報を破棄することも可能です。
- ・ **調査現場での処理**—例えば、データは対象領域外であったり、またはサンプリング基準が変更されると破棄されます（この段階で、または機材設定を変更することによって機材内処理が変更されることで、フィールドバックループが生じます）。
- ・ **後処理**—例えば、1) XYZ 座標を不規則三角形網 (TIN) に変換したり、加工データポイントを用いた数値地形モデルを作成するために使用します。2) 取得した画像からフォトグラメトリ技術によって理想的な 3 次元モデルを作成します。
- ・ **公開版の作成**—例えば、特定の公開形態として 3 次元モデルを作成します。

これは、考古学プロジェクトのライフサイクルにおいて行われるかもしれない作業段階の一部しか示していませんが、ほとんどの場合、データ作成は単純ではなく、保存用データのコピーを作成すべき段階がいくつもあることを浮き彫りにしています。

データの文書化とメタデータもまた、PIP コンセプトの重要な構成要素です。一般的には、データをできるだけもとの状態で保存することが望ましいとされ、作成後に変換・変更があった場合は適切な文書化により再現できることを前提と

しますが、あらゆるデータに当てはまるわけではありません。このアプローチがうまくいかない例として、一連の画像を使って三次元出力を構築する写真測量データがあります。このような場合、三次元出力（例えば、アンフォラのモデルやDTM）は、一連の高解像度画像から構築されますが、出力が作成される手順は、独自のものであったり再現性のないものであったりします（例えば、特定のソフトウェアパッケージ内でしか再現できない場合など）。この場合、元の画像と三次元出力の両方がデータ変換のポイントとなります。より複雑な例では、さらに多くのデータ変換のポイントが存在する可能性があります。多数の潜在的な PIP が特定されるならば、各データストリームに最適な PIP のみが選択されるように、一連の基準に照らしあわせて判断するべきでしょう。PIP を判断する大まかな基準は以下の通りです。

- ・ **保存メタデータ**：データが単に保存可能になるだけでなく、再利用可能になるよう適切なレベルのメタデータが存在しなければなりません。
- ・ **リソース探索メタデータ**：各ポイントにおいてデータが、データセットの他の部分と有意に区別し識別できるように、適切なレベルのリソース探索メタデータが存在しなければならない（これはおもにレガシーデータに適用されます）。
- ・ **識別可能な移行パス**：すべての段階でデータの移行オプションが明確でなければならない。
- ・ **再利用ケース**：これはおそらく最も重要ですが、時には判断が困難な場合もあります。データ作成者以外の研究者やデータが取得されたのとは異なる状況でもデータが確実に利用できる形になっているとき、再利用される可能性があるかどうかを検討課題です。その時に厄介な点は、ある種のデータについては、その時点で規定されていなくても将来的に再利用される可能性があり得るということです。その例として、開発中の後処理技術に適したデータや、将来的に可能になる（あるいは既存の技術が強化される）ことが想定されるデータなどが挙げられ

ます。

- ・**繰り返し**：データの作成手順は繰り返されるか。もし繰り返されるならば、より早い段階が適切な PIP であるかもしれません。もし繰り返されないのであれば、その時点を紹介ポイントとして選択すべきです。
- ・**保有ポリシー**：データはターゲットのアーカイブの保有ポリシーと一致している必要があります。
- ・**価値**：特定の時点でデータ保存のために介入する際のコストのことで、どのプロジェクトも予算には限りがあることを前提とします。同時に、アーカイブ化される資料の価値も意味しています。例えば、特に高価で再現が困難な繰り返し工程で作られデータは保存する価値があるかもしれません。したがって、価値は、アーカイブのコストに対するデータそのものに認められる価値のバランスとして考えられるべきです。

上記の基準は重要度にもとづき配列されているのではなく、各基準は他の基準とバランスをとる必要があります。プロジェクトのデータライフサイクルを検討する手順は、可能ならば、データ作成者とアーカイブとの間で協議されるべきです。

データ自体の性質に加えて、ファイルサイズ、保存コスト、著作権、機密性などの問題は、最終的なプロジェクトアーカイブに保存するデータの選択に影響を及ぼすでしょう。これについては、次の章で扱います。

[1] <http://archaeologydataservice.ac.uk/research/venus>

プロジェクトのアーカイブ:保存と公開

本章では、考古学的文書やデータセット、その他のリソースの長期保存と公開に関する主な要件と問題点を説明します。本ガイドでは、データ作成者の大多数が最終的な目的として、適切なデジタルアーカイブやリポジトリに資料を預けようとしていることを前提としています。いずれにしてもアーカイブの評価と構築、プロジェクトの準備の際に、多くの要因を検討することは有益です。検討要因には、保存、バックアップ、帯域幅、アクセス、そしてセンサースキャンの結果のように大規模または複雑なデータなどがあります。

データ、リソース、アーカイブのライフサイクルにもよりますが、保存、バックアップ、公開計画は単純で、コンピュータ、外付けハードドライブのセット、またはバックアップ媒体を含む場合があるでしょう。時間の経過とともに立ち現れるライフサイクルのあらゆる側面や、または収集されたデータの複雑さによって、いくつかの課題はより難解になることがあります。数年以上にわたるプロジェクトの場合は、ハードウェアの交換や移行にかかるコストも同様に考慮する必要があります。さらに大規模なプロジェクトでは、適切な保存、バックアップ、公開のインフラを追加する必要があります。長期的にみるとストレージは比較的安価であり、20 年以上の期間になるとそのコストは無視できるほどのものになるでしょう。

大量のデータを扱うプロジェクトでは、ネットワークインフラも重要になります。アーカイブと同様に、(バックアップ、公開、保存のための) データ転送もプロジェクト成功の要になるでしょう。特定のケースでは、ファイル、特に大容量のデータファイルを転送する際の物理媒体の使用とチェックサム(訳注: データ転送が成功しているかのチェック)が、転送の速さと正確性を確保するために有用です。

「ビッグデータ」の概要

考古学における「ビッグデータ」の概念は、並外れたサイ

ズをもつファイル形式と、特にその保存と配信に関する技術に象徴されています。このようなデータの作成と利用は、考古学や文化資源管理活動の分野(例えば、海洋考古学や測量、レーザースキャン、LiDAR、コンピュータによる可視化、その他の科学研究)で増加しています。しかし、データの保存、公開、再利用、アクセスにおけるコストやグッドプラクティスの意味はほとんど理解されていません。この理解の欠如は、考古学や文化資源管理の分野でよく見られるように新しい調査技術の導入に際して商用(プロプライエタリ)なファイル形式を使用することによってさらに悪化する可能性があります。

考古学コミュニティに対するビッグデータ技術についてのガイダンスの不足に対して、ADSは2005～06年にビッグデータ検討プロジェクト [1] を実施しました。これまでのアーカイブガイダンスでは、ビッグデータを扱う技術は取り上げられていませんでした。そして、最終的な報告書とファイル形式のレビューでは、現状の課題に対する解決策の提案に向けて大きな進展がありました。本章では、一般的なデータの保存と公開の文脈において、報告書で示された課題をいくつか取り上げます。なお読者は、ビッグデータに関するより幅広い議論と背景について知るために、ビッグデータ検討プロジェクトの報告書も参照することが推奨されます。また、特定のデータタイプについての推奨事項は、関連する技術(海洋リモートセンシングやレーザースキャンなど)の項目で取り上げるため、詳細な説明は省きます。ビッグデータ報告書では、これらの技術について、長年の経験を持つ考古学以外の調査研究組織から学ぶことができる点に着目しており、本章でも可能な限り事例を紹介します。

保存、バックアップ、データキュレーション

表面的には、テラバイトの容量を持つ外付けハードドライブが安価で入手しやすくなっているため、ほとんどのデータセットの物理的保存は問題ありません。しかし、多くのアー

カイク組織では、データのバックアップ（内部・外部）に多額の投資を行っています。例えばADSは、Legato NetworkerとAdic Scalar Tape Libraryによるヨーク大学のバックアップサービスを利用しており、イギリスデータアーカイブ(UKDA)にデータのコピーを保持しています。前章「デジタルデータ作成計画」で明示したように、プロジェクトのデータセットを保存するための基本的な戦略は、大容量のハードドライブをオフライン環境に置いて複製を保存し、定期的にマスターデータと同期するといった単純なものです。前述した通り、この戦略は、信頼性の高いバージョン管理や、ウイルス、その他の破損の原因からデータを確実に守るためのその他の手順で補足されるべきです。データの物理的保存における課題は相対的なものであり、データセットを保存する組織の利用可能なリソースにもとづいて検討されるべきでしょう。ファイル数やサイズの上限値は、考古学では、より多くのリソースを扱いビッグデータを扱うことに慣れている地球科学などの他の分野に比べるとはるかに低いでしょう。現在、数十ギガバイトに達するデータサイズは、アクセス性と長期保存（利用可能性）の点で考古学者にとって問題と考えられますが、ストレージ利用可能性とデータセットの利用の双方について、今後より大きなデータの取り扱いが想定されるでしょう。

ハードウェアとデータリフレッシュ

「ストレージ」にはデータを保存・バックアップする媒体のサイズだけでなく、データのリフレッシュメント（新しいハードウェアやソフトウェア環境へのデータセットの移行）の定期的な実施という要素も含まれていることに注意する必要があります。物理的保存のコストは下がり続けていますが、データのアクセス性と利用可能性にとって重要なリフレッシュメントと長期的な管理のコストは下がっていません。今後も、技術の進歩に対応するとともに維持管理などのコストを削減するために、アーカイブは定期的にシステム全体、または一部をアップグレードしなければならないでしょう。

テープなど、物理ディスクストレージとバックアップ媒体のコストは急速に低下しています。5年後には、ギガバイト単位のコストはゼロとみなされるほどに減少するだろうと、現在にいたる傾向の分析から示唆されています（Gilheany n.d.）。しかし、保存に関連するシステムの資産コストは、継続的なメンテナンス、バックアップ、保険コストなどが実体です。経験上、アーカイブシステムがより高度になり、定期的なリフレッシュメントに際して対象とするアーカイブが占める割合が小さくなっていく（アーカイブ全体が拡大することを想定）につれて、既存リソースのリフレッシュメントのコストは時間とともに部分的に減少していくと考えられます。しかし、リソースのリフレッシュメントのコストが徐々に低下しても、アーカイブやリポジトリの継続的で積極的な運営と管理のために必要な人的資源のコストの増加（すなわち賃金の増加）によって、その一部は相殺されるでしょう。

ADSの10年の歴史の中で、使用する機器は第3世代目に移行しており、基本的には5年ごとのアップグレード周期の中で稼働しています。新しい機器の購入だけでなく、そのセットアップや旧システムからの移行にかかる人員、作業時間という点でもコストがかかります。保存の長期的なコストの計算は概念的には難しいことが多いですが、5年周期の更新に

基づくならば、データセットが100年間維持されるには20回のリフレッシュメントが必要になります。私たちが保持し続けている、価値の高い紙媒体のアーカイブ（歴史史料）の多くがかなり古いものであることを考えると、デジタルデータセットも今後長期にわたって維持する可能性があると言えるでしょう。

現在のADSポリシーでは、いまのところ（データの保管のコストは）一回1メガバイトあたり50ペンス（約63セント）ですが、これは初期のアーカイブのものと近似しています。しかし、近年のシステムのアップグレードの進展により、コストは大幅に削減できると示唆されています。この「メガバイトあたり」のコストは、上記で説明した内容をまとめたものであり、「継続的な管理と更新」と表現したほうが良いのかもしれませんが、以下は、ライフサイクル管理の現状とそれに関連した保持・破棄についてのポリシーをより正確に、しかしシンプルに表現したものです。

保持期間	リフレッシュのコスト
5年	$R + E$
10年	$R - DR + E - DE$
15年	$R - 2DR + E - 2DE$
20年	$R - 3DR + E - 3DE$
25年	$R - 4DR + E - 4DE$

R = リフレッシュのコスト
 DR = リフレッシュのコストの減少
 E = 物理的機器・装置のコスト
 DE = 機器・装置のコストの減少

例えば、 $R = 9p$, $DR = 3p$, $E = 4p$ そして $DE = 1p$ （メガバイト当たりの価格は例にすぎません）とした場合、次表のようになるでしょう。

保持期間	リフレッシュのコスト(ペンス)	累計(ペンス)
5年	$9 + 4 = 13$	13
10年	$9 - 3 + 4 - 1 = 9$	22
15年	$9 - 6 + 4 - 2 = 5$	27
20年	$9 - 9 + 4 - 3 = 4$	28
継続	—	30

このように、1回あたりの保存コストは20年後には無視できるほどに小さくなると考えられます。この結果は、シンプルなモデルが導き出したものでもあり、実際には保存ポリシーに基づくリソースのリフレッシュメント、管理、運営などの継続的なコストもかかります。したがって、1メガバイトあたり30ペンス（約38セント）の1回あたりのコストは、この先20年にも適用できるでしょう。ADSのポリシーは、現在のところ、ADSに預けられたすべてのデータを永続的に保存するために「ベストエフォート」（最善努力）が行われる（すなわち、上記の20年間のコストモデルに従う）ことを前提としています。しかし、場合によっては、資金提供機関が一定期間を超えた保存を要求しなくなる可能性もあり、定期的な見直しが必要になるでしょう。データセットを廃棄する

理由としては、保存期間が指定されていた、他のリソースに置き換えられたり統合されたりした、価値がないと判断された、保存を継続するには現実的な方法がなかった、などが考えられるでしょう。

「ビッグデータ」特有のハードウェアとリフレッシュの問題

データが、適切なアーカイブパッケージ（適切なファイル形式、十分な文書化、メタデータ）となっている限り、ビッグデータのアーカイブは、そうでないデータに比べて人的リソースの消費を低減させるでしょう。例えば移動に使用する媒体からアーカイブ環境にデータを取り込む場合、物理的なサイズの大きさからファイルの移行に多くの時間がかかるでしょう。同様に、固定値やチェックサム値を生成して転送が成功したかを確認するのは、ファイル内の各バイトが参照されるため、はるかに長い時間がかかります。しかし、これらの手順はいずれもバックグラウンドまたは自動化タスクとして実行することができるため、「人的コスト」を削減できる可能性があります。定義上、ビッグデータの物理的なストレージ要件は、従来のデータセットよりも大きいです。しかしながら、ビッグデータアーカイブは、ストレージ料金がメガバイト単位で評価され、移動・読み出し（インジェスト）コストがファイル数に基づいているような課金体系には適しているでしょう（訳注：単一ファイルに多量のデータが収納されているため）。

データの公開

データセットを幅広く公開することは、作成者とアーカイブの間でのデータの転送と同様に問題が多いと見られています。これは特にビッグデータの場合に当てはまるでしょう。多くのユーザーが望むオンラインでのファイルへアクセスは、高帯域幅システムに組み込まれそのリソースを利用できるアーカイブ組織にとっては問題ではありません。しかし、ビッグデータの利用について、一般のユーザーが同じように高帯域幅の接続にアクセスできるとか、世界各地からのアクセスが同じ条件的に同じであると仮定するのは危険です。このため、ユーザーが意図せずにネットワークに影響を与えないように、多くのアーカイブでは、ダウンロードできるファイルサイズを制限したり調整しています。時には、適切な接続が確認されたユーザーのために、特別な対応（例えばhttpやftpダウンロード）により、より大きなファイルをダウンロードできるようにすることもあります。

多数の小・中サイズのファイルで構成されている大規模なデータセットは、アーカイブの管理者、ダウンロードしようとするユーザーにとってほとんど問題はありません。しかし、例えばレーザースキャンによる点群データ、詳細な衛星画像や水深データなどの大容量ファイルはより複雑な問題を抱えています。容量の小さい一つのファイルであれば、インターネット回線が遅い場合でも、ユーザーは個別にダウンロードでき、内容を確認できます。対照的に、大容量ファイルは、ダウンロードするのに時間がかかるため、ユーザーとアーカイブの双方のリソースに大きな負担がかかります。大容量ファイルは「要約」することが困難であり、ユーザーがダウンロードする前にファイルに何が含まれているのかを知ることが難しくなりますが、多くの場合はユーザーが必要としているのはファイル全体ではないかもしれません。

ネットワーク

大容量ファイルへの安定したアクセスを提供し、ユーザーとアーカイブ両者の負荷を軽減するという観点から、ビッグデータプロジェクトの一環として、ビットトレント（BitTorrent）^[2] などの一般的なP2P（ピアツーピア）方式を含む様々なネットワーク技術が検討されました。こうした方法は分散型アーカイブモデルをサポートする可能性がありますが、データ（特にビッグデータ）の再利用は、小さなコミュニティ内での限られた活動になる可能性が高いと結論付けられました。従って、制限されたダウンロードやP2P方式は、ダウンロード速度や信頼性の向上にほとんど寄与しないと考えられました。JANET Lightpath service ^[3] のような高速「アクセスポイント（PoA）」光ネットワークも検討されましたが、実装コスト（数千ポンド）のため、これ以上は進められませんでした。これらは、ユーザーによるビッグデータへのオンラインアクセスを可能にすることの難しさを説明する事例です。

ウェブサービスとグリッドコンピューティング

ウェブサービスとグリッドコンピューティング技術は、データ公開のより柔軟なアプローチを可能にしますが、アーカイブのコストも増加します。National Geoscience Data Centre GeoIndex ^[4] や British Atmospheric Data Centre Data Extractor Tool ^[5] などのウェブサービスを利用すると、ユーザーはオンラインでデータセットを検索したり、データのサブセットを定義し、抽出したりすることができます。さらに、NERC Earth Observation Data Centre ^[6] のような他のウェブサービスやグリッドベースのアプリケーションにはユーザーがダウンロードの段階をスキップできるツールが盛り込まれています。ユーザーはデータセット全体をダウンロードすることなく、様々な方法でデータを編集し、他の形式に変換したりすることができます。このようなサービスはデータのプレビューやダウンロード機能（さらにはユーザーのソフトウェアアクセス）の問題を改善する大きな可能性を秘めています。こうしたツール提供は、アーカイブに大きなコストをかけることになります。

物理メディア

現在、大規模なデータセットの公開に最もよく使われている方法は、DVD（小さい容量の場合）や外付けハードドライブ（大きい容量の場合）などの持ち運び可能な媒体に対応したものです。既に述べたように、1テラバイトのポータブルハードドライブは比較的安価に入手可能であり、提供することも返却してもらうことも可能です。

大容量ファイルを取得するには、帯域幅の占有や媒体の準備コストなど何らかの費用がかかる可能性が高いです。利用しようとするユーザーは、データが自分の作業に関連するかどうかを確認する必要があります。従来は、国立海洋学センター（National Oceanography Centre）SeaDOG データベース ^[7] のような大規模なオンラインの記述的メタデータカタログを利用して、ユーザーがデータ保有者を確認してデータのダウンロードを要請できるようにしてきました。サムネイル画像や動画クリップなどの「見本（tasters）」の利用も、データをダウンロードするかどうかの判断を支援する確立された仕組みです。

ビッグデータは、例えば事例研究「岩絵記録のブレイクス

ルー：Breaking through Rock Art Recording」[8]によって作られた点群モデルのような、フライスルーや点群モデルなど、従来とは異なるメカニズムを提案しています。これらは一般的にプロジェクトの成果であり、減算（デシメーション：サイズ減少）したデータセットが使用される傾向がありますが、関連する原データの情報も重要になるでしょう。

- [1] <http://archaeologydataservice.ac.uk/research/bigData>
- [2] <http://www.bittorrent.com/>
- [3] <http://www.uklight.ac.uk/>(*)

- [4] <http://www.bgs.ac.uk/services/ngdc/home.html>
- [5] <http://badc.nerc.ac.uk/data/>(*)
- [6] <http://www.neodc.rl.ac.uk/?option=displaypage&Itemid=146&op=page&SubMenu=146>
- [7] <http://www.noc.soton.ac.uk/gg/facilities/seadog.php>(*)
- [8] http://ads.ahds.ac.uk/catalogue/resources.html?btrarb_ahrb_2005

引用・参考文献リスト

Gilheany, S. (nd) The Decline of Magnetic Disk Storage Cost Over the Next 25 Years. White paper. <http://www.archivebuilders.com/whitepapers/>(*)

著作権および知的財産権(IPR)

はじめに

著作権と知的財産権（IPR）は、考古学プロジェクトのライフサイクルを通してデジタルアーカイブを計画する際の重要な検討事項です。デジタルデータの配布と再利用を容易にするためには、データ作成者と利用者の両者に関係する著作権やその他の知的財産権の問題が特に重要となります。作成者と利用者はいずれも、著作権と知的財産権があらゆるオリジナルな成果に自動的に生じる可能性があることを認識しておくべきです。著作権と知的財産権に関する法律は国によって大きく異なっているので、本ガイドは関連する問題の概要を説明することを目的とします。疑問点や個別の問題については、さらなる検討が必要になるでしょう。一般的に、アーカイブに保存する前に、権利が確立されているか、または権利問題が解決されているかを確認することが重要です。

本章では、アメリカとイギリスにおける著作権に関する問題を説明し、続いて国ごとの問題を取り上げます。本章は以下のように構成されています。

- ・ 一般的なトピック
 - ・ 著作権の保有者
 - ・ 推奨
 - ・ 原データ
 - ・ ライセンス対象物
 - ・ メタデータと著作権
 - ・ マルチメディア
 - ・ そのほかの考慮事項
- ・ イギリスとアメリカにおける個別的な項目

誰が著作権を保有しているのか

一般的に、成果の著作権は、正式に他の人や組織に譲渡しない限り、著作物の創作者、または創作者の雇用者が保有します。これは、契約や資金提供によって、資金提供者が成果の著作権を別の当事者のものとする要求を妨げるものではありません。

推奨

考古学者は、調査によって作成されたデータや文書が広くアクセス可能で利用可能にするために、調査を委託するクライアントと交渉すべきです。考古学者は、契約書に著作権の問題が明確に記載されているかを確認し、プロジェクトが完了したときにどのような情報が機密とみなされるかを事前に明確にしておくべきです。情報がアーカイブに寄託される際に、著作権保有者を特定しておくことが重要です。また、アーカイブ内でアクセスを制限しているデータ、文書、またはその一部がある場合は、誰がどのような条件でデジタルアーカイブへのアクセスを許可するのかを明らかにしておくことが重要です。

一般的に、デジタルアーカイブに保存されている考古学情報は可能な限り多くの人々が利用できるようにすることが推奨されます。しかし、極秘情報の場合には、一定期間、情報に一時的な公開禁止措置がとられることがあります。機密情報の場合、公的機関は、遺跡の保護を確保するために、情報の流通を制限することができます。

原データ

原データには、著作権に基づくアクセス・利用に関する特定の制限がある場合があります。制限や著作権は、原データが作成された国によって異なります。各国の規制に関わらず、データがデジタル化されているプロジェクト、およびデジタル化された原データが使われているプロジェクトでは、許可を取得し、資料をプロジェクト内で十分に文書化することが重要です。

ライセンス対象物

背景や文脈上の地理空間調査データなどの特定の資料は、プロジェクトの外部資料から取得されることが多いため、情報のライセンスを保持している第三者が所有している場合があります。多くの場合、外部組織はライセンス所有者の許可なしでライセンス対象のデータをアーカイブすることはできず、ライセンスを受けたデータセットに商業的価値が存在する場合、許可が承諾されることはほとんどありません。このような場合、見込みのある解決策は、「分散型アーカイブ」

を作成することです。分散型アーカイブでは、リポジトリやアーカイブ組織が作成データとともに、第三者の組織が保有する原データを参照するためのメタデータも保持します。特にビッグデータ（例：センサーによるスキャンやその他の複雑なデータソースなど）の場合、このような分散型の仕組みは、大規模データセットの維持とアクセスをアーカイブではなく外部の専門組織が担うことになるという利点もあります。欠点は、完全なデータセットへのアクセスを必要とするユーザーは、第三者から原データを購入しなければならない可能性があることです。

メタデータと著作権

アーカイブ内で確実に権利を守り、維持していくために重要な要素は、文書化です。プロジェクトの文書化とプロジェクトのメタデータに関する前章、および個別の手法に焦点を当てた後の章では、記録と個人または組織の関係を記録することが重要な要素となっています。大枠では、関連するデータ作成者や許可を受けた者の名前とともに、権利、使用、アクセス条件が明確に文書化されていることが不可欠となります。前述の通り、資料がデジタル化されたり、購入されたものである場合には、その情報を、データの利用方法の詳細とともに記録する必要があります。

マルチメディア

上記のガイドラインは、動画と音声ファイルを含んでいないことに注意してください。視聴覚資料やマルチメディア資料には複雑な著作権保護があります。これはとくに、発掘調査や講演の録画を含むような映画に当てはまり、動画、サウンドトラック、脚本などに個別の著作権保護が存在するかもしれません。

その他の考慮事項

著作権に加えて、知的財産の所有者や創作者は、その著作物に関連した著作人格権を有しています。考古学では、この権利は、大規模な発掘調査チームや作業グループの中の個人にも適用されるかもしれません。これらの人々を個別に特定できないかもしれませんが、彼らの貢献を認めることは倫理的実践の一部でもあります。ほとんどの場合、これは著作権を制限するものではありません。

英国の著作権

一般的に、正式に他人や組織に譲渡されていない限り、作品の創作者または創作者の雇用者が著作権を有します。しかし、考古学プロジェクトへの資金源によっては、異なる場合があります。データがデジタル化されているプロジェクトでは、許可が与えられ、資料がプロジェクト内で十分に文書化されていることが重要となります。デジタル化プロジェクトを実施する際に考慮すべき問題は、JISC Digital Mediaの文書「Learning Lessons from Other Digitisation Projects」[1]に記載されています。

データ作成と著作権

イギリスでは、著作権の法的枠組みは1988年著作権・意匠・特許法によって規定されています。この法律では、イギリスの元首または（政府機関の職員を含む）王室の使用人が職務中に作成した著作物に存在する、特別な王室著作権の詳細が規定されています [2]。

欧州委員会にはまた、イギリスの法律に適用された、あるいは適用されようとしているダブリンコアな著作権指示があります。これらには、作成者の死後70年間の著作権保護（指示2006/116/EC [3]）やデータベース著作権保護（EEC 1996 [4]）などがあります。

イギリスにおける著作権に関する最新情報は、以下のJISC法律文書を参照してください。

- ・ Copyright Law Essentials（著作権法の要点） [5]
- ・ 'Intellectual Property Law Essentials'（知的財産権の要点） [6]
- ・ 'e-Repositories and the Law Essentials'（電子リポジトリと法律の要点） [7]

著作権とマルチメディア（オーディオ、ビデオ、など）

JISC Digital Mediaのウェブサイトでは、デジタル映像・音声ファイルに関する問題点について、文書「Copyright and Other Rights for Creating Time-based Media Resources」[8]で詳しく説明されています。

セクション1.2 米国の著作権

アメリカでは、多くの場合、著作権は著作物の創作者、または創作者の雇用主が有しています。しかし、考古学的なデータや文書の中には、著作権によって制限されていないものもあります。これは、考古学データや文書が「パブリックドメイン」にある場合に適用されます。著作物の著作権が消滅または放棄されている、またはデータや文書がアメリカ合衆国政府の著作物の場合、著作物はパブリックドメインとなります（パブリックドメインについての詳細は以下を参照）。これは、業務がアメリカ政府機関との契約によるものであるか、アメリカ政府機関の職員によって実施された場合に適用されます（アメリカの規則ではそのような調査による成果物を「アメリカ合衆国政府の著作物（US Government Work）」と呼びます）。例として、米国森林局や陸軍工兵隊などの連邦政府機関が請け負った考古学調査や、国立公園局の考古学者が実施した調査などが挙げられます。

データ作成と著作権

アメリカの法律では、著作権保護は、あらゆる有形的表現媒体に著作者性が固定されたオリジナルの著作の創作者のため（17 USC. § 102 (a) [9]、または、これらの個人が著作物を創作している間に所属していた組織のために存在します（「職務著作物（Works Made for Hire）」、米国著作権局（2010））。特別なケースでは、著作権で保護された著作物は、著作権法のフェアユースの方針に従って利用することができます。フェアユースについての詳細は、以下の「フェアユース」項を参照するか、アメリカ著作権庁のウェブページ [10]を参照してください。

創作物には文学（ノンフィクション、フィクション、または詩歌）、音楽、演劇、画像、視聴覚、録音、建築物、またはその他の有形の創作表現などがあります。著作権問題を考える際には、科学および技芸を促進するという法の構成上の目的を想起することも重要です。アメリカにおける著作権は、アクセス制限やその他の方法で知識の普及を制限することを意図したものではありません。著作権とは、著作者が著作物に