

発掘調査報告書を各機関内で電子化する場合の方法と注意点

作成: 全国遺跡報告総覧プロジェクト

1. 電子化に使用する機器

(1) 複合機(コピー機)

使用する複合機(コピー機)が、単体で電子化機能をどの程度持っているかの確認が必要となる(Jpeg スキャニング機能だけのものから OCR 付き PDF を作成できるものまであるため)。また、読み取り解像度は、カラー600dpi に対応するものが望ましい(一部機種では、400dpi の制限有り)。スキャニング機能だけの場合には、PC に転送し OCR 付き PDF を作成する。

(2) スキャナ

電子化には PC を使う必要があるが、画質等細かな設定が可能となる。また、作業には A3 対応のオートシートフィーダ付きのものが望ましい。

2. スキャニングの条件設定

解像度は 600dpi とし、ページの情報により 2 値(文字)～カラー(写真)の設定を行う。複合機の場合には、自動で傾き補正やカラー認識を行えるものもあるが薄い色調のものは誤認識が起きやすく事前にテストして最適な設定を行っておく必要がある。

- ・地図…元図の品質が低い場合読み取れないことがある。
- ・トレース図…細かい線や文字が読み取れないものや、色つきのものがある。
- ・写真…色の再現が難しくスキャナの差で大きく異なる。
- ・文字…拡大時に文字のカスレが起きる。

3. スキャニング作業の手順及び注意点

(1) 裁断可能な報告書の場合

- ・表紙と本体を外した後に背糊部分(5mm)を裁断しスキャンを行う。
- ・再製本不要の場合には、表紙を外さずに裁断しても良い。
- ・針で止められている場合には、針を外す。
- ・背糊を十分除去しておかないとオートシートフィーダで引っかかることがある。
- ・表紙・付図は、別にスキャンしたものを該当ページに入れる。

(2) 裁断不可の報告書の場合

- ・1ページずつスキャンする。
- ・のどが不鮮明になったり、画像データが傾いたりしないよう注意する。

(3)OCR 処理

報告書は固有名詞が多いため、OCR 処理の精度は低い。

(4)公開用 PDF の作成

Adobe Acrobat を使ってデータを圧縮する。圧縮しても 100MB 以上となる場合は、100MB 未満になるよう分割する。分割する場合、ファイルサイズは各ページのサイズの総和となるが、各ページのサイズは内容によって大きく異なるためページで単純に分けることができない。文章主体のページはサイズが小さいが、図版等のカラー写真ページではサイズが大きくなるため、分割場所に注意が必要となる。

<Adobe Acrobat を使った PDF ファイルの最適化作業>

「アドバンスド」メニューにある「PDF の最適化」により、画像の種類に応じて次の処理を行う。

・カラー画像

ダウンサンプル:ダウンサンプル(バイキュービック法)150ppi

次の解像度を超える場合:150ppi

圧縮:JPEG

画質:低

・グレースケール画像

ダウンサンプル:ダウンサンプル(バイキュービック法)150ppi

次の解像度を超える場合:150ppi

圧縮:JPEG

画質:低

・白黒画像

ダウンサンプル:ダウンサンプル(バイキュービック法)300ppi

次の解像度を超える場合:300ppi

圧縮:CCITT Group 4