

国立国会図書館未収かつ入手困難資料のデジタル化データに係る収集要件

1 画像データの仕様

- ・ フォーマットは JPEG 2000/TIFF/JPEG/PDF¹を想定し、1冊の単位ではフォーマットを統一する。
- ・ 非圧縮又は可逆圧縮のファイルに加え、それを提供用に圧縮²した非可逆圧縮のファイルがあることが望ましい。
- ・ 解像度は原寸大に対し 300dpi 以上とし、スキャニング後に補間しないことを原則とする。
- ・ 階調は 24 ビットフルカラー又は 8 ビットグレイスケールとする。
- ・ 撮影については見開き/1 頁どちらでも構わないが、1冊の中で統一されていることが望ましい。
- ・ 1冊単位³で画像を 1つのフォルダに格納する。フォルダ名は資料が一意に特定できるものとし、画像のファイル名は資料の頁の並び順に合致するよう連番(0001, 0002...)を付与する。
- ・ 1冊につき、表紙等のサムネイル画像⁴が 1枚あることが望ましい。
- ・ 階調及び実寸の判断のために、スケール類が写し込まれていることが望ましい。

2 メタデータの仕様

- ・ 各資料のメタデータについては、提供機関の書誌データ(タイトル、著者、出版者、出版年等)を基に作成し、キーとなる ID 等を付与の上、画像を格納したフォルダ名と突合が可能なものとする。それに加え、画像作成時における情報(フォーマット、スキャナ、圧縮率、データサイズ等)も付与されていることが望ましい。
- ・ フォーマットはエクセル/CSV/TSV とする。

3 その他

- ・ 国立国会図書館で入手可能性調査を行い、入手困難であることが確認された資料については、図書館及び個人への送信を行う。
- ・ 当該資料のデジタルデポジットシステムへの登録後、OCR によるテキスト化を行い、検索等に用いる。

¹ PDF はデジタルデポジットシステムに格納は可能であるが、送信等への対応ができない。

² 画像のファイルサイズは、表示を考慮して 1ファイルあたり最大 5MB 程度に収まるのが望ましい。

³ 合冊製本されている場合は、合冊前の巻号ごとにフォルダを作成することが望ましいが、個別の相談事項とする。

⁴ フォーマットは JPEG。ピクセル値の大きい方の辺が 256 ピクセルとなるように作製し、ピクセル値の小さい方の辺は、上下又は左右に等幅の余白を付け、256 ピクセルとする。背景はグレイ(濃度 50%、R128 G128 B128、#808080)の単色塗り潰しとする。

- 当該資料に関する問い合わせについては、国立国会図書館では対応せず、提供機関に取次ぐこととする。提供機関が再撮影等の対応を行った場合には、差替え等の対応を行うこととする⁵。
- 提供機関が当該資料のインターネット公開を希望する場合は、提供機関で権利処理等を行い、国立国会図書館で確認することとする。

⁵ 原本を当館に寄贈した場合等の取扱については、別途協議することとする。