

資料デジタル化の技術



国立国会図書館 関西館
電子図書館課

福田 亮

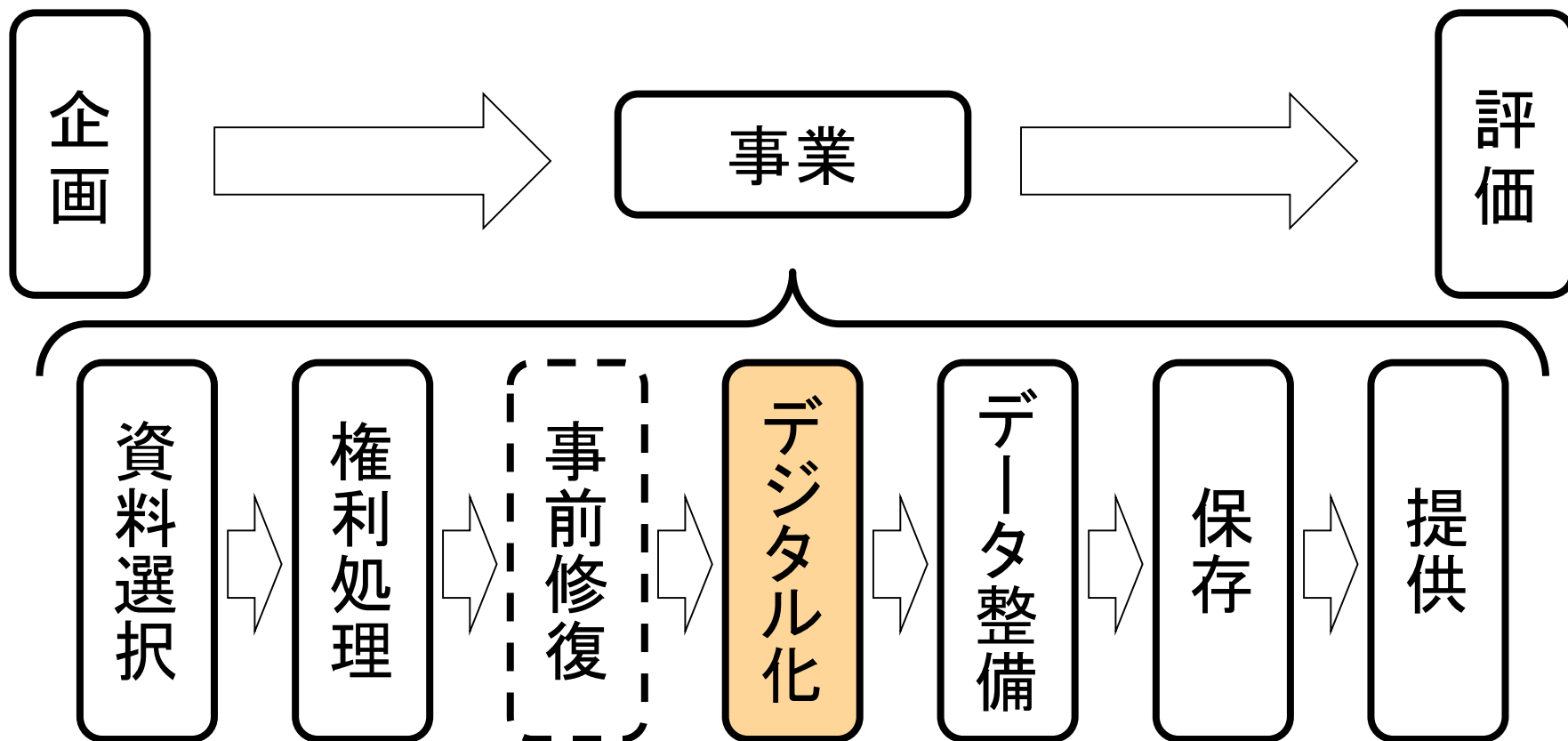
目次

1. デジタル化作業の流れ・仕様
2. デジタル画像の基礎知識
3. 画像作成
4. 画像編集実演

1. デジタル化作業の流れ・仕様

1.デジタル化作業の流れ・仕様

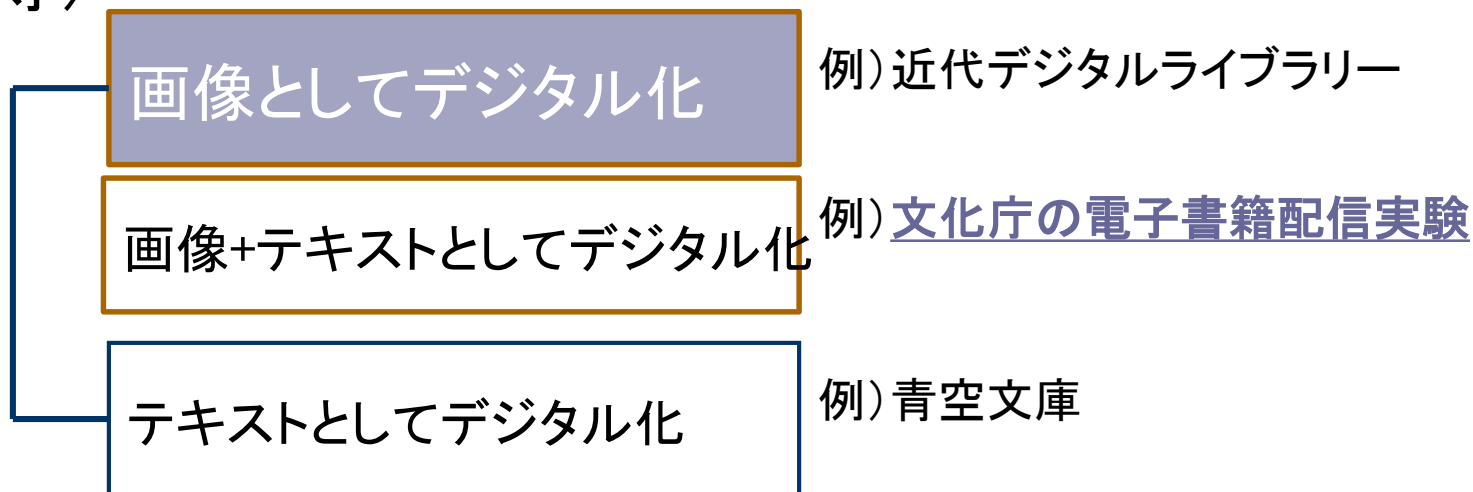
デジタル化の全体フロー



デジタル化の対象と方法

□ 図書館等で行うデジタル化

- 音声、画像、映像（オーラルヒストリー、写真、記録映像など）
- 文字と図画から構成される資料（図書、雑誌、貴重書等）



どういう手順でデジタル化するか

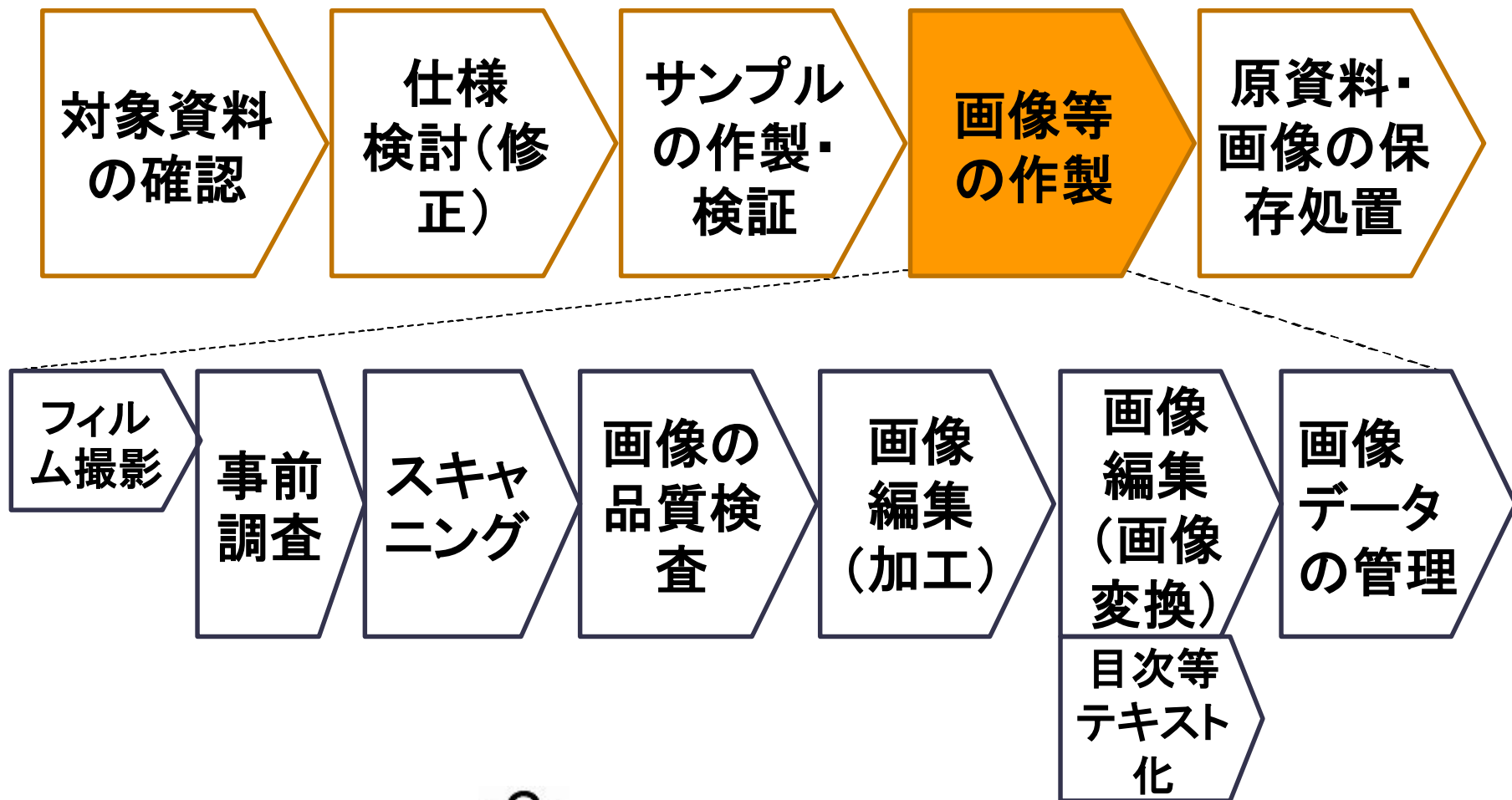
□ 資料を画像としてデジタル化する場合

対象資料のフィルム撮影を行い、
撮影フィルムをスキャンする
ことにより電子化する方法

対象資料を直接スキャンするか、
又はデジタルカメラで撮影する
ことにより電子化する方法

1.デジタル化作業の流れ・仕様

デジタル化作業の詳細フロー



仕様書作成

- ガイドライン類を参照
- 他館の先行事例からヒント
 - 「資料デジタル化の手引き2011版」(配布)
- 仕様チェックシートを作ってみるとよい。考え方の“見える化”（配布：当館サンプル）
- 迷ったら 試してみる
 - 既存のものでまかなえないか？
 - 個人的に「作ってみる」・「使ってみる」



(参考) ガイドライン

□ 知のデジタルアーカイブ～社会の知識インフラの拡充に向けて～ (平成24年3月30日)

- http://www.soumu.go.jp/menu_news/s-news/01ryutsu02_02000041.html
- 提言及び構築のためのガイドラインが掲載
- 先行事例紹介
- 参考情報多数
- デジタル化に主眼

2. デジタル画像の基礎知識

2. デジタル画像の基礎知識

2つのデジタル画像形式①

□ ペイント形式（ビットマップ、ラスター画像）

■ 点の集まりとして記録

□ 身近な使用例：デジタル写真

□ ベクタ形式

■ 図形（点、線、面の情報）として記録

□ 身近な使用例：

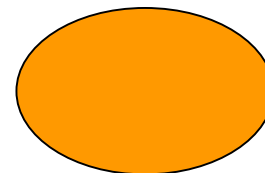
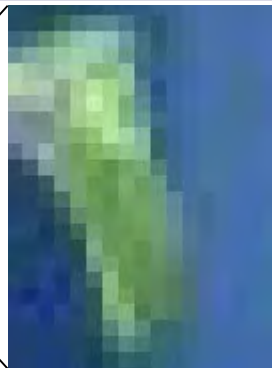
CG、MS officeの図形、文字フォント

2. デジタル画像の基礎知識

2つのデジタル画像形式②



ペイント形式



ベクタ形式



2. デジタル画像の基礎知識

画素数 — 「点」(要素)の数

- ○○× ■■ピクセル(pixel)
- △△△万画素
- 画像の最小単位＝色のついた「点」
- ドット(dot)・・・プリンタやディスプレイで画像を再現するための最小単位
 - 写真のピクセル(pixel)に相当



2. デジタル画像の基礎知識

解像度 — 表現能力の尺度 ①

□ ディスプレイ、スキャナ、プリンタ、デジタルカメラ

□ 絶対解像度＝画像のサイズ

■ 縦横のピクセル数[1280×1000=128万画素]

□ 相対解像度＝画像の密度

dpi(dots per inch) *1inch=2.54cm

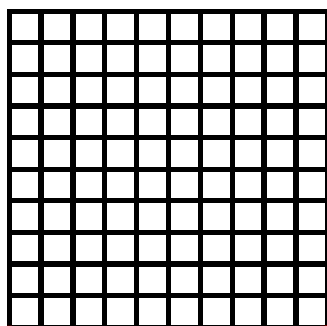
350dpi → “2.54cmの中に350の点(ドット)”



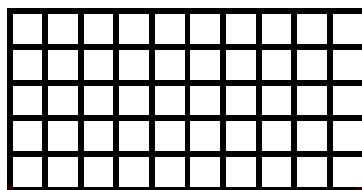
2. デジタル画像の基礎知識

解像度 — 表現能力の尺度 ②

＜絶対解像度＞

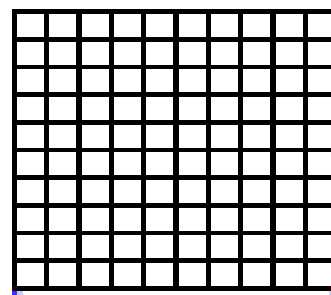


$10 \times 10 =$
100ピクセル

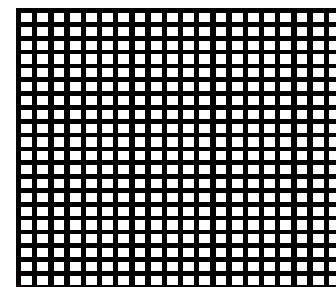


$5 \times 10 = 50$
ピクセル

＜相対解像度＞



10dpi



20dpi



2. デジタル画像の基礎知識

階調 — 色の濃淡をあらわす段階

□ 白黒

- ① 2値(モノクロ2階調)・・・白か黒 1ビット
- ② グレyscaleール・・・白と黒の階調で表現
256階調(8ビット)が一般的

□ カラー

- ③ 8ビット 256色 (インデックスカラー)
- ④ 16ビット 65536色 (ハイカラー)
- ⑤ 24ビット 約1678万色 (フルカラー)

□ X ビット = 2^X 通りを表現 8ビット = $2^8 = 256$



2. デジタル画像の基礎知識

階調（白黒 ～ カラー）



2値（モノクロ2階調）



グレイスケール



カラー（24ビット）

2. デジタル画像の基礎知識

(参考) 色彩の予備知識(1)

□ 色の3属性

- 色相 (Hue: H) 赤青など基準となる色味
- 明度 (Value: V) 色の明るさ
- 彩度 (Chroma: C) 色の鮮やかさ(純色⇔無彩色)

□ 色を表現する方法

- 減法混色(混ぜると暗く)
 - 絵画など物の色づけ、印刷
- 加法混色(混ぜると明るく)
 - 電子的なデジタル画像表示(ディスプレイ等)

2. デジタル画像の基礎知識

(参考) 色彩の予備知識(2)

□ カラースペース

- 機器や印刷で再現できる色の範囲
- CMYK、RGBなど複数の基本となる表現方式があり、さらにそのバリエーションがある。(再現できる色の範囲はそれぞれ異なる)
 - (例) sRGB (standard RGB) と Adobe RGB
 - ICCプロファイル(機器間での色表示を合わせ)
- 各色は、それぞれのカラースペース内で数値的に表現される。
 - (例) “白”のカラーコード #FFFFFF
 - CMYK | C0 M0 Y0 K0 RGB | R255 G255 B255

2. デジタル画像の基礎知識

圧縮 — ファイルサイズを縮小①

- 可逆圧縮(ロスレス(Loss less))
 - データを省かない＝元に戻せる
- 非可逆圧縮(ロツシー(Lossy))
 - (比較的不要な)データを省く＝元に戻せない
- 圧縮率
 - 非可逆圧縮 > 可逆圧縮
- 圧縮率が高いと品質劣化

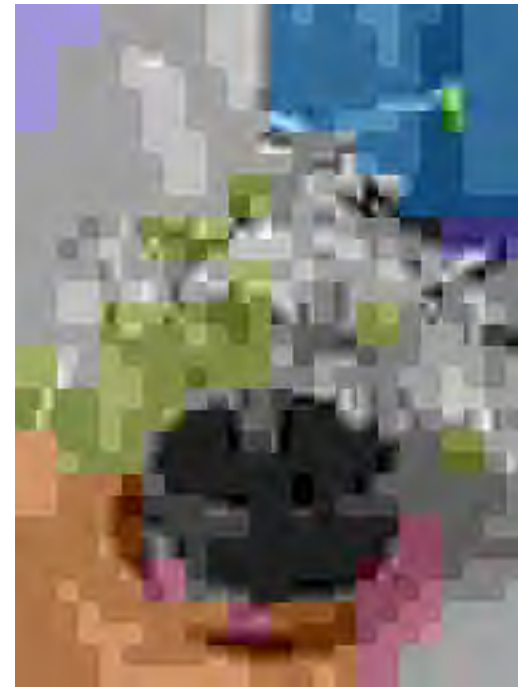


2. デジタル画像の基礎知識

圧縮 — ファイルサイズを縮小②



17.3KB



7.8 KB

2. デジタル画像の基礎知識

画像フォーマットの概要

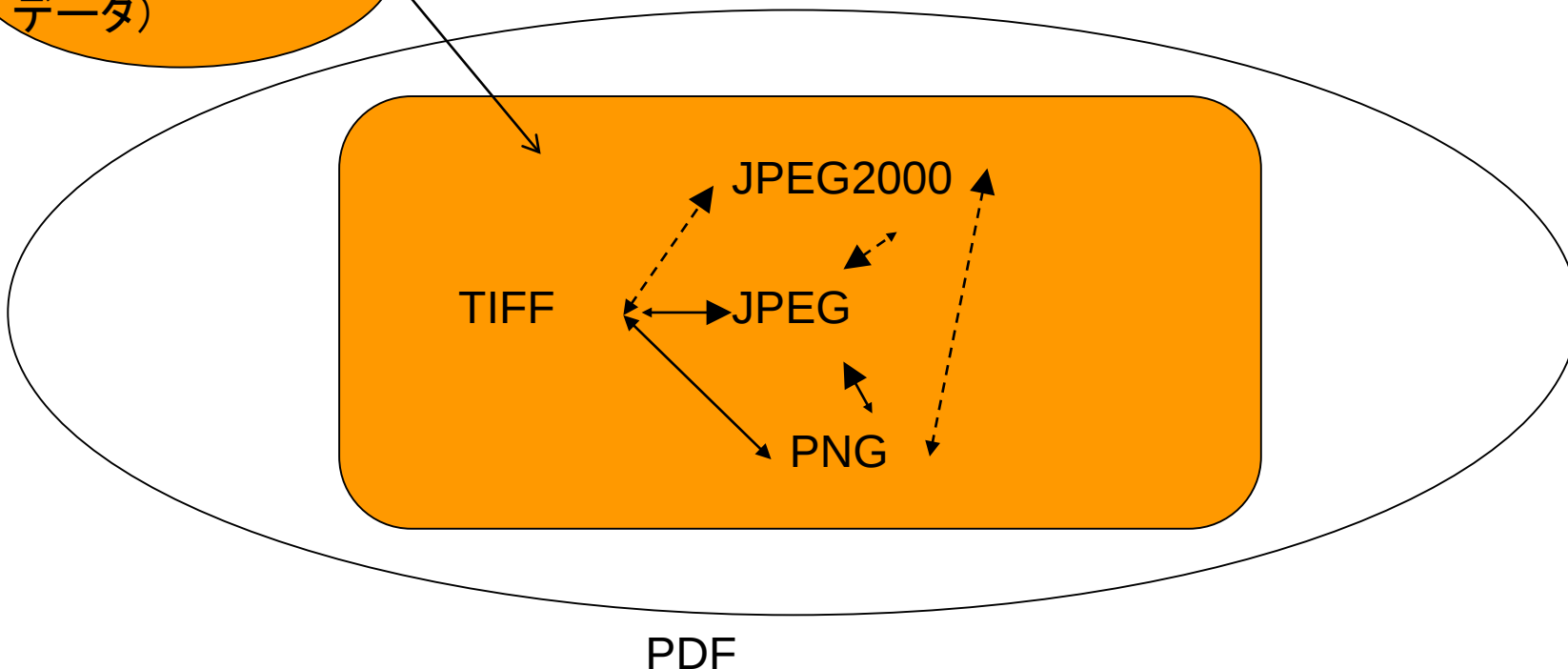
フォーマット	拡張子	圧縮	用途	備考
TIFF	.tif(f)	(非圧縮)	保存用	スキャニング時の保存用画像として利用(基本的に非圧縮で扱われる)
GIF	.gif	可逆	提供用	主にウェブサイトのアイコン等の画像として利用(色表現は8bit)
PNG	.png			GIFに替わるウェブ用画像として利用(色表現は48bitまで対応)
JPEG	jp(e)g	非可逆		圧縮率に優れ、ウェブサイトの写真画像等に標準的に利用
JPEG 2000	jp2 など	可逆	保存用	JPEGよりも圧縮率が高く、高機能だが、ウェブブラウザではまだ標準対応していない
		非可逆	提供用	



2. デジタル画像の基礎知識

デジタル化で扱う主な画像形式

RAW画像
("現像"前の画像
データ)



2. デジタル画像の基礎知識

(参考)フォーマットを考える上で

- 汎用性(普及具合と“マルチであること”)
- 複数種類の情報を格納することができる「コンテナフォーマット」(「器」のイメージ)
 - テキスト、画像、音声、動画
 - (例) MPEG 動画や音声
 - PDF、EPUBなどもそうした「器」としての特性を持つ。(電子ジャーナルや電子書籍などで使われている理由の一つ)

3. 画像作成

3. 画像作成

スキャニング機器の種類

	①資料	②解像度	③速度	④範囲
シートフィーダ型スキャナ	・資料の解体(端裁断など)が前提、リーフ資料	～600dpi	A4片面 300dpiカラー ～100枚/分	～A0 (A4が主流) ・両面読み可能
フラットベッドスキャナ	・接写面に資料を押しつけて使用	～9600dpi (超高解像度読み取り範囲は制約あり)	A4片面 300dpiカラー 6秒/枚～	～A3 ・35mmフィルム等読み取り可あり
オーバーヘッドスキャナ	・資料を上向きに置いて使用。非接触式	～800dpi (撮影サイズによる)	A2片面300dpiカラー 4秒/枚～ (機種によるばらつき)	～A1 (A3が主流) (機種によるばらつき)
(文化財用スキャナ)	・資料に応じて、スキャナの設置の仕方が変わる。非接触式	1200dpiなど	襖絵1枚にかかる撮像時間は10分程度 (600dpi時)	襖など大型の対象物も
デジタルカメラ	・資料に応じて。非接触式	被写体のサイズとカメラの最大画素数からの計算による	1ショット1秒以下	画素数・解像度による。 例:1億画素程度の場合 → 400dpi相当: A2強、B2弱程度

※ ウェブ上での簡易調査による

3. 画像作成

スキャニング解像度の考え方①

□ 文字の判読性と解像度（白黒）

- | | |
|-----------------|---------|
| ■ 10ポイント以上 | 150dpi～ |
| ■ 8ポイント以上 | 200dpi～ |
| ■ 4ポイント以下又は線画中心 | 400dpi～ |

□ （カラーの場合は、若干下げても判読可）



3. 画像作成

スキャニング解像度の考え方②

□ ディスプレイ表示と解像度

- 画面解像度 800 × 600、1024 × 768 etc.

- スキャン解像度が高い

- 全体表示は縮小表示で
- 部分拡大して詳細を見ることも可能



- スキャン解像度が低い

- 上記の裏返し

- 写真 < 文字・図表

- 保存用と提供用(電子的展示用)を分ける考え方

スキャニング解像度の考え方③

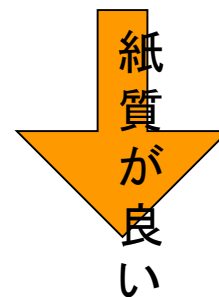
□ デジタル画像印刷と解像度

■ 紙媒体の印刷ライン数例

- 新聞など：60～80線
- 書籍・雑誌など：100～150線
- 美術書・カタログなど：150～200線

■ → 線数の2倍程度で計算

- 通常印刷 200～300dpi
- 高品質印刷 300～400dpi



4. 画像編集実演

画像編集

- 限定的な加工に留める
- フォトレタッチソフトを使用
 - 余白、不要部分の削除(トリミング)
 - 向きの変更(回転)
 - 資料が正面を向くような90度回転や傾き補正
 - ファイルフォーマット変換
 - サイズの変更(リサイズ)
 - 保存用を元に提供用の小さい画像の作成 等

4. 画像編集実演

実演① 解像度を変えてみる(リサイズ)

- サンプル 「光学解像度」400dpi
 - ↓ 「原資料に対して」200dpi相当に変更
- サンプル_200
 - ↓ 同400dpi相当に変更
- サンプル_200_400 (補間による)ぼやけ
- ソフトウェアで解像度増やす 「補間解像度」
- スキャンングの際に、「光学解像度で」「原資料に対して」という縛りが必要なわけ



4. 画像編集実演

実演②フォーマットを変えてみる

- サンプル_200 TIFF
 - ↓ JPEG・JPEG2000に変更して保存
- サンプル_200_J JPEG
- サンプル_200_J2 JPEG2000
 - ↓ PDFにバインドして保存
- サンプル_200 _J-J2 PDF
- ※デフォルト設定では、自動的にファイルが圧縮されて小さくなる場合があることに注意

4. 画像編集実演

画像の情報量

□ 画像ファイル容量

＝[ファイル情報の容量]＋[画像の容量]

□ [画像の容量]＝画素数×階調のビット数

■ サンプル画像(24ビットフルカラー)

□ $(944 \times 800) \times 24\text{ビット} = 18,124,800\text{ビット}$

□ $= 2,265,600\text{バイト} \div 2.2\text{MB}$

■ A4(21×29.7cm)を400dpi $\div 14.8\text{MB}$

□ フロッピーディスク 15枚分

□ 8ビット＝1バイト(B) 1,024バイト＝1KB



4. 画像編集実演

実演③圧縮率を変えてみる

- サンプル TIFF
 - ↓画質レベルを、「50」、「20」のJPEGに
- サンプル_J_L50 JPEG
- サンプル_J_L20 JPEG
- 元の画像容量からの圧縮率で考える
 - 上記の例 L50:1/20程度 L20:1/40程度
- ※圧縮・フォーマット・解像度変更は繰り返さない。必ず保存用画像(マスター)から

□ ご清聴ありがとうございました

■ 公開講座「資料デジタル化の基礎」も参照下さい

□ <https://ndl.secure.force.com/guidance?id=a091000000cZoXgAAK>