

# 電子化作業の概要

奈良先端科学技術大学院大学 学術情報課

雑誌、図書及び論文の電子化作業の概要について以下の要領に従って行っています。  
所蔵するすべての資料を電子化して利用者に提供したいのですが、それぞれの資料には著作権があって、勝手に電子化することは許されていません。

ひとつひとつの資料に対して著作権者の許諾を得たうえでの作業となっています。  
一般的な学術雑誌や図書の電子化のほかに、学内からの情報発信として学位論文やテクニカルレポートの電子化を行っております。

冊子体からの電子化の作業を中心に説明しましょう。

まず、受け入れられた雑誌や図書には、物理的な媒体を認識するための「バーコード」が貼られています。

これを手掛りに作業が展開されていきます。

最初に書誌情報の入力を行います。タイトル、著者、出版者等です。リコーの図書館システム「Limedio」を使っています。

次に現物の背を裁断します。裏写りがひどい場合はコピーする場合があります。

次はモノクロのスキャナで400dpi 2値で頁イメージの読み込み(TIFF)をして同時にOCR変換を行います。

濃淡がある頁は200dpi 256階調で読み込みをして(PNG) 2値のものと置き換えます。

ここで、直接利用者に見てもらう画像データと全文検索に用いる文字データのファイルを作ります。

次は頁の割付です。スキャナで読み込んだ頁イメージに冊子にふられてる実頁を付けていきます。

ここで頁順に配列もしていきます。

さらに、カラーの頁は200dpi 256階調カラースキャナで読み込み(PNG)、モノクロの頁と置き換えていきます。

次にマスク処理と名付けていますが、広告等の第三者著作物の部分があればその部分を削除します。

次は確認作業です。まず、不要な頁を削除すること、そして頁位置と文字情報が正確に対応しているかをざっとみていきます。

次に目次の入力を行います。雑誌ですと論文名、著者、頁、図書ですと章のタイトル、著者、頁等です。

最後に本構成をかけ、目次の頁情報を基に論文単位のPDF化を自動的に行い、夜間処理によって、データベースの更新を行いOPACから検索、利用できるようになります。

作成したデータの確認をWEBのブラウザで行います。

著作権者の方で保存用として電子化データを要望される場合は、CD-Rに焼き付けてお渡しします。

電子化終了した全ての資料をDVD保全の対象としています。

ファイル及びCD-ROMからの電子化については、フォーマット変換でPDF化及び全文データ化を行います。

ビデオ等の映像情報の電子化作業の概要について以下の要領に従って行っています。

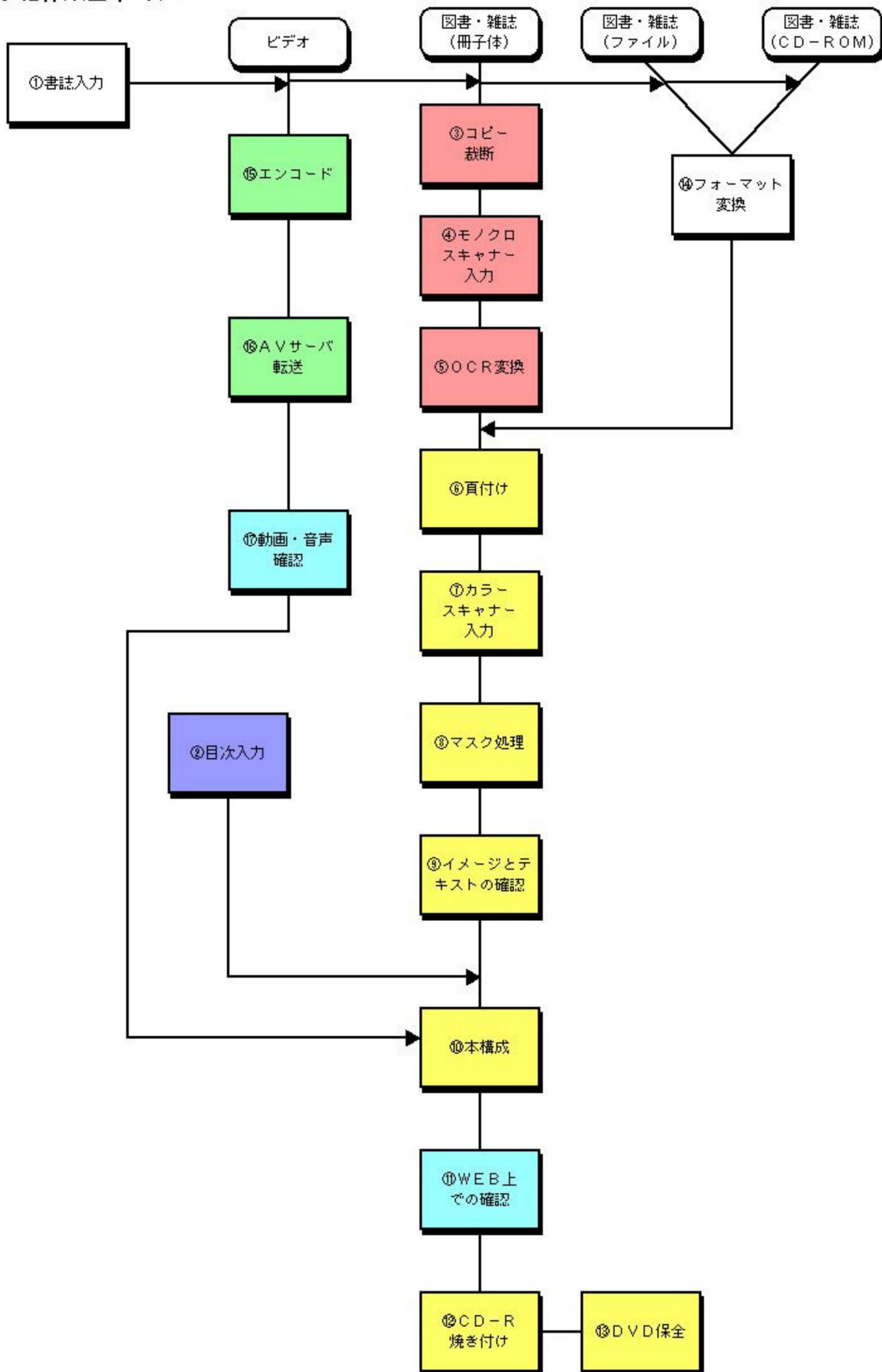
最初に書誌情報の入力を行います。タイトル、著者、出版者等です。

次に映像情報を再生しながらMPEG-2に変換(エンコード)し、ローカルディスクに蓄積します。

ローカルディスクのMPEG-2ファイルを選択し、AVサーバに転送します。

動画・音声の確認を行います。

# 電子化作業基本的フロー



## 電子化作業手順（冊子体）

電子化許諾を得た雑誌、図書、論文等については、1つの資料に対して1つの電子化整理番号がつけられます。電子化作業はその番号により管理されます。

### < 目次入力 > 目次入力

電子化資料をより参照しやすくするために、本文の画像とは別に目次を文字（テキスト）データで作成します。

これにより、目次から必要とする記事に直接リンクすることができます。

例に従って、標題、著者名、頁範囲等の目次情報を入力していきます。

#### [ 注意事項 ]

- ・著者名がない場合は省いてもよい
- ・頁範囲は半角数字、ハイフンで範囲を示す。 (例) 頁範囲：1-5
- ・1頁だけの時は数字だけでよい。 (例) 頁範囲：7

-----例-----

**標題**：Construction of an Autonomously Replicating Plasmid in n-Ankane-Assimilating Yeast, *Candida tropicalis*.

**著者名**：Akihiro Hara, Mitsuyoshi Ueda, Naoki Kanayama and Atsuo Tanaka

**頁範囲**：717-720

**標題**：Nucleotide Sequences and Characterization of Genes Encoding Naphthalene Upper Pathway of *Pseudomonas aeruginosa* PaK1 and *Pseudomonas putida* OUS82.

**著者名**：Noboru Takizawa, Toshiya Iida, Yue-Wu Wang, Masao Fukuda and Hohzoh Kiyohara

**頁範囲**：721-731

#### [ 機能解説 ]

目次テキスト取込 別アプリケーションで作成したテキストファイルを取込むことができます。

頁ノンブルチェック 目次で入力した頁範囲と、頁付け（作業）で割り付けられた数字が対応しているかチェックできます。

頁リスト 入力された目次情報が一覧表示されます。

WEB WEBで公開された目次情報をカット＆ペーストできます。

OCR 読み込んだOCRテキストを呼び出してカット＆ペーストできます。

## < モノクロスキャナ入力 > 入力作業 スキャナ読み込み/OCR 変換

本文部分の土台となる画像（イメージ）データを作成します。

電子化整理番号、ジョブ番号を入力し、書誌を確かめたら作業を開始します。電子化整理番号はバーコードの7桁の番号です。ジョブ番号は通常は01としますが、01、02...と番号をふることにより、複数のジョブに分けて作業することができます。

\* スキャナ連続読み込み/OCR 連動 ではスキャナ読み込みとOCR変換（手順）を一括で処理することができます

[ 機能解説 ]

スキャナ設定 読み取り時のスキャナ設定をします

スキャン方式（ADF、ADF両面、原稿台）

色（白黒、グレースケール）

読み取りサイズ（A3、A4、A5、B4、B5）

解像度（100? 600dpi）

\* 全文を白黒で読み込み、濃淡のある頁は後の段階で（作業）グレースケールに置き換えます。

イメージ削除 イメージとテキストを削除

テキスト削除 テキストのみ削除

## < OCR 変換 > 入力作業 OCR 連続変換

本文のテキストデータを作成します。

和文か英文か選んで処理します。

[ 機能解説 ]

OCR 設定 OCR の設定をします

日本語、英語、日英自動

## < 頁付け > 頁情報入力/修正

作業でスキャンした一枚一枚のイメージに実頁としての数字を割り付けます。

目次情報をもとにTIFF及びPNGのデータをPDFにまとめるため、さらには目次から各論文へリンクするため不可欠な情報です。

原則として頁を割り付ける際には、大きく4種類（表紙、目次、本文、頁のついていない部分）に分けます。

表紙と目次には実頁をつけません。表紙には H1、目次には M1 とそれぞれアルファベットをつけた数字を順次割り付けます。

表紙、目次以外の頁については、実頁で割り付けます。

最後に、数字のついていない頁に先頭から順に E をつけた数字を割り付けます。

-----例-----

H1 (割り付けた頁)	1 (イメージの枚数)
E1	2
M1	3
M2	4
E2	5
3	6
4	7
5	8
6	9
E3	64
E4	65

-----

この時、不要な頁を見つけたら、削除の対象にします。

例外として、ローマ数字 (i, ii, iii...)、アルファベット付き数字 (S1、L1 等) などもありますが、目次入力とそろえて割り付けすればリンクすることが可能です。

[ 機能解説 ]

**並び替え** **自動並び替え** 読込む段階で作業の都合上、頁が前後している場合などにイメージの順序を操作することができます。

**頁番号** **割付** 一枚ずつの割付

**開始頁番号**? **頁数** **割付開始** 複数枚の一括割付

**前方固定部** アルファベット等を割り付ける時に便利です。

## < カラーキャナ入力 > **カラー**

カラー頁を作成します。すでにモノクロでスキャンした画像に重ねてスキャンします。

原則として、全資料の表紙、目次（多色刷りのもの）はカラー頁となります。その他、  
学術雑誌の多色刷りの部分もすべてカラー処理します。

## < マスク処理 > マスク

広告等の第三者著作物の部分を削除する作業です。該当頁をあらかじめチェックしておきます。

該当頁でマスクをおすと Photoshop が立ち上がり、画像処理できる状態になります。

広告部分を範囲指定、カット 保存 閉じる

## < イメージとテキストの確認 > 電子化ファイルの確認 / 修正

作業の仕上げの段階です。

- ・各頁にテキストが作成されているかを確認します。
- ・削除する頁があればこの段階で削除します。
- ・濃淡のある頁はグレースケールで置き換えます。

[ 機能解説 ]

スキャナ設定 読み取り時のスキャナ設定をします

スキャン方式（ADF、ADF 両面、原稿台）

色（白黒、グレースケール）

読み取りサイズ（A3、A4、A5、B4、B5）

解像度（100? 600dpi）

イメージ削除 イメージとテキストを削除

テキスト削除 テキストのみ削除

## < 本構成 > 本構成処理

目次情報と頁イメージ情報とのリンク付けを行う作業です。

本構成を行うと、データは夜間処理にかけられます。

## < Web 上での確認 >

利用者が実際に使う場合に不備がないかどうかチェックするために、夜間処理が終了したデータの確認を Web のブラウザで行います。

・該当する書誌を検索し、「各巻号情報」へとたどります。  
(雑誌は、電子化雑誌一覧を利用してもかまいません。)

・電子化マーク及び TOC マークの確認

電子化マークが出ているか、それが学内限定もしくはフリーになっているかを確認します。「TOC」をクリックし、目次情報の確認(最後まで登録されているか、文字化け、ページ等の入力ミスはないか)をします。

・任意の目次情報から画像データを開きます。頁付けとのリンクがなされていること、画像データと目次情報が同一であること、表紙頁がカラーであることを確認します。

・必要に応じて、マスク処理画面、カラー画像をチェックします。

すべての項目が確認された資料は、**電子化終了**を行います。

## < DVD 保全 > **保全システム**

**自動保全** 作業 で電子化終了された資料は自動的に保全対象となります。

**手動保全** 電子化整理番号を手入力した資料を対象に保全できます。

**復元** 保全したデータを復元します。

## 電子化作業手順(ファイル、CD-ROM)

CD-ROM 等から、PDF ファイルの電子化を行います。1つの資料に対して1つの電子化整理番号がつけられ、電子化作業はその番号により管理されます。

## < フォーマット変換 > **入力作業** **ファイル入力**または**CD-ROM 入力**

電子化する PDF ファイルをそのまま利用します。

PS ファイルに関しては PDF に変換されます。

この段階で、PDF ファイルからテキストを抽出します。

## < 頁付け > 頁情報入力 / 修正

目次からはファイル単位でリンクするので、目次で入力した頁情報と同じようにファイルの先頭頁及び最終頁を入力します。

-----例-----

1-15 ( 割り付けた頁 )                      1 ( イメージの枚数 )

16-21                                              2

・ 3

|                                                      |

-----

## 電子化作業手順 ( ビデオ )

DV Cam、VHS テープ等から映像の電子化を行います。1つの資料に対して1つの電子化整理番号がつけられ、電子化作業はその番号により管理されます。

## < エンコード >

映像情報をデジタル化する作業です。DV Cam等のデータを MPEG-2 に変換します。1つのタイトルごとにエンコード作業を行います。

DV Cam からの電子化では、

ContLIB ( Controlled Device Select )

MPEG Encoder Device DV Cam

VTR Control

Encoder      RM-450R

EDL Editor

DVC

File Properties にファイル名 ( V00\*\*\*\*.mpg ) を設定。

VTR Control

再生、停止、巻き戻し、早送り等のボタンで、

開始コマ ( Mark in ) 終了コマ ( Mark out ) を決める。

これで自動的に時間(Duration)が決まる。

**完了**

**DVC**を反転させながら **REC**を押す。

これで自動的に頭出しが行われ、エンコードが始まります。

## < AV サーバ転送 >

ローカルディスクから AV サーバへエンコードしたデータを転送します。

**AV 入力メニュー** **AV サーバ登録**

ローカルディスクにある mpg ファイルを反転させ、

**登録**

## < 動画・音声確認 >

データがサーバへ正常に移行しているかビデオ閲覧用端末で確認します。

## Web 上での確認

電子図書館ホームページ(<http://dlw3.aist-nara.ac.jp/index-j.html>)

### 課題

#### 1. 図書

「情報管理入門：情報技術による知的生産性向上」

佐藤修、藤森洋志、阿部昭博共著. 弘学出版, 1997

#### 2. 雑誌

「バイオサイエンスとインダストリー」Vol.58 No.5

バイオインダストリー協会/Bioindustry Development Center (BIDEC). バイオインダストリー協会.

3 . 雑誌

「Cell structure and function」 Vol.25 No.3

Japan Society for Cell Biology

4 . 論文

「2値画像回転のための2パス法と3パス法での誤差の検討」

東海林健二 平野邦彦 宮道壽一. [電子情報通信学会論文誌. D-II. 情報・システム. II-情報処理](#). 電子情報通信学会. Vol.82 No.5, 1999年5月, p973-977.

5 . 論文

「自動車を取り巻く環境と材料」

加茂 尚. [高分子](#). 高分子学会. 高分子学会. Vol.49 No.5, 2000年5月, p298-301

6 . ビデオ

「IBM：革新的な半導体技術を開発」

日本工業技術振興協会, 1997(High-Tech Shower International:97111704)

7 . ビデオ

「Networked Digital Library of Theses and Dissertations」

Edward A. Fox. 奈良先端科学技術大学院大学, 1999(デジタル図書館ワークショップ:第15回(平成11年7月19日))