

歴代宝案訳注本第4・5・6・12・14冊  
デジタル化・テキスト化業務委託仕様書

- 1 委託業務名：歴代宝案訳注本第4・5・6・12・14冊デジタル化・テキスト化業務委託
- 2 委託業務内容  
県教育庁文化財課史料編集班（以下、委託者と記す）から借用した歴代宝案訳注本第4・5・6・12・14、計5冊（被写資料）を、以下の要領で画像データ化（デジタル化）し、委託者が提供したテキストデータをもとに整形し、テキストデータおよび検索可能なPDFを作成し、指定の成果物を納入すること
- 3 納入物件（各冊毎）
  - ・ 保存用画像データ\_被写資料ページ順（JPEG形式一式） DVD-R一式
  - ・ 検索性PDFデータ CD-R一式（または上記DVDに含む）
  - ・ 頁順テキストデータ CD-R一式（または上記DVDに含む）
- 4 納入期限 令和2年2月28日
- 5 デジタル化・PDF作成の内容
  - (1)被写資料
    - ・ 形態：書籍
    - ・ 撮影コマ数：第4冊525頁、第5冊516頁、第6冊540頁、第12冊492頁、第14冊573頁、計5冊2,646頁/コマ（合紙は除く）
    - ・ 原本のサイズ：B5版
  - (2) 提供した訳注本第4・5・6・12・14、計5冊をもとに頁毎の画像データ（JPEG画像）を作成する。なお本の断裁は可能とする。
  - (3) 画像データ（JPEG画像）からPDFを作成し、透明テキストを貼り付け、検索可能なデータを作成する。
    - ①頁毎のPDFデータを合体して1冊とすること。
    - ②OCRによる透明テキストデータを作成し、目視にて変換後の校正作業を行う。
    - ③頁をまたいだ文字検索を可能にすること。
    - ④文字コードは Unicode (UTF-8) とする。外字（≠文字）処理については委託者と調整する。
    - ⑤OCRソフトによる読み込みの誤変換を防ぎ、品質を担保するため、徹底して内校を行うこと。\*委託者提供のテキストデータおよび『歴代宝案訳注本』を参考に文字修正を行うこと。
  - (4) テキストデータ（巻毎）を文書毎に整形する。
- 6 デジタル化の条件  
自社内でデジタル化が可能であり、なおかつWebサイトでの掲載を前提とする

データの作成およびファイルの整理ができること。デジタル化に伴う作業を委託者以外の業者へ外注しないこと。

被写資料の取り扱いについて

デジタル化のため借用した被写資料は、破損、汚れ等のないよう取り扱いに万全を期すこと。また、被写資料は、受託業者内の撮影場所および保管場所外への持ち出しは禁止すること。

## 7 画像データ（JPEG 画像）の仕様

### ①撮影

8bit グレースケール。

- a. 原則として1頁を1コマに撮影すること。
- b. 画像データの縦横比は、被写資料と同一（B5版）とし、被写資料1頁全体を画像データの全面に配置する。また両者の水平線は一致させること。
- c. 全てのコマについて、原則として同一倍率、同一サイズで撮影すること
- d. 裏写りが確認される被写資料は、原則として合紙を入れ裏写りのないように撮影すること
- e. 画像データの「解像度」と「明暗調」は、これを被写資料の原寸大で印刷した場合において、被写資料と遜色のない「被写資料の複製」を得られる品質であること（原則として解像度はB5サイズで400dpi以上を目安とする）

・撮影画像はJPEG形式で保存し、PDFデータはJPEG形式データをもとに作成する。

・撮影方法に関して上記に因り難しい場合、または次の場合は委託者に申し出てその指示を受けること

- a. 撮影により、被写資料が破損するおそれがあるとき
- b. その他、撮影上の疑義が生じたとき

### ②格納形式

（保存用画像デジタルデータ）

- ・格納媒体：DVD
- ・DVDの収納：プラスチックケース（背面に印字できる厚さ（10mm厚）のもの）
- ・DVD盤面印刷：「DVDタイトル」「資料名」「作成年月日」を印字する。
- ・DVDプラケース内インデックスの背面に「DVDタイトル」を印字する。

8 8月中に中間報告を行う。

## 9 権利及び秘密保持

成果物の著作権は沖縄県教育委員会に帰属する。また、本業務でデジタル化する被写資料の内容及び本業務で知り得た情報、デジタルデータなどは、受託者以外の第三者に漏らしてはならない。

## 10 受託者の検査

(1) 画像データ

全画像について下記の検査を行い、不適箇所は修正する。

- ・画像データのファイル数
- ・画像データの「解像度」「明暗調」
- ・画像の向き、傾き

(2)PDF に透明テキストを貼り付けたデータ

- ・文字検索が可能かどうか検査を行う。

(3) 頁毎に整形されたテキストデータ

- ・ 頁抜けがないか、頁順になっているか検査を行う。

#### 11 納品検査

納品された成果物は委託者が検査を行い、その結果、不具合があるときは、速やかに補正して改めて検査を受けるものとする。

#### 12 その他

この仕様書について疑義が生じた場合は、受託者は速やかに委託者と協議するものとする。

## \*テキストデータ整形の留意点

・委託者提供のテキストデータを使用し、底本である『歴代宝案 訳注本』をみながら整形すること。なおテキストデータを抜き出した際、外字（作字）は削除されるため、特に留意すること。 【別紙1参照】

(1) 体裁は原則として底本どおりとする。

①テキストは横書きとする。

②本文中のルビ、注番号、空白行はテキスト化の対象としない。

③文字コードは Unicode (UTF-8) とする。

④旧仮名遣いやカナなどを含め、原則として底本の通りとする。ただし、漢字の異体字や外字等で底本通り入力できない場合、テキストで表示されない場合は、委託者と調整する（別途指示する）。

中間報告時、あるいは随時、表記できない文字の一覧を提出し、当方で代替文字または文字コードを指定する。 【別紙2参照】

⑤段落改行は底本通りとする。

⑥底本の1字アキは原則テキスト化の対象外だが、物品名・人名が列挙される場合は1字アキとする。但し姓名の間の1字アキは削除する。

⑦割書は /（全角）を入れる。

⑧文書番号、年代などの一、一、一、ハイフン - は、すべてダッシュ「一」（半角）とする。

(2) 巻毎のテキストデータを文書毎にする。

(3) テキストファイルの番号の付け方は以下の通りとする。

例：挨拶など

0-01 中扉：あいさつ

0-02 凡例

0-03 参考文献・口絵キャプション

0-04 琉球国王・中国皇帝（清代）一覧表

0-05 目次

本文（文書毎とする）

2-105-01

2-105-02

～（以下、本文は文書毎）

巻末

9-01 語注索引

9-02[1] 語句索引

9-02[2] 人名・地名索引

9-03 奥付