

# 1. 提供方法

「整理標準化データ」の提供は、DVD-R を媒体とし、下記要綱で提供する。

## 1. 1 媒体仕様

### (1) 物理フォーマット

物理フォーマットは、2000 年 5 月に DVD フォーラムで認められた追記型 (Recordable) DVD の規格である「DVD-R for General ver. 2.0」に準拠する。

### (2) 論理フォーマット

論理フォーマットは、「TR X 0006-1998 DVD – 再生専用ディスクのポリシー構造及びファイル構造」に準拠する。

また、ブリッジフォーマット (JIS X 0609 及び JIS X 0606 (ISO9660)) を前提とする。

## 1. 2 提供形態

整理標準化データの提供形態は、大別して、「SGML データ」と「イメージデータ」の 2 種類が存在する。

### (1) SGML データ

SGML \*<sup>1</sup> データは、SDIF \*<sup>2</sup> に従う様式であり、利用にあたっては、外部実体定義ファイル \*<sup>3</sup> が必要である。

**\* 1 SGML (Standard Generalized Markup Language)**

ISO 8879, JIS X 4151 で規格化されている標準汎用マークアップ言語

**\* 2 SDIF (SGML Document Interchange Format)**

ISO 9069, JIS X 4171 で規格化されている SGML 文書交換様式

**\* 3 外部実体定義ファイル**

ISOlat1 (ラテン文字 1)、ISOdia (アクセント文字)、ISOnum (数字・特殊文字)

## (2) イメージデータ

イメージデータは、バイナリ形式のファイルであり、データの圧縮方式として MMR 方式\*<sup>1</sup>と JPEG 方式\*<sup>2</sup>の 2 種類が存在する。

### \* 1 MMR方式

当該図面が「白黒」の場合に用いる符号化の方式である。

「ITU-T (旧 CCITT) 勧告 T. 6 ベーシックファクシミリ符号化方式 (二次元圧縮方式)」

### \* 2 J P E G (Joint Photographic Expand Group) 方式

当該図面が「カラー」の場合に用いる符号化の方式である。

「カラー静止画像符号化国際標準方式」

## 1. 3 格納形式

### (1) S G M L データ

四法単位/マスタ単位で計 15 種類の提供単位が存在し、「S G M L 宣言ファイル」「D T D ファイル」「目録ファイル」「提供データファイル」「削除データファイル」の 5 種が存在する。

### (2) イメージデータ

公知資料イメージファイル、及び、商標見本ファイルがこれに該当し、「目録ファイル」「提供データファイル」の 2 種が存在する。

## 1. 4 格納体裁

DVD媒体の作成単位は、四法（特許・実用新案・意匠・商標）毎とする。  
四法ディレクトリの配下に、対象となるマスタ単位のディレクトリを作成する。

格納体裁のイメージについては、「DVD 格納体裁イメージ図」、及び、「DVD 格納時の媒体跨りについて」を参照。

### (1) SGMLデータ

- ① 【ルートディレクトリ】下に、【四法ディレクトリ】を作成する。
  - ・【四法ディレクトリ】内には「SGML宣言ファイル」「DTDファイル」を格納する。
- ② 【四法ディレクトリ】下に、【対象マスタディレクトリ】を作成する。
- ③ 【対象マスタディレクトリ】下に、【目次ディレクトリ（提供案件）】と【目次ディレクトリ（削除案件）】を作成する。
  - ・【目次ディレクトリ（提供案件）】内には、「目録ファイル（全案件）」を格納する。
  - ・【目次ディレクトリ（削除案件）】内には、「目録ファイル（全案件）」を格納する。
- ④ 【目次ディレクトリ（提供案件）】下に、【提供データディレクトリ】を作成する。
  - ・【提供データディレクトリ】内には、最大 1000 案件分を 1 ファイルとした「目録ファイル」と「提供データファイル」を格納する。
  - ・「目録ファイル」には、「目録ファイル（全案件）」から、「提供データファイル」に該当する 1000 件分を抜き出した情報を格納する。
- ⑤ 【目次ディレクトリ（削除案件）】下に、【削除データディレクトリ】を作成する。
  - ・【削除データディレクトリ】内には、最大 1000 案件分を 1 ファイルとした「目録ファイル」と「削除データファイル」を格納する。
- ⑥ 四法毎に作成される【対象マスタディレクトリ】は固定とする。

例えば、特許の場合で、ある提供ロットで「サーチマスタ」の提供がなされなかったとしても、「サーチマスタ」の【対象マスタディレクトリ】のみは作成する。

## (2) イメージデータ

SGMLデータを格納した媒体（意匠、又は、商標）中の、意匠ディレクトリ、又は、商標ディレクトリの配下に、公知資料イメージファイル、又は、商標見本ファイルのディレクトリを作成する。

- ①【公知資料イメージファイル、又は、商標見本ファイルディレクトリ】下に、【目次ディレクトリ（提供図面）】と【目次ディレクトリ（削除案件）】を作成する。
  - ・【目次ディレクトリ（提供図面）】内には、「目録ファイル（全図面）」を格納する。
  - ・【目次ディレクトリ（削除案件）】内には、「目録ファイル（全案件）」を格納する。
- ②【目次ディレクトリ（提供図面）】下に、【提供データディレクトリ】を作成する。
  - ・最大 1000 図面分を 1 ファイルとした「目録ファイル」を格納する。
  - ・【提供データディレクトリ】内には、1 図面ずつの「提供データファイル」を最大 1000 図面分格納する。
- ③【目次ディレクトリ（削除案件）】下には何も作成しない。
  - ・【目次ディレクトリ（削除案件）】内の「目録ファイル（全案件）」は、既提供済みの案件に対する削除宣言をするものであり、バイナリデータの実体は存在しない。

商標見本ファイルのイメージデータは、1 案件（出願）に対し、複数図面が存在する場合がある。（立体商標）

「提供データファイル」は図面単位とするため、この場合には、1 案件に対し、複数個の「提供データファイル」が存在することになるが、同一案件の複数図面については、「目録ファイル（全図面）」間での跨りは発生させず、【提供データディレクトリ】内の「提供データファイル」の数が 1000 図面分に満たない状況でも、次の【提供データディレクトリ】を対象にして、全図面情報を格納する。

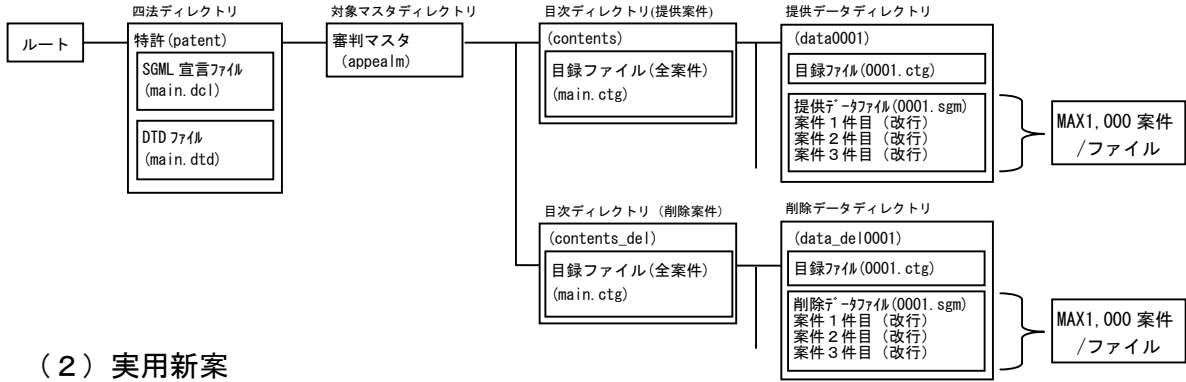
なお、削除宣言（削除情報の提供）は図面単位に行なうことはなく、1 案件単位に、複数個の図面データに対する一括削除宣言を行なう。

### 1. 5 提供周期

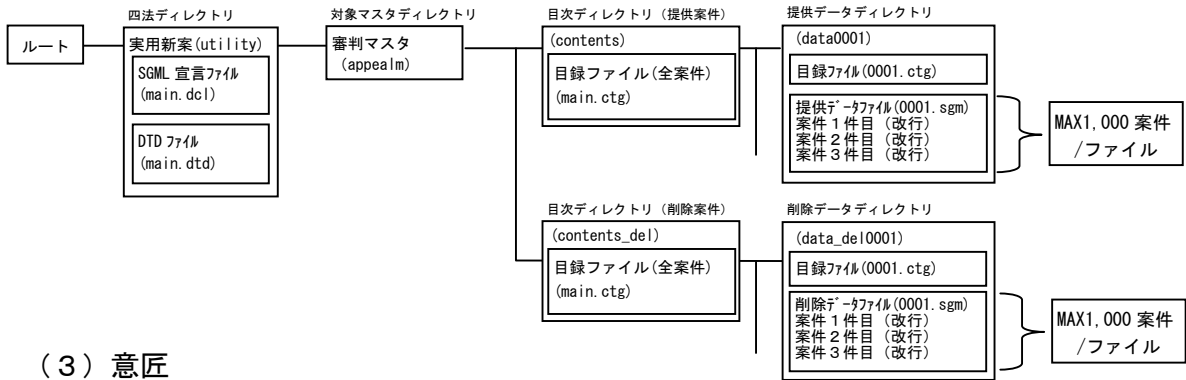
提供用DVD-Rは、基本的に毎週提供する。

# DVD格納体裁イメージ図

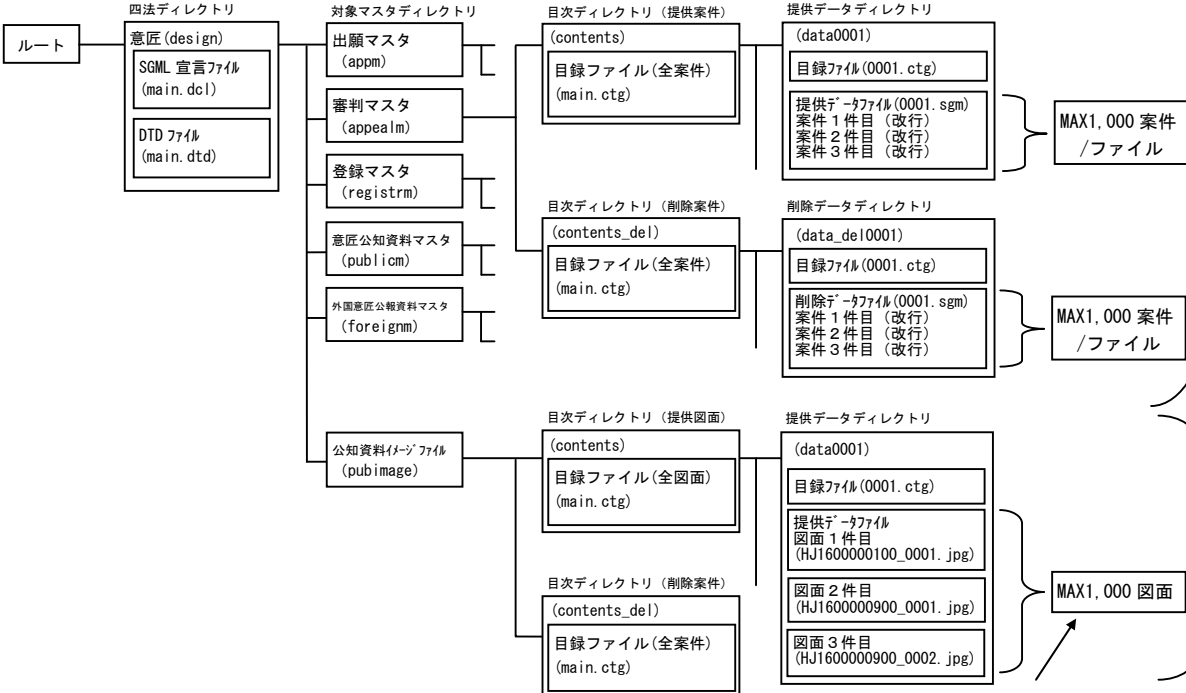
## (1) 特許



## (2) 実用新案



## (3) 意匠

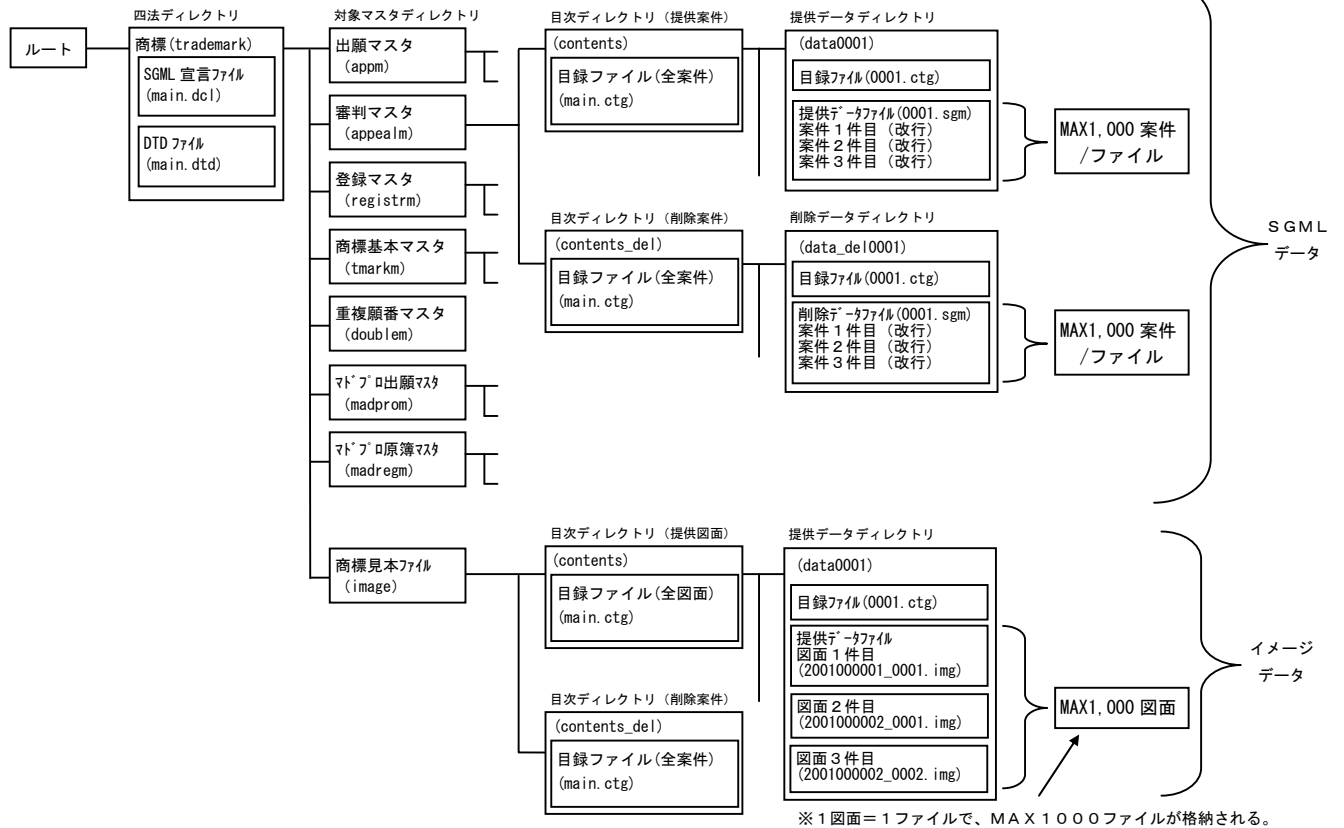


SGML  
データ

イメージ  
データ

※ 1 図面 = 1 ファイルで、MAX 1000 ファイルが格納される。

#### (4) 商標

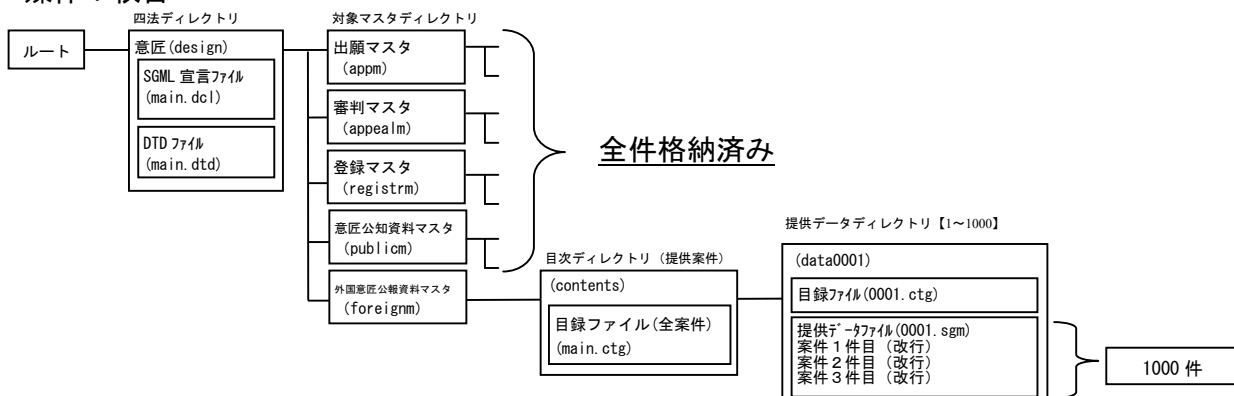


※重複願番マスタについては対象マスタディレクトリ (doublem) のみ存在する。

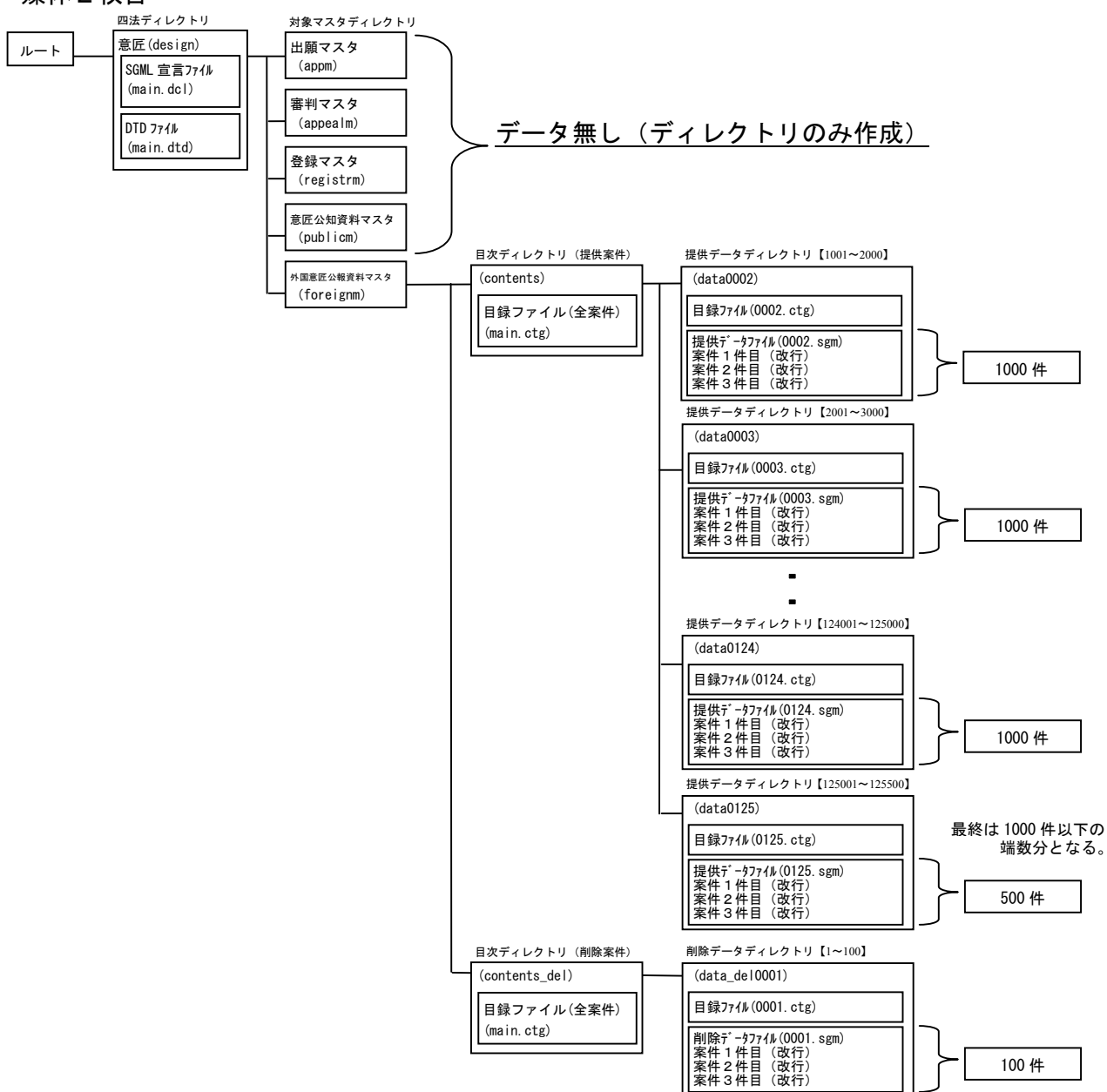
## DVD格納時の媒体跨りについて

(例) 意匠において、外国意匠公報資料マスタの提供データが 125,500 件、削除データが 100 件あり、提供データのうち 1,500 件分までで容量を超える場合

### 媒体 1 枚目



### 媒体 2 枚目



- (1) 1 枚目で格納済みの対象マスタは、2 枚目では空のディレクトリのみを作成する。
- (2) 提供データは 1000 件単位で作成する。  
 従って、外国意匠公報資料マスタは 1000 件分で 1 ファイルを作成し、1001 件目から 2 枚目の媒体に移る。
- (3) SGML 宣言ファイル、DTD ファイル、及び目録ファイル (全案件) は、同一のものを 1 枚目と 2 枚目の双方に作成する。
- (4) 削除データは、提供データを全て格納後、最後に作成するため、上記の例では 2 枚目にのみ作成する。

## 2. ファイル仕様

### 2. 1 SGMLデータ

#### 2. 1. 1 目録ファイル

「目録ファイル」には、当該ファイルに連なる「提供データファイル」内の全案件の番号情報を記録する。

記録する番号情報は案件をユニークに特定できるキー情報とする。

##### (1) ファイル形式

制御文字 (CR・LF) をレコード・デリミタとするテキストファイル形式とする。

レコードはマスタ毎に固定長とする。

##### (2) レコードの内容

以下に、マスタ毎のレコードの内容を記す。

#### A. 出願マスタ(固定長:26byte)

項番	項目名	属性	エリア長	備考
1	レコード長	C	3	“026”
2	国コード	C	2	“JP”:日本
3	四法コード	C	1	“P”:特許、“U”:実用新案、“D”:意匠、“T”:商標
4	出願番号			
	年	C	4	西暦4桁
	番号	C	6	前ゼロ付き6桁
5	データ作成日	C	8	YYYYMMDD
6	レコード・デリミタ	C	2	CR LF



B. 審判マスタ(固定長:44byte)

項番	項目名	属性	エリア長	備考
1	レコード長	C	3	“044”
2	国コード	C	2	“JP”:日本
3	四法コード	C	1	“P”:特許、“U”:実用新案、“D”:意匠、“T”:商標
4	出願番号			
	年	C	4	西暦4桁
	番号	C	6	前ゼロ付き6桁
5	審判番号			
	年	C	4	西暦4桁
	番号	C	6	前ゼロ付き6桁
6	庁内整理番号分割記号コード	C	1	【マドプロ案件】アルファベット1桁又はスペース 【マドプロ案件以外】初期値(スペース)
7	国際登録番号	C	7	【マドプロ案件】号:前ゼロ付き7桁 【マドプロ案件以外】初期値(スペース)
8	データ作成日	C	8	YYYYMMDD
9	レコード・デリミタ	C	2	CR LF

C. 登録マスタ(固定長:64byte)

項番	項目名	属性	エリア長	備考
1	レコード長	C	3	“064”
2	国コード	C	2	“JP”:日本
3	四法コード	C	1	“P”:特許、“U”:実用新案、“D”:意匠、“T”:商標
4	出願番号			
	年	C	4	西暦4桁
	番号	C	6	前ゼロ付き6桁
5	登録番号	C	7	前ゼロ付き7桁
6	分割番号	C	31	“1”又は“2”の繰返し 31桁(左詰め後スペース)
7	データ作成日	C	8	YYYYMMDD
8	レコード・デリミタ	C	2	CR LF

D. サーチマスタ

提供対象外のため欠番

E. 引用文献マスタ

提供対象外のため欠番

F. 意匠公知資料マスタ(固定長:38byte)

項番	項目名	属性	エリア長	備考			
1	レコード長	C	3	“038”			
2	国コード	C	2	“JP”:日本			
3	四法コード	C	1	“D”:意匠			
4	出願番号	C	10	“000000000”			
5	公知資料番号	C	2	“HA”:内国雑誌 “HB”:外国雑誌 “HC”:内国カタログ “HD”:外国カタログ “HJ”:インターネット情報 “HN”:新製品カタログ			
	資料区分						
	和暦年				C	2	前ゼロ付き2桁
	通番				C	6	前ゼロ付き6桁
	枝番	C	2	前ゼロ付き2桁			
6	データ作成日	C	8	YYYYMMDD			
7	レコード・デリミタ	C	2	CR LF			

G. 外国意匠公報資料マスタ(固定長:38byte)

項番	項目名	属性	エリア長	備考			
1	レコード長	C	3	“038”			
2	国コード	C	2	“JP”:日本			
3	四法コード	C	1	“D”:意匠			
4	出願番号	C	10	“000000000”			
5	公知資料番号	C	2	“HH”			
	文献種別						
	和暦年				C	2	前ゼロ付き2桁
	通番				C	6	前ゼロ付き6桁
	枝番	C	2	前ゼロ付き2桁			
6	データ作成日	C	8	YYYYMMDD			
7	レコード・デリミタ	C	2	CR LF			

H. 商標基本マスタ(固定長:72byte)

項番	項目名	属性	エリア長	備考
1	レコード長	C	3	“072”
2	国コード	C	2	“JP”:日本
3	四法コード	C	1	“T”:商標
4	出願番号			(マドプロ案件の場合:整理標準化マドプロ管理番号)
	年	C	4	西暦4桁
	番号	C	6	前ゼロ付き6桁
5	登録番号	C	7	【国内案件の場合】 号:前ゼロ付き7桁 未登録の場合は初期値(スペース) 【マドプロ案件の場合】 初期値(スペース)
6	分割番号	C	31	【国内案件の場合】 “1”又は“2”の繰返し 31桁(左詰め後スペース) 【マドプロ案件の場合】 初期値(スペース)
7	国際登録番号	C	7	【国内案件の場合】 初期値(スペース) 【マドプロ案件の場合】 号:前ゼロ付き7桁
8	国際登録番号分割記号	C	1	【国内案件の場合】 初期値(スペース) 【マドプロ案件の場合】 アルファベット 1桁又はスペース
9	データ作成日	C	8	YYYYMMDD
10	レコード・デリミタ	C	2	CR LF

I. 重複願番マスタ

提供対象外のため欠番

J. マドプロ出願マスタ(固定長:27byte)

項番	項目名	属性	エリア長	備考
1	レコード長	C	3	“027”
2	国コード	C	2	“JP”:日本
3	四法コード	C	1	“T”:商標
4	整理標準化マドプロ管理番号			
	年	C	4	西暦4桁
	番号	C	6	前ゼロ付き6桁
5	整理標準化マドプロ分割記号	C	1	アルファベット 1桁又はスペース
6	データ作成日	C	8	YYYYMMDD
7	レコード・デリミタ	C	2	CR LF

K. マドプロ原簿マスタ(固定長:45byte)

項番	項目名	属性	エリア長	備考
1	レコード長	C	3	“045”
2	国コード	C	2	“JP”:日本
3	四法コード	C	1	“T”:商標
4	整理標準化マドプロ管理番号			
	年	C	4	西暦4桁
	番号	C	6	前ゼロ付き6桁
5	整理標準化マドプロ分割記号	C	1	アルファベット 1桁又はスペース
6	国際登録番号更新回数記号	C	2	英数字2桁 国際登録番号更新回数記号がない場合は 初期値(スペース)
7	国際登録番号	C	7	号:前ゼロ付き7桁
8	国際登録番号分割記号	C	1	アルファベット 1桁又はスペース
9	事後指定日	C	8	YYYYMMDD 事後指定日がない場合は初期値(スペース)
10	データ作成日	C	8	YYYYMMDD
11	レコード・デリミタ	C	2	CR LF

## 2. 1. 2 SGML宣言ファイル、DTDファイル

「SGML宣言ファイル」「DTDファイル」には、各々、SGML宣言とSGMLのDTDを記録する。

### (1) ファイル形式

テキストファイル形式とする。

レコードは可変長とする。

### (2) レコードの内容

レコードの内容は、「SGML宣言リスト」及び「DTDリスト」を参照。

## 2. 1. 3 提供データファイル

「提供データファイル」には、1案件のSGMLを1SDIFレコードとして記録する。

### (1) ファイル形式

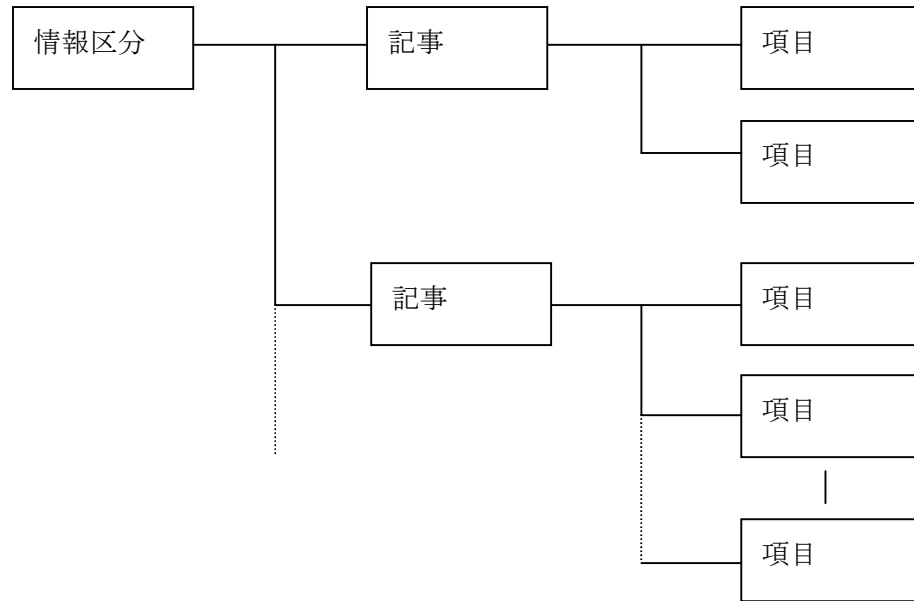
1提供データファイル中には目録ファイルのキー情報にて特定される1案件毎のSDIFレコードをMAX1000件格納する。

案件と案件の間には、制御文字（CR・LF）をレコード・デリミタとして挿入して、別レコードとして格納するが、1案件に対してのレコード分割は発生しない。

レコードは可変長とする。

## (2) レコードの構造

1SDIFレコードは1つの情報区分と、その情報区分に定義されている記事、項目により構成される。以下に、その構造について示す。



1SDIFレコードの基本構造概念図

### (a) 情報区分

情報区分には、以下の8種類のいずれかが存在する。

#### ① 基本情報

主に、出願（国際商標登録出願を除く）に関する書誌事項等の記事の上位概念。

#### ② 審判情報

当該出願に対する審判情報関連の記事の上位概念。

#### ③ 登録情報

当該出願（国際商標登録出願を除く）に関する登録情報関連の記事の上位概念。

- ④ サーチ情報  
提供対象外のため欠番
- ⑤ 国内出願引用文献情報  
提供対象外のため欠番
- ⑥ 意匠公知資料情報  
当該出願に関する意匠公知資料関連の記事の上位概念。
- ⑦ 外国意匠公報資料情報  
当該資料に関する外国意匠公報関連の記事の上位概念。
- ⑧ 商標公開情報  
当該資料に関する商標独自に発生する記事の上位概念。
- ⑨ マドリッドプロトコル情報  
日本を指定国とする国際商標登録出願に関する書誌事項及び登録情報関連の記事の上位概念。
- ⑩ マドリッドプロトコル原簿情報  
日本を指定国とする国際商標登録出願に関する原簿情報関連の記事の上位概念。

(b) 記事

主に、各情報区分に対応した同一事象又は連続事象で発生しうる項目の上位概念。案件毎の状況で、存在状況に有無がある。

(c) 項目

整理標準化データの提供対象項目群。  
記事同様、案件毎の状況で、存在状況に有無がある。

上記、情報区分・記事・項目については「第Ⅱ章 3. データ項目一覧表」を参照。

## 2. 1. 4 削除データファイル

「削除データファイル」には、「提供データファイル」と同様、1 案件の SGML を 1 SDIF レコードとして記録するが、その内容は、各マスタ毎に目録ファイルに定義されている番号情報に対応する記事・項目、及び更新日付に限定されている。

上記、情報区分・記事・項目については「第Ⅱ章 4. 削除データ項目一覧表」を参照。



## 2. 2 イメージデータ

### 2. 2. 1 目録ファイル

#### (1) ファイル形式

制御文字 (CR・LF) をレコード・デリミタとするテキストファイル形式とする。

レコードはファイル毎に固定長とする。

#### (2) レコードの内容

以下に、ファイル毎のレコードの内容を記す。

#### A. 商標見本ファイル(固定長:63byte)

項番	項目名	属性	エリア長	備考
1	レコード長	C	3	“063”
2	国コード	C	2	“JP”:日本
3	文献種別	C	2	“T1”:商標
4	出願番号			(西暦4桁が“0000”の時は回復番号)
	年	C	4	西暦4桁
	番号	C	6	前ゼロ付き6桁
5	図面番号	C	4	前ゼロ付き4桁 *1
6	更新日付	C	8	YYYYMMDD (特許庁内でのデータ作成日)
7	図面数	C	4	“0001”~“9999” *1
8	原寸の大きさ			
	縦	C	3	“001”~“999” (mm) *1
	横	C	3	“001”~“999” (mm) *1
9	圧縮方式	C	2	“M2”:MMR方式 “JP”:JPEG方式 *2
10	解像度	C	2	“16”又は“00” (16:MMR方式、00:JPEG方式) *2
11	ライン数			
	縦	C	4	“0001”~“9999” (本) *1
	横	C	4	“0001”~“9999” (本) *1
12	イメージデータ長	C	10	0以上 (ゼロの場合は“削除案件”を意味する)
13	レコード・デリミタ	C	2	CR LF

\*1 削除案件の場合、エリア長分、全てゼロ。

\*2 削除案件の場合、エリア長分、全てスペース。

B. 意匠公知資料イメージファイル(固定長:42byte)

項番	項目名	属性	エリア長	備考
1	レコード長	C	3	“042”
2	国コード	C	2	“JP”:日本
3	四法コード	C	1	“D”:意匠
4	出願番号		10	“0000000000”
5	公知資料番号			
	資料区分	C	2	“HA”:内国雑誌 “HB”:外国雑誌 “HC”:内国カタログ “HD”:外国カタログ “HJ”:インターネット情報 “HN”:新製品カタログ
	和暦年	C	2	前ゼロ付き2桁
	通番	C	6	前ゼロ付き6桁
	枝番	C	2	前ゼロ付き2桁
6	図面番号	C	4	前ゼロ付き4桁 *1
7	データ作成日	C	8	YYYYMMDD
8	レコード・デリミタ	C	2	CR LF

\*1 削除案件の場合、エリア長分、全てゼロ。

備考

\* 圧縮方式:JPEG 方式

\* ライン数:縦 480 本、横 640 本

2. 2. 2 提供データファイル

(1) ファイル形式

1 図面毎に 1 提供ファイルを作成する。

レコードは可変長とする。

(2) レコードの内容

1 提供ファイル= 1 レコードとし、レコード分割は行なわない。

バイナリ形式の「イメージデータ」のみで構成される。

### 3. 文字コード

#### (1) 1バイトコード

『JIS X 0201-1976 7ビット及び8ビットの情報交換用符号化文字集合』のローマ文字用図形文字集合（JIS X 0201-1976 表2）をG0集合に指示し、列2～7へ呼び出したものを使用する。

#### (2) 2バイトコード

『JIS X 0208-1997 7ビット及び8ビットの2バイト情報交換用符号化漢字集合』をG1集合に指示し、列10～15へ呼び出したもの（JIS X 0208-1990の図形文字符号の各バイトの8ビット目を“1”にしたもの）を使用する。

尚、従来から、特許庁内で独自にフォントを割り当ててある特定文字（特許庁外字）については、整理標準化データ仕様書に記す基準で、該当文字列に変換した値を提供データとする。

当該文字列（「特定文字の提供形式一覧」参照）は、基本情報、登録情報の各情報区分に存在する。

審判情報及び商標公開情報の各情報区分についても、平成11年までの新規・更新データを含む提供データ上は、上記基準が適用されるが、特許庁の審判マスタ及び商標基本マスタの蓄積方式が平成11年末に変更され、その移行時に、特殊文字（特許庁外字）については別基準で変換・移行されたため、平成12年以降の新規・更新データを含む提供データ上では、既に変換済みの文字列がそのまま提供されることとなる。

上記、移行時の変換基準については「特定文字の提供形式一覧Ⅱ」を参照。

### (3) 制御文字コード

改行 (CR), 復帰 (LF) を用いる。

名称	記号	符号化表現
改行	CR	(0D) <sub>16</sub>
復帰	LF	(0A) <sub>16</sub>

### (4) 文字の置き換え

以下の文字列はタグではなく、タグの開デリミタ“<”と閉デリミタ“>”を、データとして用いることができるように、表示または印刷時に文字を置き換えて使用する文字列である。

以下の文字列がデータ中に現れた場合、文字列を表示または印刷時に置き換え文字に置き換える。全て1バイトコードである。

文字列	置き換え文字
&gt;	>
&lt;	<
&amp;	&
&apos;	'
&quot;	“

### (5) ラテン文字 (実体参照)

マドリッドプロトコル情報区分及び、マドリッドプロトコル原簿情報区分においては、文字コード ISO-8859-1(Latin1)に含まれるラテン文字を表現するために、ラテン文字実体集合一覧に示す文字実体集合を使用している。これらの文字実体集合は、マドリッドプロトコル情報区分及び、マドリッドプロトコル原簿情報区分上のデータに実体参照として記録される。

尚、これらの文字実体集合は ISO-8879 (JIS X 4151) に掲載されている、

『ISO 8879-1986//ENTITIES Added Latin 1//EN』

『ISO 8879-1986//ENTITIES Numeric and Special Graphic//EN』

『ISO 8879-1986//ENTITIES Diacritical Marks//EN』

で示す公開識別子で指定される公開実体宣言に準拠している。

上記、文字実体集合については「ラテン文字実体集合一覧」を参照。