

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 (IS09660)) 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 *¹ データは、SDIF *² 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 *³ 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*¹と JPEG 方式*²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 2 0 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

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

【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。

【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。

【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。

【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。

【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。

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

S G M L データを格納した媒体(商標)中の、商標ディレクトリの配下に、商標見本ファイルのディレクトリを作成する。

【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。

【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。

【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。

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。

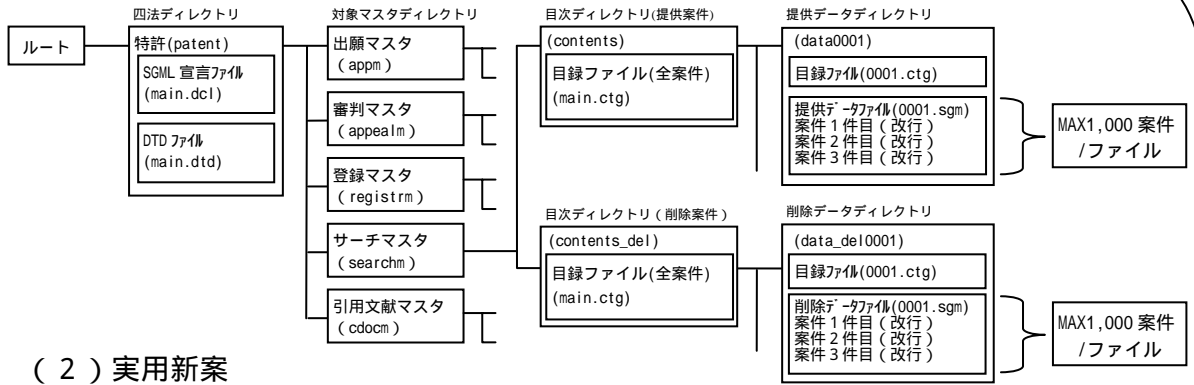
なお、削除宣言(削除情報の提供)は図面単位に行なうことはなく、1 案件単位に、複数個の図面データに対する一括削除宣言を行なう。

1.5 提供周期

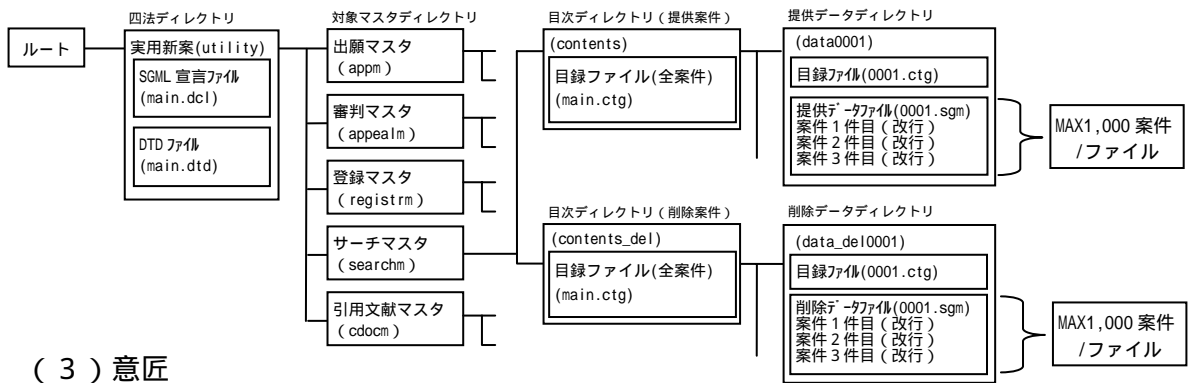
提供用 DVD - R は、基本的に隔週で提供する。

DVD格納体裁イメージ図

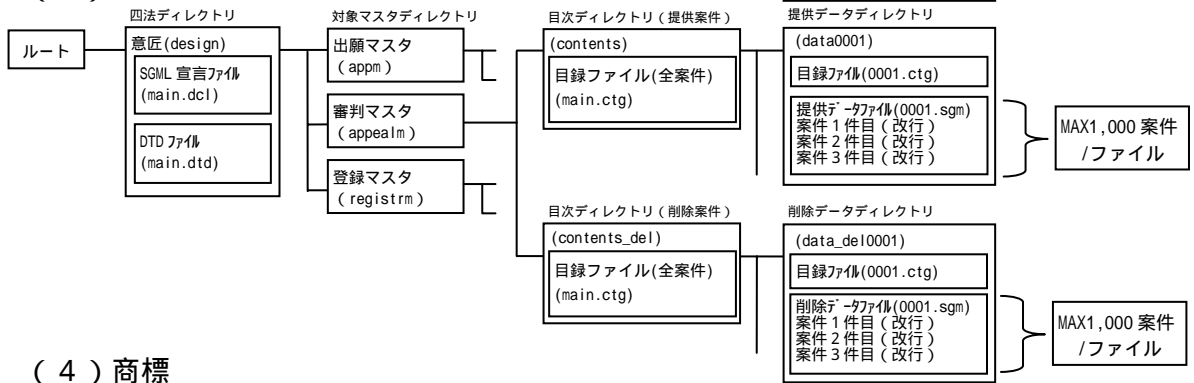
(1) 特許



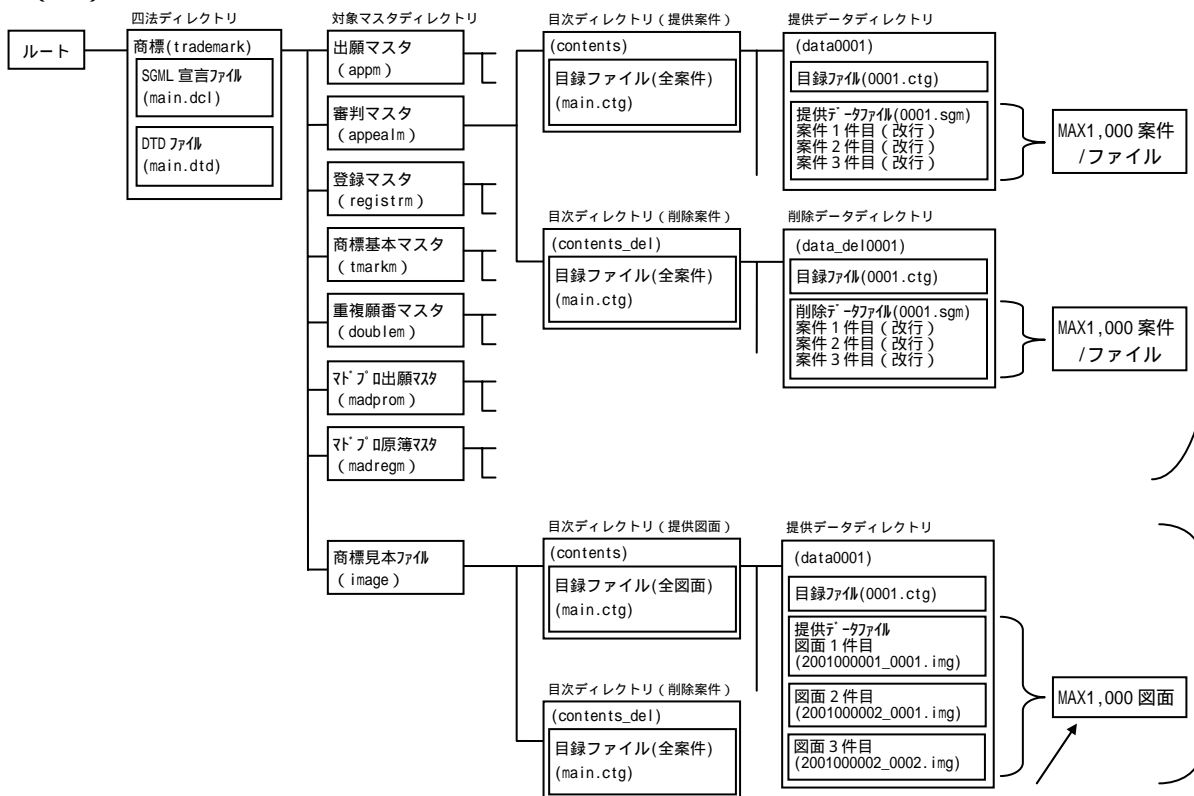
(2) 実用新案



(3) 意匠



(4) 商標



SGML
データ

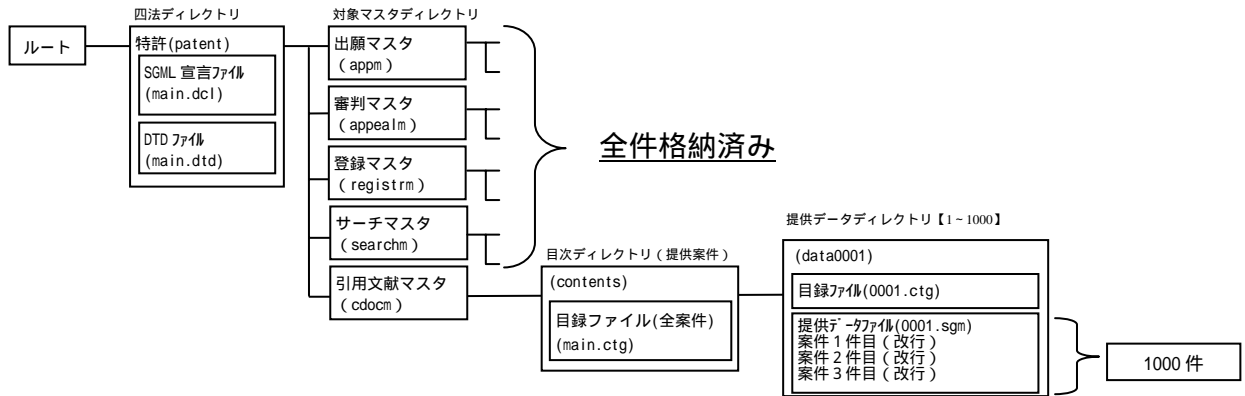
イメージ
データ

1 図面 = 1 ファイルで、MAX1000ファイルが格納される。

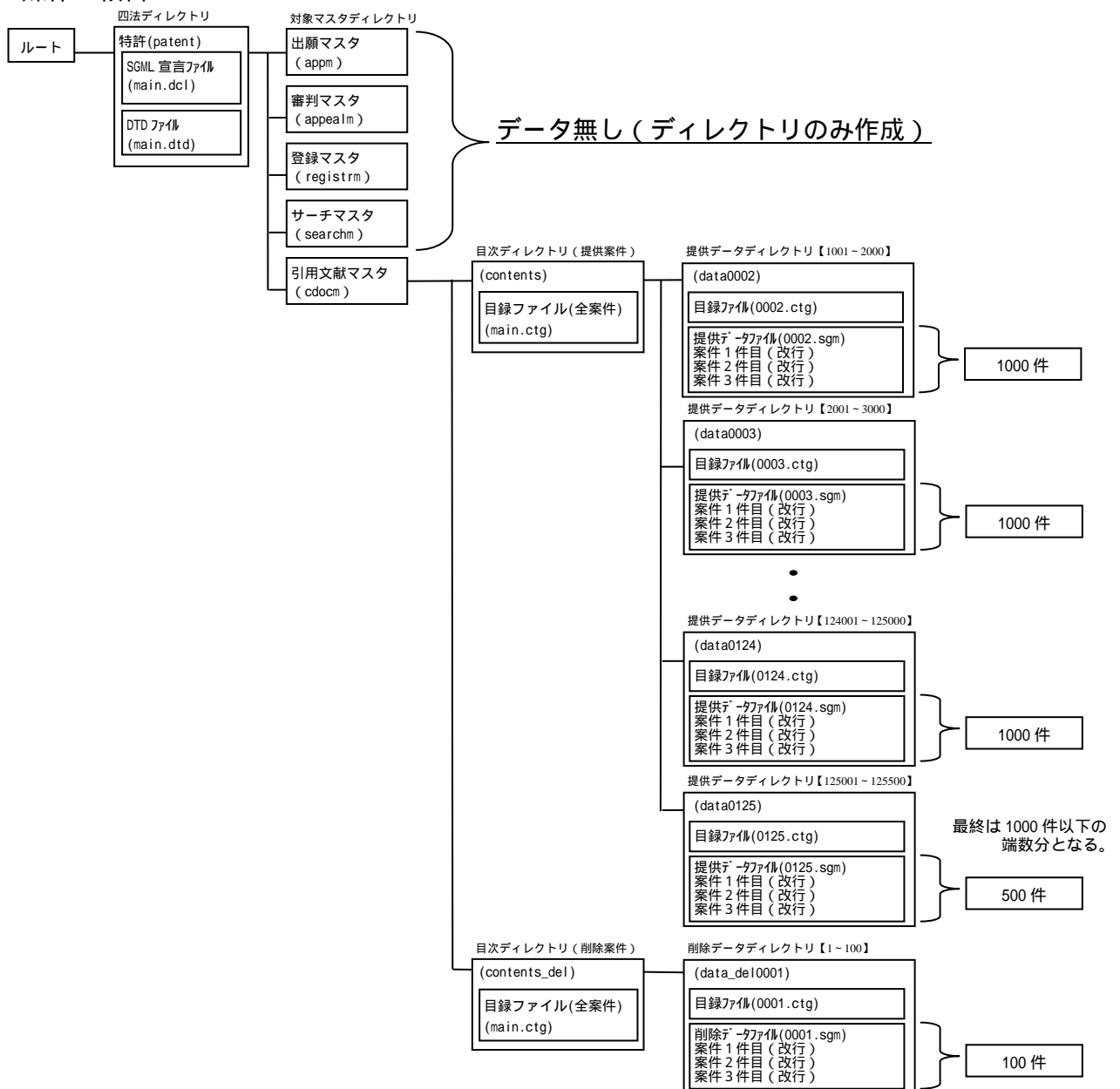
DVD格納時の媒体跨りについて

(例) 特許において、引用文献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、引用文献マスタは 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 枚目でのみ作成する。

2 . ファイル仕様

2 . 1 S G M L 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 (C R ・ L F) 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 : 2 6 byte)

項番	項目名	属性	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:36byte)

項番	項目名	属性	IR7長	備考
1	レコード長	C	3	“036”
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	データ作成日	C	8	“YYYYMMDD”
7	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	IR7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) 商標:(権利分割がある場合のみ発生) 他の場合は初期値(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(固定長:37byte)

項番	項目名	属性	IR7長	備考
1	レコード長	C	3	“037”
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
4	出願番号			*1
	年	C	4	西暦4桁
	番号	C	6	前ゼロ付き6桁
5	文献情報 種別	C	1	“A”:特許(公開、公表、再公表) “B”:特許(公告、登録) “U”:実用新案 (公開、公表、再公表、登録実用新案) “Y”:実用新案(公告、登録) “C”:特許(明細書) “Z”:実用新案(明細書)
	年	C	4	西暦4桁
	番号	C	6	前ゼロ付き6桁 *2 登録番号の場合、 “000”+登録番号(7桁)
6	データ作成日	C	8	“YYYYMMDD”
7	レコード・デリミタ	C	2	CR LF

*1 項番5の種別が“C”または“Z”の時、サーチ情報独自の番号。

以下の形式で記録される。

項番	項目名	属性	IR長	備考
4	出願番号			(サーチ情報独自の番号)
	文献種別	C	1	“C”:特許(明細書)、“Z”:実用新案(明細書)
	番号部	C	9	9桁の数字

また、種別が“C”または“Z”の時には、提供データファイルまたは、削除データファイル内の当該案件には「出願記事」は存在しない。

*2 「再公表番号」の場合、以下の形式で記録される。

先頭1桁：“8”

残り5桁：国際公開番号の番号部5桁(前0付き)

但し、5桁中の先頭1桁目が英字で記録されている場合がある。

その場合の英字は「A 10」～「Z 35」を意味する。

【具体例】 8A0001 8100001

8E0010 8150010

E. 引用文献マスタ(固定長:26byte)

項番	項目名	属性	IR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
4	出願番号			
	年 番号	C C	4 6	西暦4桁 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

F. 商標基本マスタ(固定長:72byte)

項番	項目名	属性	IR7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、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) 権利分割がある場合のみ発生 他の場合は初期値(スペース) [マドプロ案件の場合] 初期値(スペース)
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 権利分割がある場合のみ発生 他の場合は初期値(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

G. 重複願番マスタ(固定長:64byte)

項番	項目名	属性	IR7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	“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 未登録の場合、初期値(スペース)
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) 権利分割がある場合のみ発生 他の場合は初期値(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

H. 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 権利分割がある場合のみ発生 他の場合は初期値(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	

I. 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 権利分割がある場合のみ発生 他の場合は初期値(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 権利分割がある場合のみ発生 他の場合は初期値(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

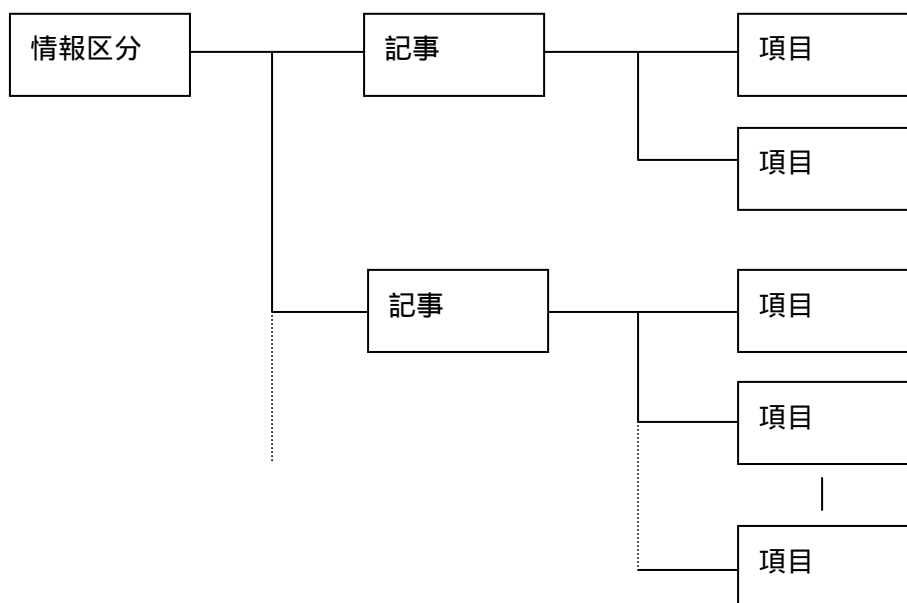
1提供データファイル中には目録ファイルのキー情報にて特定される1案件毎のSDIFレコードをMAX1000件格納する。

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。

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。

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。

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。

サーチ情報

当該出願に関する検索用情報関連の記事の上位概念。

国内出願引用文献情報

当該出願に関する国内引用文献情報関連の記事の上位概念。

商標公開情報

当該出願に関する商標独自に発生する記事の上位概念。

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。

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 3 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 S G M L を 1 S D I F 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 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。

レコードは固定長(63byte)とする。

(2) レコードの内容

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	日本:“JP”
3	文献種別	C	2	商標:“T1”
4	出願番号			(西暦4桁が“0000”の時、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	MMR方式 :“M2” JPEG方式 :“JP” *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。

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 1 0 ~ 1 5 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、サーチ情報、国内出願引用文献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 1 1 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 1 1 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 1 2 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」を参照。

(3) 制御文字コード

改行 (C R) , 復帰 (L F) を用いる。

名称	記号	符号化表現
改行	C R	(0D) ₁₆
復帰	L F	(0A) ₁₆

(4) 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。