

# 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

特許・実用新案における整理標準化データの内、「出願マスタ」「登録マスタ」「サーチマスタ」「引用文献マスタ」の各情報区分をXML<sup>\*1</sup>データ形式により提供する。

**\* 1 XML (eXtensible Markup Language)**

## 1. 3 格納形式

特許・実用新案を対象とし、マスタ単位で計8種類の提供単位が存在し、「DTD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4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）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 格納納時の媒体跨りについて」を参照。

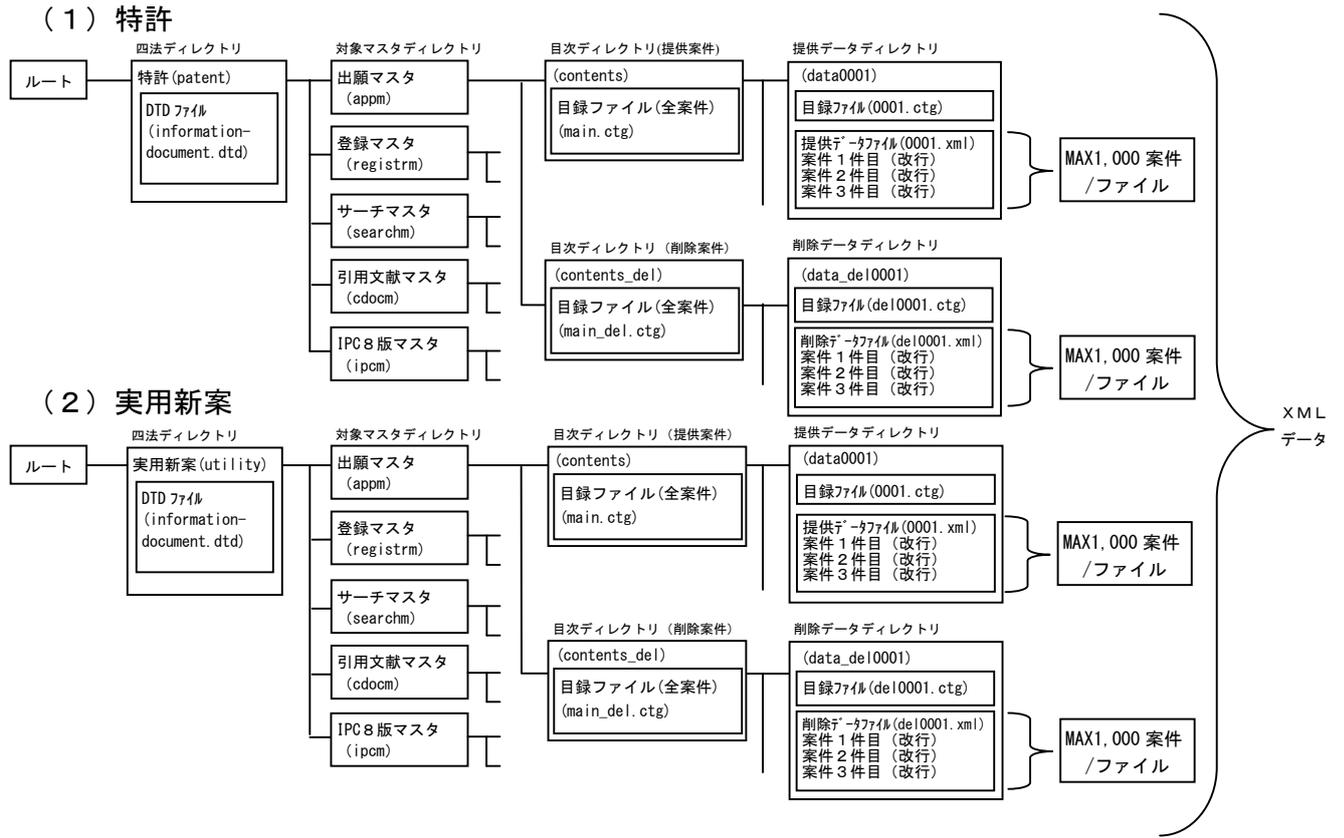
- ① 【ルートディレクトリ】下に、【四法ディレクトリ】を作成する。
  - ・【四法ディレクトリ】内には「DTDファイル」を格納する。
- ② 【四法ディレクトリ】下に、【対象マスタディレクトリ】を作成する。
- ③ 【対象マスタディレクトリ】下に、【目次ディレクトリ（提供案件）】と【目次ディレクトリ（削除案件）】を作成する。
  - ・【目次ディレクトリ（提供案件）】内には、「目録ファイル（全案件）」を格納する。
  - ・【目次ディレクトリ（削除案件）】内には、「目録ファイル（全案件）」を格納する。
- ④ 【目次ディレクトリ（提供案件）】下に、【提供データディレクトリ】を作成する。
  - ・【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。

## 1. 5 提供周期

平成 26 年度第 20 回までの提供用 DVD-R は、基本的に隔週で提供する。  
平成 26 年度第 21 回以降の提供用 DVD-R は、基本的に毎週提供する。

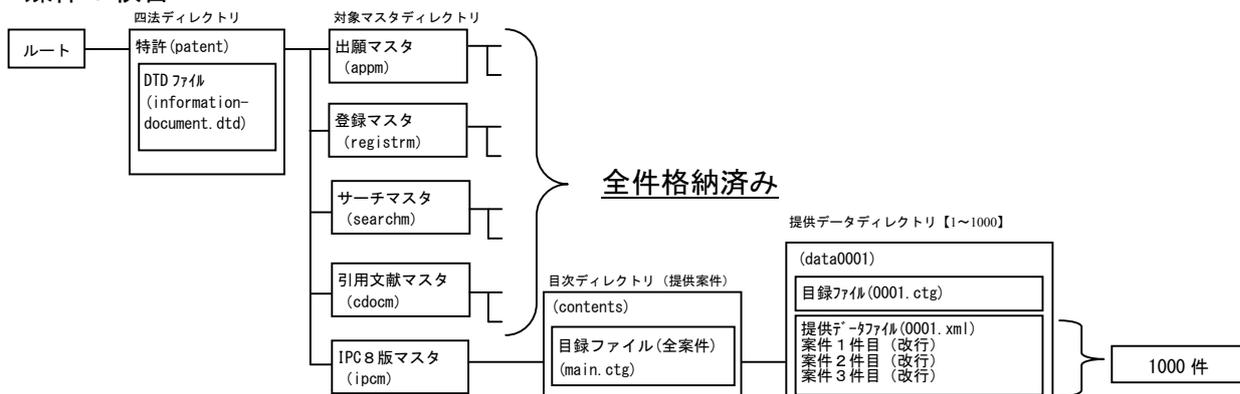
# DVD格納体裁イメージ図



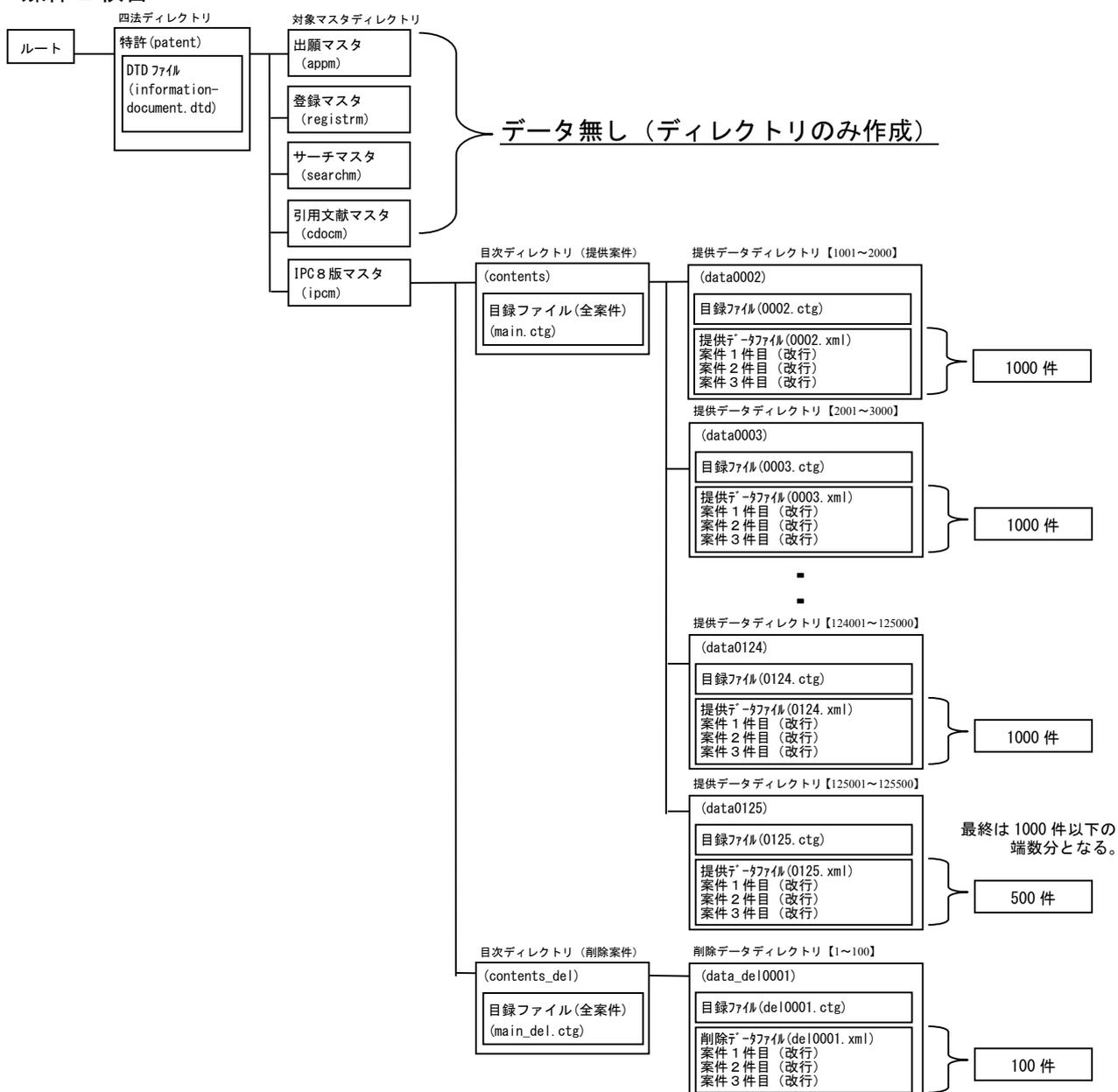
## DVD格納時の媒体跨りについて

(例) 特許において、IPC8版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、IPC8版マスタは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) 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 XML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。

※以下の表記中における“△”は1文字の空白(スペース)を表す

#### 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
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. 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
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	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

C. サーチマスタ(固定長:37byte)

項番	項目名	属性	エリア長	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
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	エリア長	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

\*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

8F0001 → 8150001

D. 引用文献マスタ(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
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

E. IPC8版マスタ(固定長:40byte)

項番	項目名	属性	エリア長	備考
1	レコード長	C	3	“040”
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	2	“A $\Delta$ ”:特許(公開、公表) “B1”:特許(登録) “B2”:特許(公告、登録) “U $\Delta$ ”:実用新案(公開、公表、登録実用新案) “Y1”:実用新案(登録) “Y2”:実用新案(公告、登録) “A1”:特許(再公表)、実用新案(再公表) “C $\Delta$ ”:特許(明細書) “Z $\Delta$ ”:実用新案(明細書)
	文献番号	C	12	*2
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 $\Delta$ ”または“Z $\Delta$ ”の時、国際特許分類(IPC)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	エリア長	備考
4	出願番号			(国際特許分類(IPC)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

\*2 同項番内の種別によって異なる。

種別により以下の形式で記録される。

項番	項目名	属性	エリア長	備考
“A $\Delta$ ”、“B2”(公告)、“U $\Delta$ ”(公開、公表)、“Y2”(公告) の場合				
	文献番号			
	年	C	4	西暦4桁
	番号	C	6	前ゼロ付き6桁
	スペース	C	2	“ $\Delta\Delta$ ”
“B1”、“B2”(登録)、“U $\Delta$ ”(登録実用新案)、“Y1”、“Y2”(登録) の場合				
	文献番号			
	前ゼロ	C	3	“000”
	番号	C	7	登録番号7桁
	スペース	C	2	“ $\Delta\Delta$ ”
“A1” の場合				
	文献番号			
	国コード	C	2	“WO”
	年	C	4	西暦4桁
	番号	C	6	前ゼロ付き6桁
“C $\Delta$ ”、“Z $\Delta$ ” の場合				
	文献番号			
	番号	C	10	前ゼロ付き10桁
	スペース	C	2	“ $\Delta\Delta$ ”

## 2. 1. 2 DTDファイル

「DTDファイル」には、XML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「別添 DTDリスト」を参照のこと。

## 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XMLを1レコードとして記録する。

### (1) ファイル形式

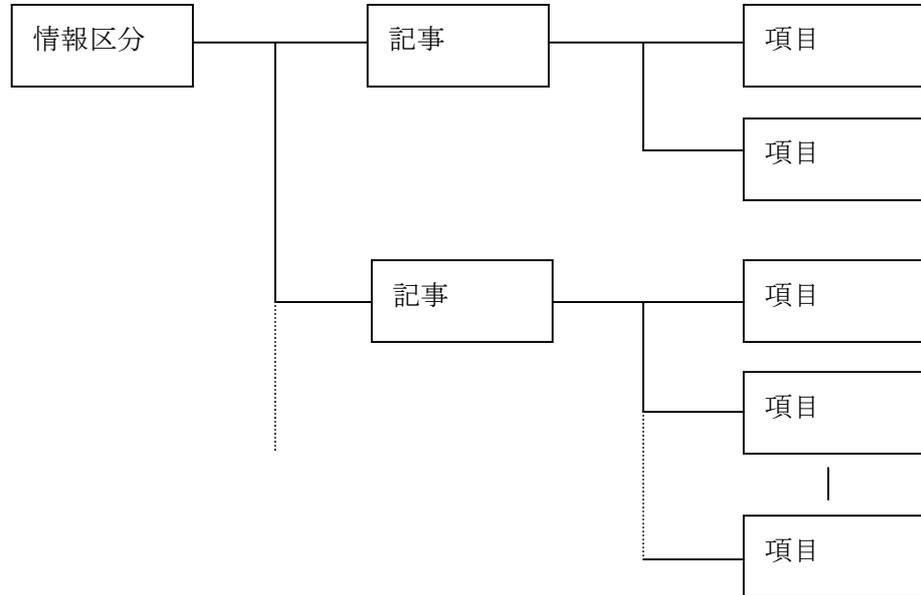
1 提供データファイル中には目録ファイルのキー情報にて特定される1案件毎のレコードをMAX1000件格納する。

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

1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。



1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5種類のいずれかが存在する。

① 基本情報

主に、出願に関する書誌事項等の記事の上位概念。

② 登録情報

当該出願に関する登録情報関連の記事の上位概念。

③ 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。

④ 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。

⑤ 国際特許分類(IPC)情報

当該出願に関する国際特許分類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。

主な項目は、記事単位で ST. 36、DVD-ROM 公報仕様および特許庁 XML 交換仕様に準拠している。

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 XML を 1 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。

### 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）は各情報区分に存在する。

#### (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;	"