

情報知識学会
SGML研修フォーラム

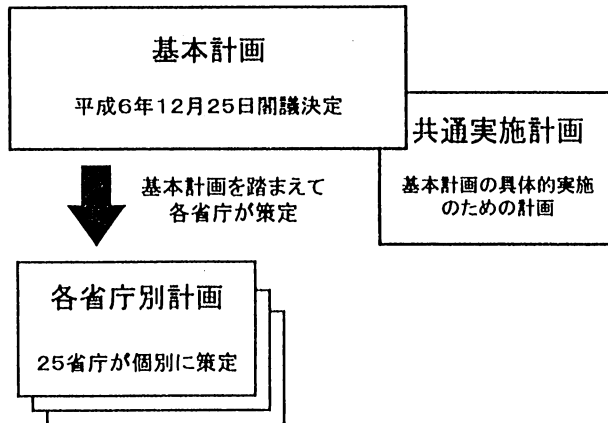
白書等データベースにおける SGMLの採用

平成9年2月5日
総務庁行政管理局
行政情報システム企画課
山本 寛繁

項 目

- 行政情報化推進計画
- 情報化推進基盤の整備
- 白書等データベースの整備
- SGMLを採用した統一的な仕様
- 各省庁における取り組み
- 今後の課題

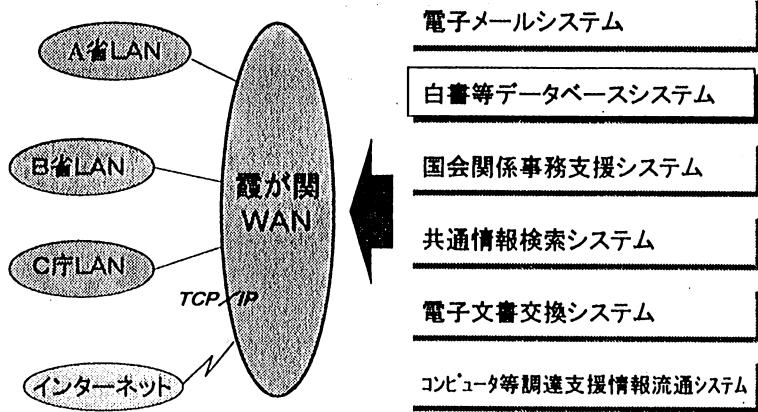
行政情報化推進計画の構成



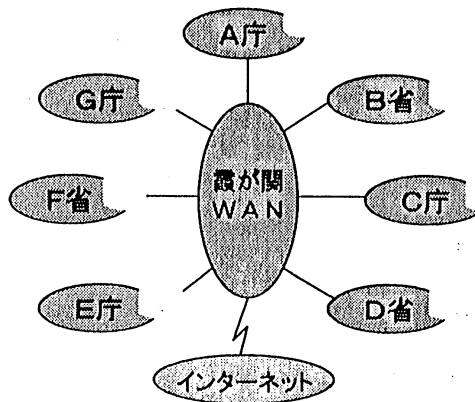
情報化推進基盤の整備

- 中央省庁におけるパソコンの整備
 - パソコン配備率 1.31人/台(平成8年3月)
- 各省庁施設内ネットワーク(各省庁LAN)の整備
 - 25省庁のうち22省庁整備(平成8年3月)
- 省庁間ネットワーク(霞が関WAN)の整備
 - 国際的な標準仕様を基本的に採用
 - 電子メールシステムから10機関でスタート(平成9年1月~)
 - 中央省庁等31機関の利用を予定

霞が関WANの各種システム



データベース省庁間利用の概念図



白書等データベースの整備

◆行政情報化推進計画◆

白書・年次報告書、調査研究報告書等の
統一的な仕様に基づくデータベース化

白書
データベース
(SGML)

パイロットシステム
(情報処理振興事業協会:IPA)

成果

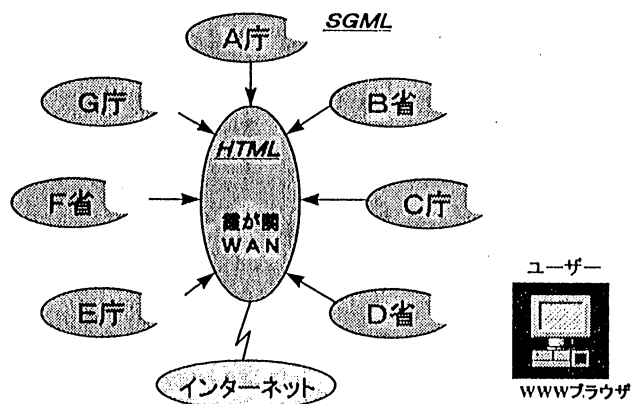
「白書等データベースの
統一的な仕様」の決定

(平成8年6月18日)

統一的な仕様とは

- 対象情報 ⇒ 白書・年次報告書、調査研究報告書…
- 電子化形式 ⇒ 国際的な標準形式(テキスト、GIF…)
- 使用文字 ⇒ 原則JIS第1水準・JIS第2水準(漢字)
- 文書構造形式 ⇒ SGML
- システム機能 ⇒ 全文検索、目次検索、HTML変換
- 霞が関WANとの連携 ⇒ TCP/IP、WWWブラウザ

白書等データベースの省庁間利用

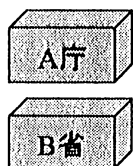


文書構造規格としてSGMLを採用

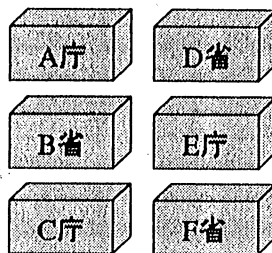
- 統一的なSGML宣言の作成(標準SGML宣言)
- 統一的なDTDの作成(標準DTD)
- ISOの「Book DTD」に準拠

白書等データベースの予算措置

平成8年度
2省庁：44百万円



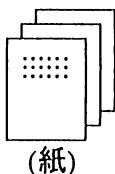
平成9年度
6省庁：81百万円



(注)平成9年度は政府案ベース

SGML化の実施

白書等原稿



(紙)



(FD)

SGML化委託

大蔵省印刷局

民間業者

大蔵省印刷局は

- ほとんどの白書を発行
- 電算写植校了データを保有
- IPAのパイロットシステムを受注



統一的な仕様に基づくSGML化受注へ

白書のCD-ROM提供

○国会・閣議に報告する白書 35白書

CD-ROM化された白書 9白書

経済企画庁:経済白書・国民生活白書
科学技術庁:科学技術白書
環境庁:環境白書
国土庁:防災白書
文部省:教育白書
通商産業省:通商白書・中小企業白書
郵政省:通信白書



今後の課題

- SGML関連ツールの普及
- SGMLを前提とした共同執筆、改訂作業
- 文書自体の構造の標準化
- 使用文字（外字の取り扱い）
- 他システムへのSGML活用

白書等データベースの統一的な仕様

平成 8 年 6 月 1 8 日
行政情報システム各省庁連絡会議幹事会了承

白書等データベースの統一的な仕様は、以下のとおりとする。

第1 対象となる行政情報

本仕様は、各省庁が共通して保有する行政情報のうち、白書・年次報告書、調査研究報告書及びこれらに類するものを対象とする。

第2 電子化方法

紙、フロッピーディスク等により保有しているデータは、次の方法により電子化を行う。

1 電子化形式

(1) 次のデータは、テキスト形式により電子化する。

- ア 本文等
- イ 図、表及び写真の標題、注書等

(2) 次のデータは、国際的な標準の形式によりイメージデータとして電子化する。

- ア 図、表及び写真等
- イ その他テキスト形式による電子化が不可能なもの

(3) 動画像及び音声のデータは、国際的な標準の形式により電子化する。

(4) 表計算ソフトウェア等により作成された数値等データをファイル転送により交換する場合においては、国際的な標準の形式により電子化する。

2 使用文字

(1) 文字セット

- ア 1バイト文字の英数字及び記号は、J I S X 0 2 0 1を使用する。
- イ 2バイト文字は、J I S X 0 2 0 8を使用し、漢字については、J I S 第1水準漢字及びJ I S 第2水準漢字とする。

(2) 文字コード

文字コードについては、J I Sコード、シフトJ I Sコード又はE U Cコードを使用する。

3 外字

- (1) 外字については、可能な限り2に定める文字のうち任意の文字に置き換える。
- (2) 置き換えが困難な外字のうち、利用頻度が高いものについては、第3の2の(1)に定める標準DTDに定義する。標準DTDに定義されている当該外字については、必要に応じて電子化を行い、その場合は、標準DTDに定められたタグに置き換え、国際的な標準の形式によりイメージデータ化する。
- (3) 標準DTDに定義されていない外字については、標準DTDに定義を追加することができる。電子化する場合は、(2)に準じて取り扱う。

第3 文書構造形式

文書構造の形式を定める規格として、JIS X 4151 (SGML)を採用することとし、次に定めるSGML宣言及びDTDにより、SGML化を行う。

1 SGML宣言

(1) 標準SGML宣言

SGML宣言は、部会で了承する標準SGML宣言を使用する。

(2) 標準SGML宣言の定義変更

標準SGML宣言の定義変更については、基本文字集合又はアプリケーション情報の定義の変更を除き、行わない。

変更した標準SGML宣言は、ファイル転送等により他省庁に提供する。

2 DTD (文書型定義)

(1) 標準DTD

DTDは、部会で了承する標準DTDを使用する。

(2) 標準DTDの定義変更

標準DTDの定義の変更については、タグ又は外字の追加を除き、行わない。
変更した標準DTDは、ファイル転送等により他省庁に提供する。

3 SGML関連ツールの選定

SGMLに関連するツールの要件は次のとおりとし、各省庁において選定する。

- (1) 標準SGML宣言及び標準DTDに対応できること。
- (2) 各省庁において定義変更した標準SGML宣言及び標準DTDに対応できること。
- (3) HTML変換が行えるツールが利用できること。

第4 データベースシステム

データベースシステムについては、国際的な標準に対応したソフトウェア、ハードウェアを積極的に活用することにより、オープンなシステムを整備するものとし、当該システムが保有し、提供する機能は、次のとおりとする。

1 処理対象データ

データベースシステムは、次のデータに対応できることとする。

- (1) SGML宣言
- (2) DTD (文書型定義)
- (3) 文書インスタンス (文書実現値)
- (4) テキスト形式のデータ
- (5) イメージデータ
- (6) 動画像・音声データ
- (7) CSV形式、SYLK形式等の数値等データ

2 検索等機能

- (1) 次に掲げる検索機能の提供については、必須とする。

ア 特定された一つの白書等の目次から、任意の目次項目を選択して、該当する箇所を出力できる機能

イ 一つの任意語又は複数の任意語の組み合わせにより、特定された白書等から当該任意語を含む箇所を出力できる機能

- (2) その他の検索機能の提供については、任意とする。

- (3) シソーラス機能、検索結果件数の出力機能、ヘルプ機能等検索を支援する機能の提供については、任意とする。

3 ファイル転送機能

データの改ざん、棄損等の防止を確保しつつ、ファイル転送機能を提供する。

4 データ変換機能

SGML形式のデータをHTML形式のデータに変換する機能を保有する。

第5 霞が関WANとの連携

次により、霞が関WANとの連携を図り、データベースシステムが保有する情報を霞が関WANにおいて円滑に流通させる。

1 通信プロトコル

- (1) 霞が関WANのネットワークプロトコルであるTCP/IPに対応すること。

- (2) データベースシステムにアクセスするプロトコルは、HTTPとすること。

- (3) データベースシステムが保有するファイル転送機能において使用するプロトコルは、HTTP又はFTPとすること。

2 データの変換

SGML形式のデータは、HTML形式に変換した上で、霞が関WANに流通させる。

3 セキュリティ

他省庁の利用に供するデータベースシステムを自省庁LANに接続する場合には、当該システムをファイアウォールの外に配置するなど、自省庁のためのデータベースシステムのセキュリティを確保する。

また、データベースアクセスにおいては、特に必要とする場合を除き、利用者認証は行わない。

第6 出力システム

出力システムについては、データベースシステムが提供する機能を利用するため、次の機能を保有することとする。

- 1 HTML形式のデータに対応できること。
- 2 国際的な標準の形式により電子化されたイメージデータに対応できること。
- 3 国際的な標準の形式により電子化された動画像及び音声データに対応できること。
- 4 ファイル転送に対応できること。

第7 その他

- 1 本仕様は、情報通信技術の動向等により、適宜見直しを行う。
- 2 本仕様に基づく整備によって蓄積された成果を各省庁において積極的に活用することにより、各省庁における整備を一層推進する。
- 3 本仕様に基づく白書等データベースの省庁間利用を円滑かつ効率的に推進するため、別途、要領を策定する。
- 4 本仕様に基づいて白書等データベースを整備するに当たり、各省庁間で統一性、整合性を図るべき事項については、適宜、部会了承を行う。