公文書目録情報のデータベースモデル ~階層構造を持つ目録情報のリレーショナルデータベースでの実装~

大城 博光

はじめに

県公文書館は平成7年の開館と同時にコンピュータシステムを導入し、収蔵資料目録の作成や検索及び所在管理等をおこなって保存と閲覧サービスに応えてきた。開館から約4年が経過したが、これまでの業務経験を活かして、公文書館業務に最適なシステム像の設計をおこなってきた。

通常、設計時の成果となる設計書や仕様書には、決定に至るまでの経緯となる「どのような手法で設計したのか。」、「何故、そのような造りにしたのか。」等の改めてシステムの評価を行う際にもっとも重要となる内容がほとんど見えてこない。

そこで本文では、設計時におけるデータベースのモデル化に絞り、検討過程で使用したワークペーパーを 用いて、設計の根拠と手法を記述していきたいと思う。

1 リレーショナルデータベースの設計手順

リレーショナル型データベース(以下: RDBMS)は、貯える情報を行と列で構成される2次元の表で現し、その表を構成する要素(属性)を自由に組み合わせ、目的の形(仮想表)で表現することができる。全ての情報を1つの表で現すと同じ記述を何度もくりかえすことになるがRDBMSでは、複数の表を自由に結合、分解することも可能にしている。その為RDBMSの設計では表同士のリレーションシップ(相互関係)が効率のよいデータデースシステムを構築する重要なポイントになる。

一般的なRDBMSの設計手法は、実在する帳票類を実体表として捉え、それを正規化(*1)という便利な指針に従って最も無駄のない構成で、目的の帳票となる仮想表(*2)を生成していく方法である。

本設計では、既存データベースの性能向上という点もあり、仮想表の分析をはじめにおこなった。(*3)まず、業務プロセス毎に必要とする仮想表を洗い出し、その仮想表を構成する要素を記述の繰り返しが存在しないように関連する要素同士のグループ化をおこない、そのグループ間のリレーションシップを確立しデータベースモデルを構築していく手順をとった。

^{*1} 効率的なデータベースを設計する為の方法論とその基準。

^{*2} 仮想表には、印刷帳票だけでなくコンピュータ画面に出力される情報も含む。

^{*3} 一般的に、データベースの実装・稼働までには、実体表の正規化までの概念設計、業務プロセス毎の仮想表の検討までの論理設計、ハードウェア構成に準じておこう物理設計のステップでおこなう。物理設計においてはデータベースの専門知識が必要となるが、概念設計、論理設計までは業務を熟知していることが重要となる。概念設計でおこなう「正規化」は、きちんと手法と業務を理解していれば、誰がやってもほぼ同じ結論を導き出せ、データベースを構築したことがある人には自然と身についている技法である。しかし1度もデータベース設計を試みたことのない人には、正規化の目的である「繰り返しをなくす」作業がどういう意味を持つのかピンとこないであろう。そこで全専門員の共通認識を得やすい仮想表の設計から取り組むのが日々の業務と密接し、より理解しやすい設計手順を踏めると判断した。

2 仮想表の分析と構成要素のグループ化

データベース化すべき項目を洗い出す目的で、まず検索用目録の要素の検討をおこなった。

資料を探し出す目的で利用される目録は、検索者へどのような情報を与えるか、またその情報にどのような索引を設けどのように並べるかが重要である。検索用目録の要素の検討には、当館の既存の目録構成要素の見直しからおこない、それに対し他の資料保存機関の採録項目も調査し必要と思われる要素を追加した。(*4)また、検索用目録以外にも、公文書館業務では様々な帳票類が必要になる。そこで、システム改善要件(*5)を考慮し、各種業務の流れを図式化した業務フロー図(*6)を作成し、業務のどのプロセスでどのような帳票類を必要とするのかを分析した。

このような流れで検討された検索用目録や各種帳票類を仮想表として捉え、仮想表を構成する要素を関連する物同士グループ化した。(*7) その結果を図1に示す。

^(*4) 資料の受入先で、その資料目録が既にデジタルデータとして存在する場合は、その目録データを自館用目録様式に変換し利用することにより入力作業を軽減している。例えば県文書学事課の廃棄台帳データ、県行政情報センターの引渡データ、米国立公文書館のマイクロフィルム目録データ等である。このような他機関の目録様式との変換も考慮し汎用性を持たせる目的で、国際標準記録史料記述(ISAD-G)との対応も検証した。

^(*5) データベースモデルに影響する主なシステム改善要件

①目録の検索時において、キーワード検索に頼り切っているため、検索者が何らかの文字列を想定しなければいけない。また、想定した文字列に合致する情報が膨大な量の場合には絞り込みが面倒になる。

②目録検索の結果において、箱単位、簿冊単位、件単位の情報が混在して表示されるのがわかりづらい。

③ある特定した資料の重複物、複製物等の関連資料の情報検索ができない。

^(*6)業務フロー図では、最も効率的な作業の流れを検討し作り上げた。また、その際にコンピュータ処理として、 現在システム化されていなかった各種帳票類も仮想表として取り込んだ。

^(*7) グループ化の際には正規化の技法にも従った。

図 1 グループ要素と仮想表の構成

受入管理台 マイクロ撮影台 修 復 仮想表 カルテ

グループ	-	要素		日録		テ票		要素説明
クループ	01	受入コード			1,20	/		要素説明 受入毎の一連番号
受入情報		受入コート	0		-			受入時の一連番号 引渡:移管:寄贈:寄託:購入:複製作成:その他
		受入田付	0	0				対後: 移官: 計贈: 計記: 購入: 複聚作成: ての他 資料を受け入れた年月日
	03	受入日刊	-	0			L	
	04	受入量	0	U				寄贈者、移管先(文書学事課)等 受入単位の総量・規模
	05	受入資料	0					受入単位の総重・規模 受入資料群を総称する名称
	06	受入情報	0					
分類情報	07	分類コード	\cup				_	受入時の約束事や受入先が所有するオリジナル情報等
	08	資料分類		0	1550			
		資料分類解説	_	0				
		内容コード		$\overline{\bullet}$				 資料の内容を特定する記号
内容情報	11	タイトル						資料の主題
	12	サブタイトル		0	0	0		資料の副題
		出所/編・著者		-		0		
		出所/編・者石 資料日付(自)		0	0	-		出所または編集者、著者 本料の作成された日仕(白)
		資料日付(目)		0	0			資料の作成された日付(自) 資料の作成された日付(至)
	_	資料的 (主)		0	0		-	資料の作成された日刊(主) タイトルでは不足する内容に関する情報。
	17	資料解說 発行/出版		0	0			資料の発行/出版者
	18	<u> </u>		0	0			文字情報: 頁数 映像/音声情報: 時間
	19			0	0			文書,図書,地図,図面,空中写真,写真,新聞,映像,音声
	20	資料種別 言語種別		0	0			文音,図音,地図,図画,空中与具,与具,利面,映像,音戸 資料の記述に用いられている言語
	21 22	個人情報識別		0	0			無:有
	23	資料コード		0	•	•		記録媒体を特定する記号
形態情報	24	用途識別		0				閲覧用:保存用
	25	媒体種別		0	$\frac{1}{2}$	0		風鬼州・体行州 紙,マイクロフィルム,映写テープ,ビデオテープ,カセットテープ,CD
	26	媒体形態		0	0			媒体種別毎の媒体形態
	27	媒体表題	-					記録識別が1媒体複数内容の際に、媒体につけた名称
	28	記録日付				0		媒体に情報を記録した日付
	29	複製元資料コード			0	0		複製物の複製元資料コード
	30	検査日付				0		媒体の保存状態最終確認日
	31	媒体情報			-	0		媒体の再生、修復、整理、利用/公開制限理由・措置情報
保存情報	32	修復日付				0	_	媒体を劣化処置した日付
	33	修復措置情報				0		媒体の保存状態最終確認日
利用情報	34	閲覧回数			_	9		資料が閲覧された回数
	35	複写回数						資料が複写された回数
		利用識別						通常: 閲覧制限: 複写制限(保存状態から判断)
	37	公開識別						通常: 一部公開: 非公開 (資料内容から判断)
	38	公開目付						即時公開:公開制限解除西暦年度
	39	最終閲覧日付						資料の最終閲覧日付
所在情報	40	所在識別					-	責件の取称阅見ロリ 書庫:閲覧:館外:修復:複製:整理:展示:事務:編纂
		所在藏別						青卑: 見見: 貼外:
	41					-		取終の所任変更日13
	42	所在備考			-	2000		
	43	収納コード		0				収納容器記号
	44	書架コード		0		L		収納棚記号

- 注)1)本表では1部の仮想表のみ記載。
 2)●: 仮想表を一意に決定する要素(主キー)
 ○: 仮想表の構成要素
 ◎: 仮想表の構成要素
 ③: 仮想表の構成要素で繰り返される項目
 3)分類情報と内容情報においてはその後の設計で構造を再検討した。その構造については第4項「分類情報と内容情報の構成」で述べる。

3 表同士のリレーションシップとその検証

次にグループ化した受入情報:分類情報:内容情報:形態情報:利用情報:所在情報:保存情報を1つの表として捉え、それぞれの表の相互関係を分析した。結果を図2に示す。

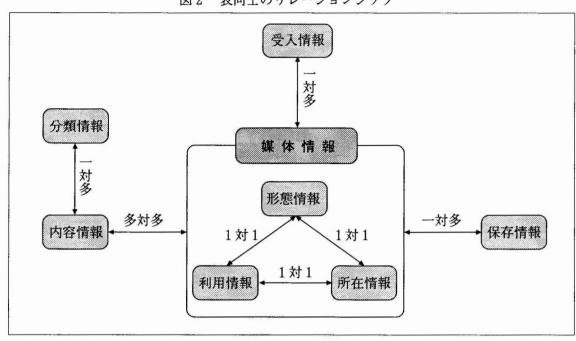


図2 表同士のリレーションシップ

①「一対一」(形態情報:利用情報:所在情報)

一対一の関係では、片方の表の1情報(行)に対して、もう片方のテーブルも1情報(行)だけ関連づけられる。この形態は、1つの表の要素が多い場合やセキュリティを保つなどの理由で、一つの事柄に関する情報を複数の表に分けて保存する場合などに使われる。

形態情報・利用情報・所在情報は1つの媒体を特定するとそれぞれの情報が一意に決定し、また前述した 理由もあてはまらないのであえて分割管理する必要性もないことから媒体情報として1つの表にした。

②「一対多」

一対多の関係では片方の表の1情報(行)に対して、もう片方の表は複数の情報が関連付けられる。

· 受入情報: 媒体情報

1つの受入先から複数の媒体を受け入れる為、1つの受入先が決まると複数の媒体が存在する。逆に1つの媒体には受入先は一意に決定する。

·媒体情報:保存情報

1つの媒体の保存情報(保存状態記録、修復措置記録等)は、検査または修復を実施する毎に複数の記録が存在する。逆に1つの保存情報は1媒体毎に管理される為、媒体は一意に決定する。

· 分類情報: 内容情報

分類情報は、公文書においては出所(主管部課)による分類をとっている。よって1つの分類には複数の 資料内容が存在し、また資料内容は1つの分類にあてはまる。

③「多対多」(内容情報:媒体情報)

多対多の関係は「一対一」「一対多」とは異なる特殊な形態である。2つの表が多対多の関係を持つ場合、RDBMSでは、この関係をそのまま実装することはできない。表同士のリレーションシップは必ず1対多

(または多対1)でなければいけない。しかし、資料は利用または保存の目的で複数の複製物が作成される。 また、刊行物においては、同じ物が何冊も重複して存在する。その場合、内容情報と媒体情報は多対多の関係をもつことになる。そのような場合に、多対多の関係を実装するには、2つの表に対して1対多の関係を もつ結合表を作成する必要がある。図3で示す記録アドレス表がその結合表になる。

内容コード 資料コード タイトル 資料種別 媒体種別 書架コード C001 人事関係書 文書 M001 冊子 A - 05C002 予算執行綴 文書 M002 A - 06冊子 C003 工事設計書 文書 M003 マイクロフィルム C - 02C004 工事図面 図面 M004 冊子 A - 07C005 会計報告書 刊行物 M005 冊子 A - 08内容コード 資料コード 記録アドレス C001 M001 C002 M002 C002 M003 0001コマ C003 M003 0050コマ C004 0100コマ M003 C005 M004 C005 M005 仮想表 (展開イメージ) 資料種別 資料コード 内容コード タイトル 記録アドレス 媒体種別 書架コード C001 人事関係書 文書 M001 紙 A - 05文書 A - 06C002 予算執行綴 M002 紙 C002 予算執行綴 文書 M003 0001コマ マイクロフィルム C - 02C003 工事設計書 文書 M003 0050コマ マイクロフィルム C - 02工事図面 図面 M003 0100コマ マイクロフィルム C - 02C004 A - 07会計報告書 M004 C005 刊行物 紙 A - 08C005 会計報告書 刊行物 M005 紙

図3 記録アドレス表による多対多関係表の結合

「内容コードと資料コード」

内容コードは、資料の内容を特定するコードで、同じ内容の資料には同じコードが用いられ、図書類の出版時に印刷されているISBNコード等と同様の機能を持つ。これはシステム内部での管理に用い表面的には意識されない。

これに対し資料コードは、その資料の保存や物流管理をおこなう為、それぞれの媒体自体に資料コードラベルを貼付し、媒体を特定するコードになっている。

「記録アドレス表の仕組み」

記録アドレス表は、内容コードと資料コード及び媒体に内容が記録されているアドレスの3つの項目で構成される。記録アドレス表により、内容情報と媒体情報をそのどちらからでも操作できることが可能になる。図4の例で説明すると、資料タイトル「予算執行級」を特定すると、その情報が記録されている2つの媒体(資料コード:M002,M003)の情報まで取得することができる。また、資料コードM002の媒体を特定すると

媒体には1コマ目に「予算執行綴」と50コマ目に「工事設計書」と100コマ目から「工事図面」が記録されていることがわかる。これはマイクロフィルムだけでなく、複数の資料の複製を1つの媒体に納める記録媒体(CD等)でも同様の管理が行える。また、情報の登録時においては、従来は記録した複数資料の総称をその媒体のタイトルにしていた為、内容情報が不足していたのに対し、その媒体自体の媒体情報を登録し複製元の情報を選択設定するだけで内容情報の設定が行えるうえ、媒体間の関係(正副)までシステム管理できるようになった。

4 分類情報と内容情報の構成

検索者が資料を特定する上で提供すべき情報と方式の検討を進めていき、分類情報を資料ガイドとシリーズ、内容情報をタイトルと内容情報に分割し、より整理された方式で情報を提供できるように改善した。

「資料ガイド」: 当館でどのような資料を収蔵しているのかを利用者に把握させる為、資料群を階層的に分類する。(*8) 機能的には、検索時のキーワードとの組み合わせまたはによる絞り込みやキーワードが想定できない時には目次的役割を果たす。

「シリーズ」: 資料ガイドの最下位レベル配下で、その資料群の個別の分類としている。公文書においては 事務分掌に基づいた仮想の標準フォルダーとなる。

「タイトルとサブタイトル」: 行政が日常の慣例業務から発生した文書を綴る簿冊は、逐次刊行物と同様に 年度や分冊記号を除くと殆ど同じタイトルになる。そこで資料の主題を示すタイトルと検索時の主題の特定 には直接結びつかない年度や分冊記号等のサブタイトルに分割し、同一タイトルでグループ化してサブタイ トルとの階層構造を持たすことにより、検索の応答時により整理された帳票形式で応答することができる。 上記で述べたそれぞれの表の関係を図4に示す。

レコードゲループ シリーズ 記 述レベル 述 アイテムグループ アイテム シリーズ 資 料 タイトル サブタイトル 沖縄県 総務部 ○○○関係書類 記要 述素 総務課 00000簿 No. 1 No. 2 人事課 △△△関係書類 $\triangle \triangle \triangle \triangle \triangle$ 1998年度 " 1999年度 " リレーショ 一对多 ·対多 シリーズ 資料ガイド タイトル 内容情報 ンシップ

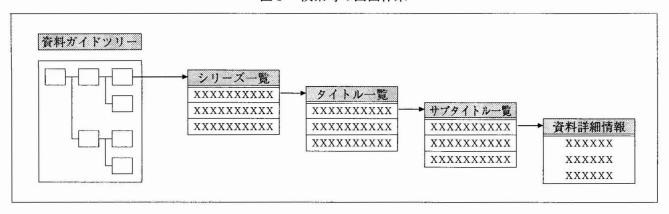
図4 資料内容に関する情報の構成

「検索時における画面体系 |

資料ガイド→シリーズ表→タイトル表→内容情報表は全て1対多の階層構造で結びついている。この階層構造を持つ資料目録を図5に示す階層的な画面体系で表現/応答する。

(*8) 資料ガイドは2次元の表で管理するが、最大10レベルの階層体系をアプリケーションで動的に表現する仕組み になっている。

図5 検索時の画面体系



設計を終えて。

行政にもリストラの波は押し寄せてきている。リストラクチャ(restructure:再構築)は人減らしばかりが手段ではない。例えばこれまで3人月かかっていた作業を1人月で済むよう業務の効率化をはかるのもリストラである。その際に夜通し作業していてもコストのかからないコンピュータを利用し、作業の効率を上げ、サービスを向上させるのは大きな意味を持つと思う。そういう意味で今回のシステム設計も稼働時にはリストラとしての成果があがることであろう。

最後に、21世紀のスタートに合わせ、ここで紹介した検索システムを'アーカス21' (Archas: Archives Administration System) という名称(仮)で当館のホームページでの稼働を予定している。このような設計を経て構築された検索サービスを実際に利用していただき、評価していただきたい。

今後はアーカス21はもとより、IT(Information Technology)時代における歴史情報発信機関のあるべき姿を模索し、そのモデル機関となれるよう目指していきたいと考えている。

(おおしろ ひろみつ:財団法人沖縄県文化振興会公文書管理部 公文書専門員)