

地域データベースモデルの実験的開発

An Experimental Development on the Model of Regional Data Base

伊藤陽介(*), 八重樫純樹(**)

(*) NEC ソフト株式会社静岡支社

(**) 静岡大学情報学部

Yosuke Ito(*) and Jyunki Yaegashi(**)

(*) Shizuoka Branch, NEC Soft, Ltd.

(**) Faculty of Informatics, Shizuoka University

[要旨] : 本論文はインターネットでの地域情報発信において、データ構造の違いやデータ間の不整合の問題が発生していることを問題として取り上げ、これらの問題を解決するためにXML技術を用いた地域データベースの構築を提案する。さらに浜松市の地域情報データをもとにデータのモデリングと地域データベース、地域データベースアプリケーションの構築を行う。この過程と地域データベース応用アプリケーションの評価をもとに、地域情報データベース構築における現在の問題点について考察を行う。

キーワード : 地域データベース, 横断検索, データ統合, XML, XSLT

[Abstract]: This article will show descriptions about problems of sending regional information in the internet, and propose an idea to solve these problems. This idea is development of a regional database using xml technology, then develop model of regional information, regional database based on regional information of Hamamatsu city and regional database application. Through these experiments and evaluation of regional database application, this article examines the current problems of development of regional database.

Keywords: Regional Database, Cross-Searching, Data Unification, XML, XSLT

1. 地域データベースとは

1.1 地域情報とは

現在、Web上で地域振興のためやデジタルアーカイブ事業の一環としてさまざまな地域情報が公開されている。そもそも地域とは「政治・経済・文化の上で、一定の特徴をもった空間の領域 (文献 (1))」であり地域情報とはその地域がどういうところかを示す情報、つまり「地域独自の過去から蓄積されたものが他の地域と何がどのように違うか (文献 (2))」を示す情報の

ことをいう。さらに、その地域に住む人のための行政サービス情報やイベント情報なども地域を説明する情報となるため、これらも地域情報のひとつであると定義できる。つまり地域情報とは地域を語るのに必要な情報、個性のことである。これら地域情報となりうるモノのことをまとめて地域資源と定義する。また本研究では「自治体」を地域の枠組みとしてとらえる。

1.2 地域情報発信の現状

現在、Web上で公開されている地域情報は各

機関が各々の用途に特化した形で公開を行っている。そのため、それぞれのデータ構造が違だけでなく同一の地域資源に関連する情報であっても相互参照が行えず、複数のWebページを探索しなければならないことが多い。その例を図1に示す。「A公民館」でのイベント情報が「イベント」ページに掲載されているとする。しかし「A公民館」の所在を知りたくても場所情報へのリンクが存在していない場合、「公共施設」のページから別途「A公民館」の情報をたどる必要がある。また、逆に「A公民館」でどのようなイベントが開かれているかを知りたい場合は「A公民館」のページではなく「イベント」ページを参照するといった手間がかかる。これはページの設計だけで解決できる問題ではなく、あらかじめデータを整理・管理しておく必要がある。

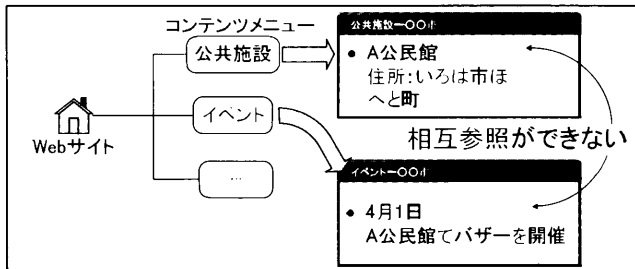


図1 相互参照ができない例

また、同一の地域資源に対する情報がページによって異なる場合があり、整合性がとれず閲覧者が混乱をまねく原因となっている。

1.3. 地域データベースの必要性

先に述べた現在の地域情報発信の問題点をまとめると以下のようなになる。

- ・ データ構造統一の概念が希薄
- ・ 同一の地域資源に対する情報がWeb上の各所に分散している
- ・ 地域資源同士の相互関係が未整理

これらの問題を解決するためには全ての地域情報を横断的に俯瞰し、整理・管理を行う仕組みが必要となる。そのためにはデータベースの構築が理想的である。地域情報をデータベース

化することでデータは一元管理され地域情報間の矛盾・重複がなくなり相互関連の把握が容易になる。本研究ではこれらを踏まえ浜松市をモデルに地域データベースを構築する。また、これを利用した応用システムの開発を行うことで地域データベースの利用法を示す。さらに、この過程を通して地域データベースの効果的な構築と利用に関しての研究を行う。

2. 地域データベースの構築

先の章で地域データベースの必要性について述べた。本章では、実際に地域データベースを作成するためにはどのような手法で構築すればよいか考察を行う。

2.1. データの統合方法

各機関にちらばる地域情報データをひとつの地域データベースとして活用するためにはデータの統合が必要となる。データを統合する手法にはデータを物理的に統合する方式「集中型」と仮想的に統合する方式「分散型」の2種類の方法がある。

- ・ 集中型クリアリングハウス方式

集中型クリアリングハウス方式(以下集中型)とはデータを物理的に1箇所で管理する方法である。統一されたスキーマを用いてデータを記述するためデータ管理が容易である一方、データのテーマが考慮されないためテーマの特色を失ってしまうという欠点がある。集中型による例を図2に示す。さらに、集中型はデータ収集の方法によって受動的クリアリングハウス方式と能動的クリアリング方式の2種類に大別できる。



図2 集中型

- ・ 受動的クリアリングハウス方式。

各機関があらかじめ決められたフォーマットに基づいてデータをクリアリングハウスに提供する方式。

例：デジタル岡山大百科（資料（1））、文化遺産オンライン（資料（2））

- ・ 能動的クリアリングハウス方式

クリアリングハウスが各機関のデータをトロールして台帳を構築する方式。

例：統合人文科学データベース（資料（3））

- ・ 分散型横断検索システム

分散型横断検索システム（以下分散型）とは横断検索用サーバを用いるかクライアント側でデータの加工編集を行って仮想的に統合する方法である。横断検索用サーバを用いる例を図3に示す。各データベースがテーマの特色を活かしたスキーマを採用することができるというメリットがあるが、検索対象となるデータベースすべてにアクセスし結果を取得するためネットワーク帯域の混雑などの影響を受けやすい。

分散型で主に用いられる技術がANSIによって規定されたZ39.50プロトコル（以下Z39.50）である。Z39.50により異なったシステムの複数のデータベース間においても1度の検索で検索結果を一括して得ることが可能になる。国内では図書館同士がお互いの所蔵図書館の横断検索を行う際に用いることが多い。

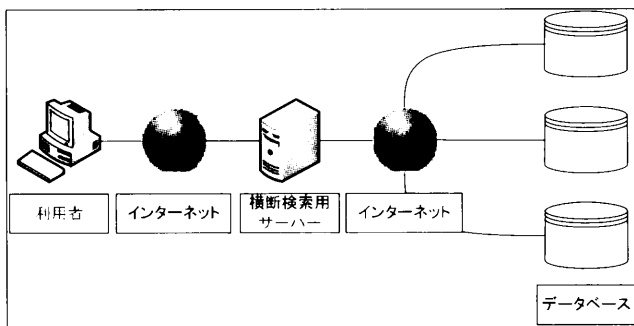


図3 分散型

2.2. 地域データベースの事例

現在、Web上で公開されている地域データベースの中でも岡山県立図書館のデジタル岡山大百科（文献（3）、資料（4））が提供している

「郷土情報ネットワーク」は収録データが4万件を超える国内随一の地域情報データベースとなっている（資料（5））。これは「岡山県のありとあらゆる情報をデータベース化し公開すること」を目的として積極的に地域情報を収集・公開しているためである。郷土情報ネットワークのデータベースは集中型で構成され以下の特徴を持つ。

- ・ データの管理

郷土情報ネットワークの情報管理にはメタデータ記述のための共通語彙であるDublin Core（文献（4））を用いている。図書データとも内部的にマッピングされ、これによってデジタル岡山大百科全体の横断検索が可能となっている。

- ・ データの分類

郷土情報ネットワークに登録されているデータはNDC（日本十進分類法）分類と岡山県独自の分類法である夢づくり分類・こども分類の3種類の分類法でそれぞれ分類されている。この情報を用いて利用者に応じた分類検索が可能になっている。

- ・ データの提供

データ提供の概念図を図4に示す。メタデータはWebコンテンツ作成者自身がデータを提供・登録することで受動的クリアリングハウス方式の運用となっている。これによりリンク切れの問題やデータの品質の問題も解消している。また、行政関係者だけに限らず岡山県民にも広く情報の提供を呼びかけている。

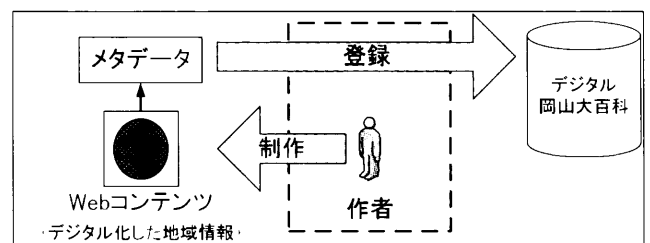


図4 デジタル岡山大百科のデータ提供

2.3. データベース構築法の決定

本章では地域データベースの構築に最適な手法を考察する。最初に考えるのは現状のWeb上

にあるデータをそのまま用いる手法である。これは分散型によって実現が可能である。しかし、同じ地域資源に対する情報が分散・重複して存在している現状でデータ毎にマッピングを行う作業の手間を考えると現実的ではない。また、地域情報間の関連情報の機能を実現するためには、関連を記憶しておくためのデータを別途作成し保存しておく必要があるためデータの管理が煩雑になってしまう。

次に集中型によるデータ統合について考察する。統合の際に用いるデータフォーマットには、新たなものを提案するのではなく、デジタル岡山大百科などで利用されているようなメタデータ概念とDublin Core語彙を利用する手法の方が比較的、実現が容易である。データ統合の際には、地域情報データをDublin Coreの元素にマッピングしていく必要があるが、歴史研究データベースにおける安達らの研究(文献(5))から分かるようにDublin Coreの元素の意味を拡大解釈したり元素を新たに追加したりしなければならない。解釈の仕方は人によって異なる可能性がありDublin Coreの元素にデータ作成者が意図していない情報、Dublin Core元素が想定していない情報を含むこともある。マッピングの作業を機械的に行うことができれば主観の入り込む余地を排除し、解釈のブレを減らすことが可能となるが、機械的に行うためには多種多様なフォーマットで記述されている地域情報それぞれのデータ構造を解析しなければならない。つまり地域情報のフォーマットが変更するたびに解析作業が必要となる。この問題を解決するために地域情報データをコンピュータの可読性が高いXMLで書き直す必要があると考えた。XMLで記述することでワンソースマルチユースという特性が利用可能となる。つまり地域情報データをXMLで作成し(ワンソース)その文書からメタデータを抽出(マルチユース)するという手法が採用できる。抽出処理にはXSLT(文献(6))のXMLのための文書構造変換言語を用いることが可能

である。この特性を活かし図5のようにデータをコンテンツの一部としてそのまま利用することが容易となるため、地域データベースの内容とコンテンツの内容を一貫性のあるものとすることができデータベースが地域情報発信の基盤として機能することが可能となる。

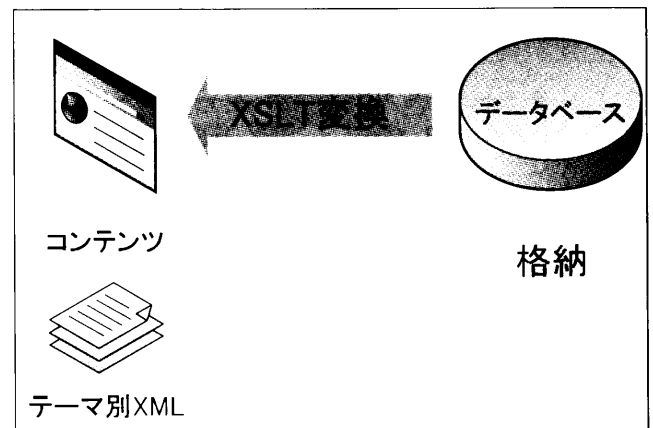


図5 地域情報発信の基盤としての利用

以上を踏まえ、具体的な記述方法を以下のものとする。

- ・テーマ別のデータ記述

XMLを用いて記述する。XMLを用いた表記は自由度が高いため、テーマごとにあらかじめ記述のルールを定めておく必要がある。これにはメタデータの表現方法であるRDF(Resource Description Framework)(文献(7))と特定の語彙を用いる場合、各種の団体等が定めた独自のフォーマットを用いる場合の2種類あるが適宜テーマに適した記述方法を採用する。複雑なデータはRDFと適切な語彙を用いて記述することで、データの意味を確実に解釈のブレが少なく記述することが可能となる。またRDFを用いるほど複雑ではないデータは記述に適したフォーマットを選定、使用する。

- ・メタデータの記述

データから自動的にメタデータを抽出する。また、メタデータの記述にDublin Coreを用いることで横断検索を実現する。

以上の手法をまとめ図6に示す。地域情報のデジタル化(XML化)と同時に統合用のメタデータが生成することができるので、別途メタデー

タを作成する手間を省くことができる。また、テーマ別のXMLを用いた詳細な検索とメタデータを用いた地域情報全体の横断検索、双方を実現することが容易となる。

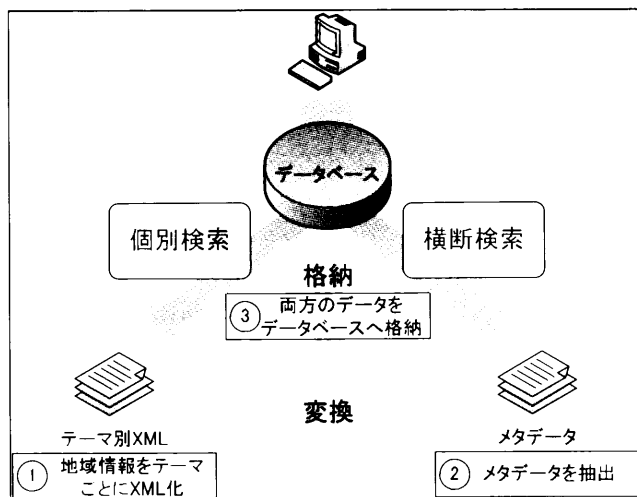


図 6 地域データベース構築モデル案

さらに、地域情報相互の関連情報の整理を実現するために地域情報データ記述の段階で地域情報相互の関連情報の記述を行う。メタデータにもそのデータを反映させることによって横断検索を用い、関連のある地域情報同士の相互閲覧が可能となる。

3. 地域情報データの作成

本章では先の章で決定した手法を用いて地域データベース構築法に従って地域情報データの作成を行う。

3.1. 使用した地域情報

地域情報データのデジタル化を行うにあたり本学においてデジタル化の作業が進められている浜松市の石造文化財に関連する情報と基本的な地域情報として4つのテーマをピックアップした。選んだテーマは「文化財」「目標物」「人物」「イベント」である。また、データは以下のものを参考にした。

○文化財

- ・石造文化財

「浜松市石造文化財所在目録」(文献(8))

- ・犀ヶ崖資料館天井絵

「犀ヶ崖資料館天井絵データベース」

(静岡大学情報学部によって調査・構築されたデータベース)

○目標物

- ・小中学校・公共施設

「浜松市」

(<http://www.city.hamamatsu.shizuoka.jp/>)

- ・高等学校

「まなびどっとこむ」

(<http://www.man-abi.com/>)

- ・大学

「静岡県」

(<http://www.pref.shizuoka.jp/>)

- ・神社

「静岡県神社庁」

(<http://shizuoka-jinjacho.or.jp/>)

- ・寺院

「iタウンページ」

(<http://itp.ne.jp/>)

○人物

- ・犀ヶ崖資料館天井絵の寄贈者

「犀ヶ崖資料館天井絵データベース」

- ・犀ヶ崖資料館天井絵の制作者

「犀ヶ崖資料館天井絵データベース」

○イベント

- ・公民館での催し情報

「浜松市」

(<http://www.city.hamamatsu.shizuoka.jp/>)

各データのIDはURI (Uniform Resource Identifier) (実際にはURIに含まれるURLを用いる) を用いて一元管理しこれを付与する。データとなる地域情報の公式Webページが存在する場合そのURLも併記する。

モデリングの結果に従ってデータをMicrosoft Excelに打ち込み、適宜改変した鈴木(文献(9))のVBAを実行しXML文書を生成する。以下にテーマ別のモデリングの過程を概説する。

3.2. 地域情報データのモデリング

3.2.1. 文化財情報のモデリング

文化財の中でも有形のものを主眼においてモデリングを行う。文化財はひとつひとつが多く、その情報は複雑である。これは文化財が数値化できる情報（重量や寸法など）と意味的な情報（文化財が制作された背景や目的など）を混在して保持しているためである。そのため、文化財情報をXMLで記述するためには複雑な情報が整理・表現が可能なモデルが必要となる。それに適していると考えられるのが国際ドキュメンテーション委員会 (CIDOC: International Committee for Documentation) の CIDOC-CRM (CRM: Conceptual Reference Model) である (文献 (10))。CRM は国際博物館会議 (ICOM: International Council of Museum) の中のCIDOCによって考案されたデータベース構築変換モデルで、各地に分散的に存在している既存のデータベースを利用してこれまで蓄積されてきた知識・データ・情報を地球規模で共有することを目的としている。既存の文化財データベースのスキーマはバラバラであるため結合を行いたい場合、相互のスキーマ情報を収集して動的で大域的なスキーマの構築が必要になる。この際、データベース間の差異を吸収するために属性と値との概念間の関係を定義しておく必要がある。このような定義関係をオントロジーと呼びCRMはこのオントロジーの考え方に基づいている。CRMの最新版は4.2版で、2006年9月にはISO 21127として国際標準規格に採用されている。CRMはデータの表現に関して特に定められていないが、記述にはRDFを使用することが提案されている。よってRDF/XMLを用いて記述が可能であるため文化財情報の記述にCRMを用いる。

・CRMによるモデリングの例 (資料 (6))

CRMでは対象領域とする領域で使われる概念をClass (クラス) と呼びその概念の属性や概念同士の関係をProperty (プロパティ) と呼

ぶ。概念には現実世界に対応する具体的な対象がある。例えば「伊藤陽介」はPersonというクラスの具体物である。この具体物をInstance (インスタンス) と言う。「草花鳥獣文小宝箱」を「松田権六」が作成したというモデルの例を図7に示す (資料(6)より引用)。工芸品である草花鳥獣文小宝箱はMan-Made Objectクラスのインスタンスであり、制作者である松田権六はPersonクラスのインスタンスである。両者を結びつけるのは「草花鳥獣文小宝箱の制作」というクラスである。また、「草花鳥獣文小宝箱」と「草花鳥獣文小宝箱の制作」はwas produced byというプロパティによって、「草花鳥獣文小宝箱の制作」と「松田権六」はcarried out byというプロパティで関連付けられる。括弧の中はプロパティを逆からたどった場合の名前である。CRMではこのようにクラスとプロパティを結び合わせていくことで意味のネットワークを構築していく。

・石造文化財のモデリング例

浜松市は10年ほど前に浜松市立博物館の主導で市内の石造文化財の調査を行い、これを「浜松市石造文化財所在目録」としてまとめた。この目録データをもとにCRMによるモデリングを行った。その結果が図8である。また凡例を図9に示す。

3.2.2. 目標物情報のモデリング

施設、遺跡など土地の上に存在している建築物などをまとめて「目標物」として扱いモデリングを行った。目標物情報の中で1番重要な情報は、その目標物がどこにあるのかということ、つまり位置情報である。RDFとCRMを用いても目標物情報の表現は可能であるが、保持すべき情報が文化財情報ほど複雑ではないため位置情報の扱いを主体としたXML記述のためのフォーマット、LBCS (Location Based Content and Service) を用いることとした。LBCSとは財団法人日本情報処理開発協会 データベース振興センターによって開発・策定されているフォーマット

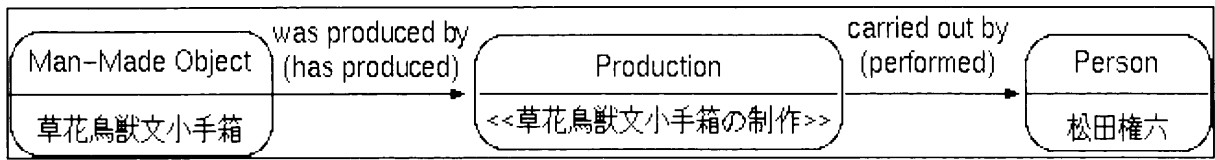


図 7 CRM モデルの例

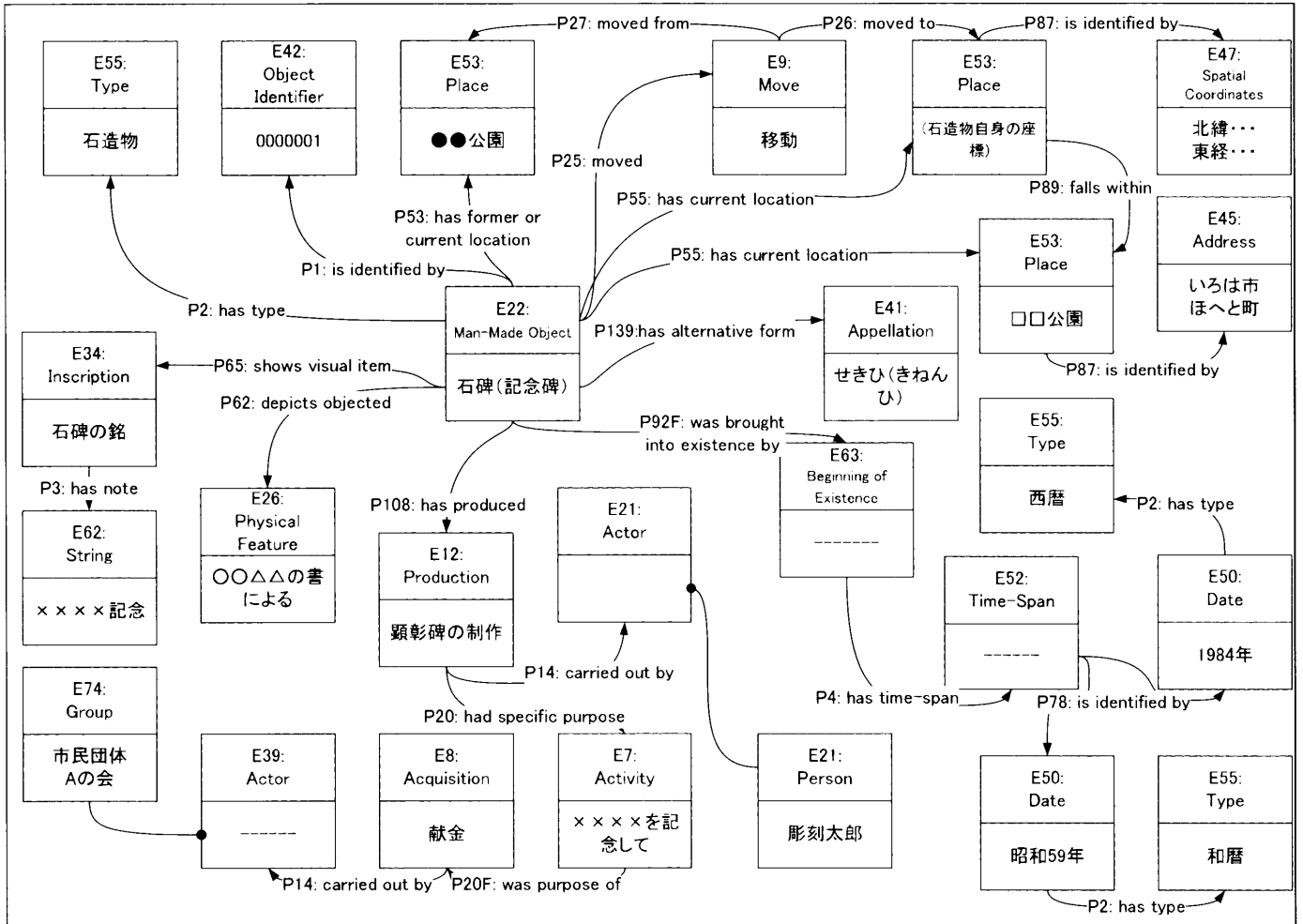


図 8 石造文化財モデル

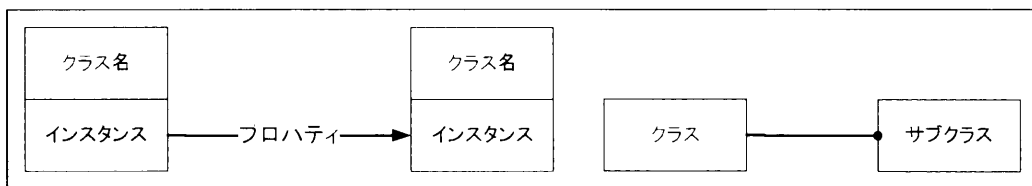


図 9 CRM モデリング図の凡例

トである (資料 (7))。「電子データに位置情報を付けることによって地図を含む実世界空間に關係付けられたデジタルコンテンツとして活用し、利用者に、そのデジタルコンテンツと連携する、位置情報に関連したサービスを提供する IT の応用領域の一つ」とされ「XML に基づいて符号化する形式で規定する (文献 (11))」とされている。図 10 は LBCS による記述の例である。

各要素の解説

- ・ poi 要素: デジタルコンテンツへのリンクのために必要な情報を記述する。
- ・ identifier 要素: 親要素に対するグローバルに有効な識別子を記述。
- ・ title 要素: 親要素に対する名称を記述。
- ・ description 要素: 親要素に対する補足的な内容を記述
- ・ category 要素: 親要素のカテゴリ情報を記

```

<?xml version="1.0" ?>
- <lbc:lbc xmlns:lbc="http://www.dpc.or.jp/lbc">
- <lbc:poi>
  <lbc:identifier
    lbc:uri="http://titan.ia.inf.shizuoka.ac.jp/digital_encyclopedia/japan/shizuoka/hamamatsu/lbc/uni_0001" />
  <lbc:identifier lbc:uri="http://www.shizuoka.ac.jp/" />
  <lbc:title lbc:name="静岡大学" lbc:ruby="しずおかだいがく" />
  <lbc:place lbc:placeId="静岡県浜松市中区城北三丁目5-1" />
  <lbc:description lbc:content="" />
  <lbc:category lbc:categoryName="教育機関"
    lbc:type="http://titan.ia.inf.shizuoka.ac.jp/digi_hamamatsu/dictionary/landmark" />
  <lbc:category lbc:categoryName="大学"
    lbc:type="http://titan.ia.inf.shizuoka.ac.jp/digi_hamamatsu/dictionary/landmark" />
- <lbc:propertyGroup>
  <lbc:title lbc:name="データ諸元" />
  <lbc:property lbc:name="データ元" lbc:value="浜松市" />
</lbc:propertyGroup>
</lbc:poi>
</lbc:lbc>

```

図 10 LBCS データ記述例

述

- ・ place 要素：デジタルコンテンツの位置情報を記述
- ・ property 要素：さまざまな属性値を記述

3.2.3. 人物情報のモデリング

人物情報の表現には FOAF（文献（12））という語彙が存在する。FOAFとは RDF/XML を利用して人々とそのつながりを公開・共有しようというプロジェクトである。

しかし、人間のネットワークの表現を目的としたプロジェクトであるため、人物データベースの観点から見た場合、語彙が不十分であると思われる。例えば名前の「ふりがな」に相当する語彙が存在しないため、別途語彙を追加・拡張する必要がある。本研究ではできる限り既存の語彙を利用することをスタンスとしているので FOAF ではなく CRM の語彙を利用することとした。CRM には人物に関してある程度の語彙が用意されているため本研究で構築するデータベース程度なら問題なく表現が可能である。モデリングの結果が図 11 である。これをベースとして人物にまつわる情報を適宜、追加していく。

3.2.4. イベント情報のモデリング

イベント情報ではお祭りやコンサートだけでなく公民館の〇〇教室や博物館の特別展など

地域で発生するイベント全体をとりあげて表現する。イベント情報の発信に多く利用されるのが RSS である。RSS は Web ページの見出しや更新日時などのメタデータを構造化して記述するための XML ベースのフォーマットである。また、RSS に基づいた文書を RSS フィードと呼ぶ。RSS はバージョンによって文法などが異なるが、本研究では RSS 1.0 を用いる。これは、このバージョンが Dublin Core による拡張を想定して定義されているためである。ニュースの配信や Blog の記事情報の配信に使われることが多い RSS だが、福井県の「おでかけふくい」（資料（8））のように自治体レベルでイベント情報の発信のツールとして使用している例もある。しかし、イベント情報の構造化表現には RSS だけでは不十分なので神崎（資料（9））の提案を参考にして RSS の item データのみを記述する。RSS 配信の際には Web サーバ上で必要に応じて item データを収集・加工を行うものとする。

3.3. メタデータの生成

地域情報をデジタル化した XML から XSLT を用いてメタデータを抽出して利用する。メタデータ表現の語彙には Dublin Core を用い、RDF/XML によって記述する。

- ・ Contributor エレメントの記述

情報資源の所在地情報をここであらわす

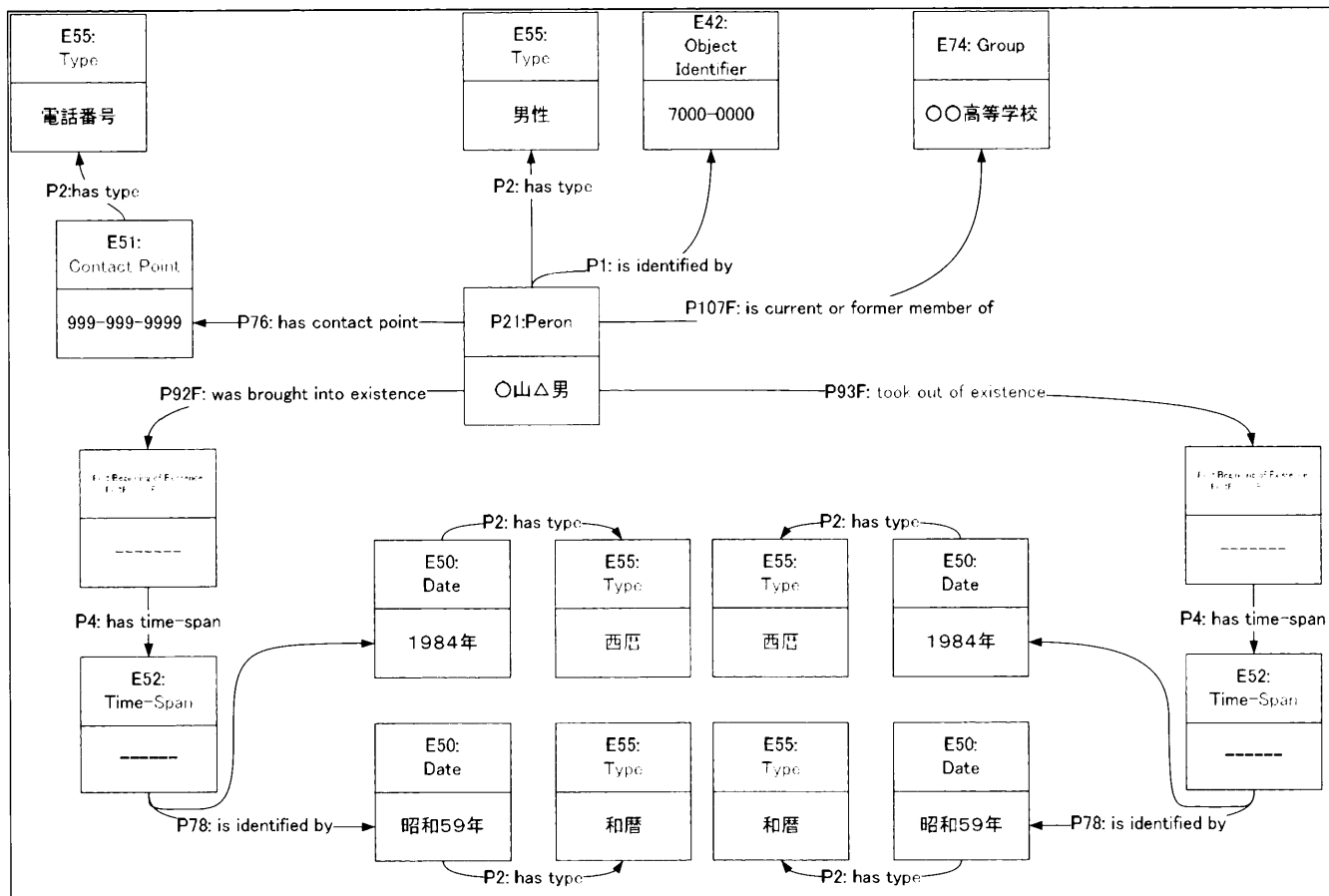


図 11 人物情報モデル

ものとする。Contributor エレメントは所在地情報を直接表す語彙ではないが、リソースのもととなる情報資源の協力者であると解釈した。

- Subject エレメントの記述

リソースのテーマを示す Subject エレメントを記述には独自の分類法に基づいた語彙を用いた。1件のメタデータに記述する Subject エレメントの数に上限は設けない。必要ならばいくつでも記述することができるため柔軟にテーマを表現できる。これによって「〇〇高校中等部」のようなデータを「高等学校」「中学校」の2つに分ける必要がなくなり1件のデータに「高等学校」「中学校」の2つの Subject エレメントを付与することで表現することが可能となる。

- Relation エレメントの記述について

各メタデータ間の関連を示すために利用する。例えば犀ヶ崖天井絵の場合、1枚の天井絵に制作者と寄贈者が存在している。こ

の情報を表現するために天井絵のメタデータでは制作者と寄贈者の URI、制作者・寄贈者のメタデータでは天井絵の URI を保持する。これによって各メタデータ間の関連情報の参照が可能となり相互をたどることが可能となる。

4. 地域データベース応用システムの実装

制作したデータをどのように保持・管理するかを決定する必要がある。データは XML で制作されているため、これを保持・管理するためには XML 文書に対応したリレーショナルデータベースか XML 文書をそのまま格納・管理することが可能なネイティブ XML データベースのどちらかを利用することとなる。しかし、検索応答速度といったパフォーマンスの問題や XML のワンスマルチユース特性をそのまま利用したいという考えからネイティブ XML データベースを利用した。そのため、データの問い合わせに

