

ネットワークを利用するデジタルアーカイブ共同構築システム Network-based Collaborative System for Digital Archives

池崎 友博†
Tomohiro Ikezaki

ドミニク ダフ†
Dominique Deuff

小野 欽司†
Kinji Ono

北本 朝展†
Asanobu Kitamoto

大西 磨希子†
Makiko Onishi

1. はじめに

本論文では、文化遺産関連の様々な分野の研究者がデータの入力から公開までを行える Web データ管理システム (ASPICO) の開発について報告する。本システムは文化遺産情報をオントロジー、エンティティ、リソースという3つの階層で表現し、それらの関係を RDF (Resource Description Framework) [1]によって記述するという設計に基づいており、部分画像へのメタデータ付与やエンティティの地図へのマッピング等の特徴的な機能を備えている。また、RDF の意味情報付リンクの性質を生かした検索や、Web の特徴を生かした遠隔/共同作業、一般公開、教育への利用、さらには異分野のデータ間の関連から新しい学問的知見を得ることなどを狙っている。

2. データモデル

本研究では、文化遺産関連のデータを取り扱う上で効果的なデータモデルについて検討を重ねてきた。以下ではそのデータモデルの説明と、このデータモデルを文化遺産データにどう適用し、それによって具体的にどのような効果が得られるかを述べる。

2.1. 3層の構造

データモデルを検討するにあたっては主に画像データを対象にしたが、被写体である文化財は多種多様であることから、実際の文化財を分析/モデリングしたものを画像データとリンクしていく方法を採用した。これは、従来の「画像+メタデータ」型の管理方法を拡張したもので、画像データをリソース層、文化財をエンティティ層、その概念をオントロジー層にそれぞれ記述し、それらを関連付けていくことにより画像データを管理していく方法である。

- ・リソース層：本システムにアップロードされたファイルをリソースと呼ぶ。リソースは主に媒体についての情報を記述した一つのメタデータを持ち、リソースの内容を表す一つ以上のエンティティとリンクする。また、データの部分を表すリソースは親リソースともリンクする。
- ・エンティティ層：実世界の対象を、対応するクラスの定義に従って記述したものをエンティティとする。一つ以上のクラスに所属し、全てのクラスに共通の標準属性とそのクラスに応じた専門属性を持つ。
- ・オントロジー層：概念を表すクラス間の関係（現在は抽象・具象関係に限定）を記述したものである。また、クラス毎にその概念に応じた属性を定義している。基本的なクラスを定義した標準部分とその下位クラスとして分野ごとに専門的なクラスを定義した専門部分から成る。

以下にクラス、エンティティ、リソースの関係を示す(図1)。

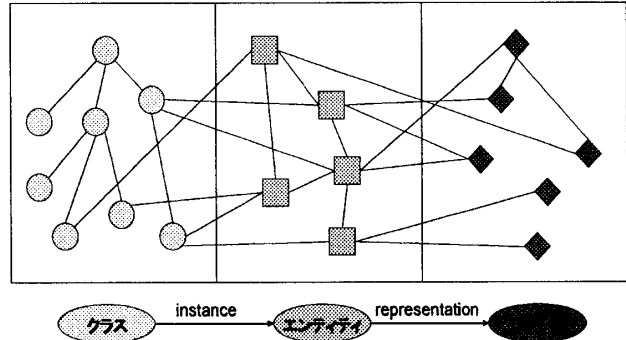


図1 クラス、エンティティ、リソースの関係

2.2. データモデルの適用

ここでは主に上述したデータモデルで画像データを管理していく際に得られる効果について述べる。

- (1) 画像を管理、公開する上では、様々な観点での分類や検索が可能であることが求められる。例えば絵画ならば、画家の出身地での分類や絵の題材での検索を指す。この例を上述のデータモデルに当てはめるには、デジタル化された絵画の画像データをリソースとし、そのメタデータには、どのようにデジタル化したかなど、主に媒体に関する情報を記述する。そして、その絵画の実物を表す絵画クラスのエンティティを作りリソースとリンクする。さらに、その絵画を分析し、画家情報や絵の題材情報などユーザが言及したい単位で、オントロジーによって意味付けされたエンティティを作成していき、従来は属性値として文字列で記述していた情報をこれらのエンティティへのリンクとして記述する。

このようにモデル化することで、画像データはエンティティ単位で明示的に分類されることになる。また、画家の情報を用いての検索、例えば「この絵画を描いた画家と同派の画家の作品を一覧する」など、エンティティ同士の関連を生かした検索が可能となる。
- (2) 画像は画像上の位置によっても異なる意味の要素を持ち、その要素単位での検索や分類が必要である。これは例えば、あるストーリーの場面を描いた絵の組み合わせが一つの壁画を構成しているような場合を指す。このような場合、画像全体と部分ごとの画像をそれぞれリソースとして取り扱うことによって、各々を上記(1)で述べたようにモデリングすることができる。これによりストーリーをキーに壁画の一部を取り出すような検索が可能となる。
- (3) 上記(1)(2)では分類対象になる要素を細分化したが、さらにこれらの要素の一つ一つは様々な視点での分類が求められる。例えば、銀の皿を考えると、金属製品として見た場合や食器として見た場合など、視点によ

って様々な分類が行えることが分かる。この例は、銀の皿を、皿クラスと銀製品クラスに属するエンティティとすることで、食器、金属製品の両面から検索や分類が行えるようになる。

以上、ここでは主に画像データについて取り上げたが、テキストデータなどの他の媒体のデータも同様にリソースとして取り扱うことができる。これらの特徴を総合すると、ある特別な装飾模様で検索し、壁画や器具、建造物の部分的な装飾や、その装飾模様についての論文などを一括して出力する、というような異分野/多媒体のデータを関連付けて取り出すことが容易になる。

3. ASPICO の概要

ASPICO は、様々な分野の研究者が遠隔/共同でデータの入力から公開までを行えるデータ管理システムを目指し開発を進めてきた[2]。ここでは、上述したデータモデルを用いてシステムを構築する際の、その記述方法とシステムの概要について述べる。

3.1. RDF による記述

データモデルの記述には RDF を用いた。RDF は URI による名前付けの仕組みを持ち、共通するノード（本データモデルではエンティティを指す）を介して情報を連結していけるような構造である[1][3]。RDF のこの特徴は共同作業においてユーザ間でのデータの統合などに最適であると考えた。また、処理のための基本的なライブラリが存在すること、他のアプリケーションでのデータの利用が期待できることも RDF を採用した理由である。

さらに、RDF の検索言語（SPARQL[4]など）を使うことで 2.2 で示したような関連を利用した検索も可能となる。

3.2. ワークスペース

本システムはパスワード認証で保護されたパーソナルスペースとグループスペース（開発中）で構成される。パーソナルスペースは個人のデータ管理を行い、グループスペースはディスカッションを行いながら、データの共同構築をしていくスペースである。また、データの公開に向けたデータの検証作業なども行う（3.4 参照）。

3.3. ツール

上記のワークスペースはそれぞれ以下の3つのツール群で構成される。これらのツールは多言語による記述が可能である。

- ・オントロジーツール群：オントロジーの編集、管理を行うツール群（開発予定）。現在、オントロジーに関してはオントロジー開発ツール Protégé[5]を使って編集したものを取り込んで使用している。
- ・エンティティツール群：エンティティの作成、削除等の基本的な管理ツールに加えて、エンティティの地図上での位置を表示するツールを含む。
- ・リソースツール群：ローカルファイルのアップロード、メタデータ付与、検索ツール、部分画像へのメタデータ付与を行うツール（図2）などから成る。

3.4. データ公開までの流れ

本システムでリソースとエンティティは `private`、`group`、`public` の3つの状態を持つ。パーソナルスペースで作成されたデータは `private` の状態を持ち、ユーザが編集を終え、そのデータの状態を `group` にすることでグループスペースに提出される。グループスペースでは公開のために内容

の検証や他のデータとの調整、データのどの部分をどのように公開するかを決定する作業を行い、それが完了するとデータの状態は `public` となり、外部からのアクセスが可能な状態となる。

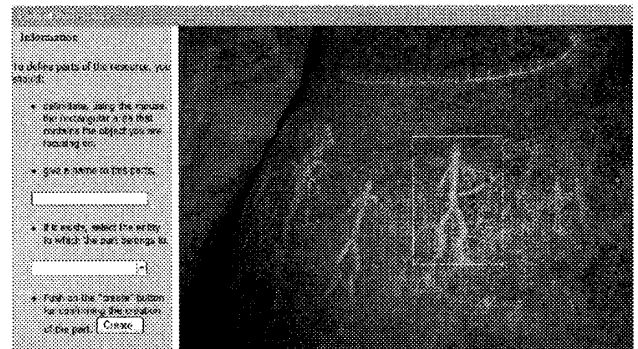


図2 部分画像へのメタデータ付与ツール

4. おわりに

本論文では、画像データをより詳細に説明、分類するために考案したデータモデルと、それを用いたデータ管理システムについて説明した。

現在、本システムはパーソナルスペースまで実装を終えており、人文科学系の研究者に実際にデータを入力してもらい意見を聞くことで改良を重ねている。

今後の課題としては、2.2 で挙げたような複雑な検索をどのようにユーザに提供していけるか、主にユーザインターフェイスの面から検討していく必要がある。

また、実際のデータ入力テストの際、クラスが具象的になるに従ってユーザがクラスとエンティティを正しく区別しにくくなる傾向が見られた。これはユーザインターフェイスの工夫によって対処できるものと思われるので今後検討していく。

さらに、「ダイヤモンド」のような個体として識別できない記述単位を扱う方法について、これはオントロジーとも密接に関係しているので、オントロジーツール群の整備と並行して検討、実装していく予定である。

そのほか、文化遺産情報を扱う上で時間・空間情報は重要であるので、これらをエンティティとしてどう効果的に扱えるか検討中である。

参考文献

- [1] G. Klyne, J. Carroll "Resource Description Framework (RDF): Concepts and Abstract Syntax", W3C Recommendation 10 February 2004, <http://www.w3.org/TR/rdf-concepts/>
- [2] K. Ono, T. Yamamoto, T. Kamiuchi, F. Andres, A. Kitamoto, S. Sato, E. Andaroodi, "Progress of the Digital Silk Roads Project," Progress in Informatics, Number 1, pp. 93-141, March 2005.
- [3] 神崎 正英, 「セマンティック・ウェブのための RDF / OWL 入門」, 森北出版, 2005
- [4] E. Prud'hommeaux, A. Seaborne, "SPARQL Query Language for RDF", W3C Candidate Recommendation 6 April 2006, <http://www.w3.org/TR/rdf-sparql-query/>
- [5] Protégé, <http://protege.stanford.edu/>