

分野連携による時空間データ構築の試み

関野 樹[†]

様々な分野の情報を、時空間情報を接点として連携させてゆくための課題が検討された。時空間解析で用いるデータについては、基準となる時空間情報の種類と記述方法をどのように統一するかが問題である。また、分野間でのデータ共有を進めるためには、各分野のデータベースが連携してクリアリングハウス機能などを提供する必要があり、そのための既存のメタデータ規格の取り扱いに工夫が必要である。さらに今後は、分野間での概念や言葉の違いを吸収するためのオントロジー技術の活用が期待される。

Trial construction of geo-temporal data in interdisciplinary studies

Tatsuki Sekino[†]

Requirements to information analysis based on geo-temporal data in interdisciplinary studies were considered. Appropriate selection of type (datum, coordinate system and calendar) and format of geo-temporal data is important for effective usage of the data. It is necessary for data sharing to provide clearinghouse functions cooperating with various types of database. It is expected that ontology technology will be applied to reduce difference in concepts and in semantics of terms among disciplines.

1. はじめに

1.1 分野連携と情報

分野連携、文理融合、学際、様々なかたちで学問分野の壁を越えた連携が模索されている。これらの中には、連携そのものを起爆剤として新たなシーズを見出そうとするものもある。一方で、地域研究のように初めから複数の学問分野の知識を総動員することが織り込まれている場合もある。特定の地域の現象に様々な角度から光を当て、そこから見えてくる事象を使いながら総合的に地域を理解しようとするものである。

そのための情報学的な手段を提供する「地域情報学」も生まれつつある。また、環境問題に関わる諸課題も多く、学問分野の知識を必要とする。環境問題の表面的な現象は、自然科学を中心とした学問分野の手段により数値化され、理解、共有することが可能である。しかしながら、その環境問題の根源には人間活動が複雑に係わっており、社会科学、人文科学諸分野の知識と手段が必要である。

1.2 接点としての時空間

複数の分野の知識や情報を結びつける接点として用いられるのが主題、時間そして空間である。主題を接点として情報を結びつける手段としては、広く普及しているキーワードを使った検索が実質的にその役割を一部を担っている。特定のキーワード(主題)に関する情報を多数の情報資源の中から見つけ出す方法であり、その検索結果は指定されたキーワードにより結び付けられた一連の情報群とみることもできる。空間情報に関しては地理情報システム(GIS)が近年その手段として躍進している。地図上に示された様々な情報は、場所を接点として結び付けられた情報である。我々はこのGISを用いることにより、地図という視覚的な方法で場所に基づく情報同士の関連性を確認することができる。では時間はどうだろうか。時間を接点としながら情報を整理する手法として古くから年表が用いられてきた。これは空間情報における地図と同様に情報同士の時間に基づく関係を視覚化する有効な手法である。また、折れ線グラフ、棒グラフなどの図表も数値化された事象の経時的な変化を表現する手段として広く用いられている。このような年表やグラフを生成するソフトウェアはそれぞれ数多くが世に出されている。ところが、時間情報に基づく情報同士の関係解析をGISのような視覚的な方法で行う解析ツールとなると皆無に等しい。ESRI社のArcGIS Tracking Analystなど、時間と空間を同時に扱うことを可能とするソフトウェアも普及し始めているが、これらもやはり「時間情報も扱えるGIS」であり、時間情報は脇役である。

この点で、有志の研究会であるH-GIS研究会が中心となり、人間文化研究機構(大学共同利用機関法人)の研究資源共有化事業のなかで作り出したGT-Map/Time(HuMap/HuTime)は、時間情報の本格的な解析機能(GT-Time)と空間情報の解析機能(GT-Map)を連携させることができる[1]。このため、GT-Mapで地理的に絞り込んだデータをGT-Timeで時間的に解析するといった、時空間という視点で情報同士の関係を解析してゆくことを可能である。しかしながら、解析ツールという点では環境が整いつつあるものの、依然として、さまざまな分野の情報を時空間の視点で1つのプラットフォームに統合し、関連性を見出そうとすることは、時空間の表現方法やメタデータの構造など、それぞれの分野の時空間情報の扱いの違いにより必ずしも容易ではない状況にある。本報告では、時空間データに基づいて異なる分野の情報を連携させようとした時に起きる問題点や課題を整理し、その解決への取り組み事例を紹介する。

[†] 総合地球環境学研究所 〒603-8047 京都市北区上賀茂本山 457-4
Research Institute for Humanity and Nature, 457-4 Motoyama Kamigamo, Kita-ku, Kyoto 603-8047 Japan

2. 時空間データの実例

まず、実際に GT-Time を使って行われた研究事例の流れを見ながら、複数の研究分野にまたがるデータを時空間情報に基づいて連携させる過程を検証する。なおこの題材は、滋賀県の自然災害に関するデータを時間情報に基づいて解析したものである。

2.1 データ作成

解析に用いられたデータは、気象庁が提供する彦根地方気象台の気象観測データ（1901 年～）[2]、滋賀県が県内で起こった災害の概要をまとめた『滋賀県災害誌』（第三部および第四部－1978～1998 年）[3][4]、滋賀県琵琶湖研究所が県内の出来事を年表としてまとめた『琵琶湖年表』（1982～2002 年）[5]である。気象観測データは気象庁の Web ページで公表されているものであり、HTML のテーブルとして提供されている。一方、『滋賀県災害誌』と『琵琶湖年表』は紙媒体しか提供されていないため、それぞれ、OCR 等を用いてデジタル化された。

GT-Time 用のデータは XML で表現される。（図 1）。それぞれのレコードは record 要素として表され、時間情報（datetime 要素）、空間情報（location 要素）、記事（event 要素）を含んでいる。時間情報は期間の始めと終わりを西暦で入力する。また、空間情報は UTM 座標を入力する。event 要素には、GT-Time の年表上に表示される title 要素、記事の詳細を示す detail 要素などが含まれる。データセットに関するメタ情報は別のファイル（tmm ファイル）で管理されており、GT-Time の場合はこの tmm ファイルの情報に基づいて、年表表示とグラフ表示の切り替え、表示範囲やデータセット（レイヤ）の名称の設定が行われる。

この GT-Time 用のデータ形式にそれぞれのデータを変換してゆくことになるが、ここで変換を行う対象期間が問題となった。複数のデータを組み合わせようとするのであれば、それぞれのデータがオーバーラップする期間のみを対象にすればよく、全ての期間のデータを変換してしまうのは解析時に不必要な処理速度の低下を招く恐れがある。また、今回用いたデータを別のアプリケーションで利用することも想定されるため、全てのデータを GT-Time に特化したデータ形式に変換してしまうのは好ましくない。このため、今回は対象となる全てのデータをいったん汎用性の高い CSV 形式（RFC4180 に準拠）にして整理・保管、その後、必要に応じて CSV ファイル内の必要なデータを必要な期間だけ専用のソフトウェアを用いて GT-Time 用の XML データに変換し、その XML データに対応する tmm ファイルを生成する仕組みとした。この CSV ファイルを生成する過程で、日付や場所をそれぞれ西暦や緯度経度に統一する処

```
<record>
  <datetime>
    <from>1923-09-01T11:58:32+09:00</from>
    <to>1923-09-01T11:58:32+09:00</to>
  </datetime>
  <location>
    <point>
      <zone>54</zone>
      <x>327267</x>
      <y>3907580</y>
    </point>
  </location>
  <event>
    <item name="title">関東大震災</item>
    <item name="number">0</item>
    <item name="detail">死者・行方不明者 105,000 人；全壊家屋 109,000 戸；
      半壊家屋 102,000 戸；焼失家屋 212,000 戸</item>
  </event>
</record>
```

図 1 GT-Time 用のデータの例

理が行われた。また、時空間情報の書式（日付であれば 7/25, 7.25, 7・25 の違いなど）を一定の範囲内に揃える処理もこの CSV ファイルを生成する過程で行われた。さらに、日付の表記を GT-Time で用いられる書式に統一する処理や緯度座標から UTM 座標に換算する処理は、GT-Time 用の XML ファイルを専用のソフトウェアで生成する過程で行われた。

2.2 データ解析

図 2 は GT-Time を使って「雨と風のある日は滋賀県内でどのような災害起きているだろうか」という観点で解析を試みた例である。ここでは、雨の条件として日降水量 10mm 以上、風の条件として平均風速が 5m/s 以上を設定した。まず、日降水量 10mm 以上、平均風速が 5m/s 以上の日をそれぞれ抽出した。この結果が図 2 の上から 4 番目と 5 番目のレイヤであり、それぞれ日降水量 10mm 以上の日、平均風速が 5m/s 以上の日を示している。対象となった期間は災害誌に記録のある 1978 年 6 月 21 日から 1998 年 10 月 18 日の 7,375 日で、そのうち日降水量 10mm 以上の日は 1,066 日、平均風速 5m/s 以上の日は 629 日であった。検証しようとする条件は、これらの風と降水量の条件を同時に満たす日であるから、この 2 つ結果の論理演算（論理積）により両方の条件を満たす日を確定した。この論理演算により生成された日が図 2 の 6 番目（下から 2 番目）のレイヤである。ここで示されている期間が、条件である日降水量 10mm 以

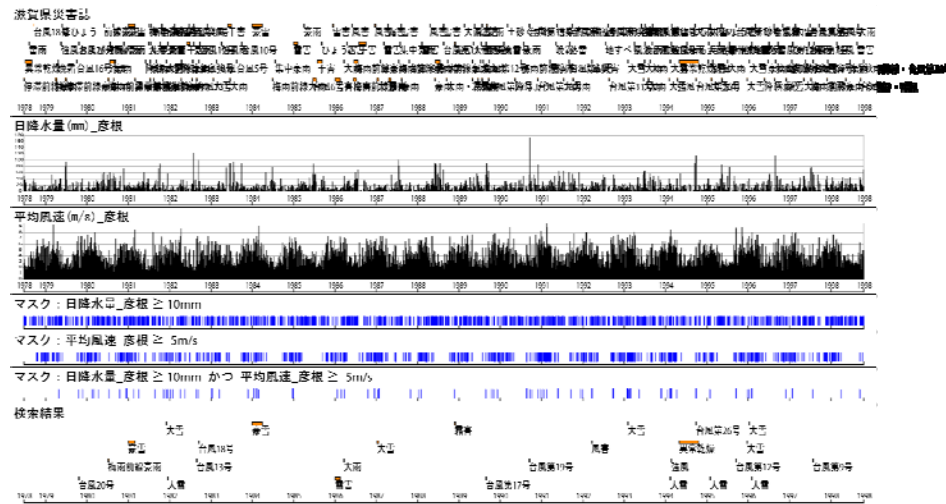


図2 GT-Timeによる解析例。上の3つの年表およびグラフは元データ（滋賀県災害誌、彦根の日降水量、彦根の平均風速）。その下3つはマスキレイヤーで一定の条件を満たす期間を示す。最も下の年表は、滋賀県災害史から条件を満たす日のデータを抜き出したもの。

上、平均風速が 5m/s 以上の日（雨と風のある日）であり、81 日がこの条件に該当した。最後に、最終的に出力された「雨と風のある日」に該当する災害の記録を滋賀県災害誌から抽出した。この結果、26 件の災害情報が抽出された（図2、最下段）。最も多かったのは雪害で、12 件であった。その次が台風で8件、さらに大雨と強風がそれぞれ2件、霜害と乾燥がそれぞれ1件と続く。一般に、雨と風という条件では台風を想像されることが多いと思われるが、滋賀県の場合は雪害が最も多いという結果であった。

では、このよう気象条件の日に県内ではどのような出来事が記録されているだろうか。生成した「雨と風のある日」を表す結果を使って『琵琶湖年表』から記事を抽出してみると、やはり雪に関する記事が4件、台風や大雨に関する記事が5件見つかった。対策本部の設置や取水制限などの行政の対応、さらには民間の状況など、災害そのもののだけでなく、関連する様々な状況を知ることが出来た。

雪に関する記事

- ・冬将軍彦根市内で 32 年ぶりの積雪 67 センチメートル。（1984 年 2 月 8 日）
- ・滋賀県、雪害対策本部設置。（1984 年 2 月 9 日）
- ・どか雪、湖北、湖西で、除雪車フル出動。（1984 年 12 月 25 日）
- ・彦根・長浜など湖北で引き続き寒波襲来、この冬 2 番目の積雪となる。（1996 年 1 月 10 日）

台風・大雨に関する記事

- ・恵みの雨で水位はマイナス 75cm。前日より 6cm 回復。（1986 年 12 月 15 日）
- ・台風 19 号の大雨で、愛知川堤防決壊。（1990 年 9 月 20 日）
- ・近畿地方を台風 26 号が縦断し大雨をもたらす。この雨で、淀川の水量が増えたため、近畿地方建設局は取水制限を一時的に全面解除するが、滋賀県は、琵琶湖からの取水制限 8%を継続。（1994 年 9 月 29 日）
- ・台風 26 号により、永源寺町の蛇砂川が決壊。彦根市の犬上川も堤防がえぐられる。（1994 年 9 月 29 日）
- ・台風 9 号の接近に伴い、鳥人間コンテストが初めて完全に中止となる。（1997 年 7 月 26 日）

2.3 課題

上述の実例では、気象観測データと2つ文書資料を組合せ、特定の気象条件に合致する日の出来事を抽出する解析が行われた。これを、自然災害だけでなく農業、経済など他の分野にも広げてゆくことで、気象とそれらの分野に関する事象との関係、さらには、気象条件を介してそれぞれがどのように関係しているのが明らかになることも期待される。しかしながら、一般に対象とする分野が広がるほどデータの多様性は増すであろうし、また、量も多くなる。このような場合、時空間情報に基づく解析を行う上でどのような問題が生じてくるだろうか。ここでは、GT-Map/Timeなどで直接利用する時空間データそのものに関する課題を2つ、データの検索や共有に関する課題を1つ挙げ、次節でこれらの課題をとりまく現状について紹介する。

（1）基準となる時空間情報の種類

実例では、中間的な CSV ファイルを生成する段階で、時間情報として西暦、空間情

報として緯度経度が用いられ、元データの時空間情報はこれらに変換された。しかしながら、目的が異なればそれに応じた時空間情報を基準にした方が良い場合も考えられる。例えば、日本の歴史に関する主題を扱うのであれば、元号を伴う和暦が便利であろうし、空間情報でも緯度経度ではなく地名が使いやすいこともある。また、データを複数のアプリケーションで共用する場合も、それぞれのアプリケーションが求める形式に変換しながら使うことになる。このような汎用的な使い方を前提とした場合、時空間情報をどのような形に統一して整理・保管するのが適切であろうか。

(2) 時空間情報の表記

基準となる時空間情報の種類が決まったとしても、その表記方法（日付であれば 7 月 25 日、7/25、7.25、7・25 の違いなど）が異なれば、アプリケーションは読み取ることができない。結局表記方法に合わせてその都度変換処理をしてゆく必要が生じてしまう。今回は、CSV ファイルを生成させる過程で表記を統一する処理を行ったが、あらかじめ決まった形（標準化された形）になっていれば、データの読み間違いを防ぎ、処理もスムーズに進めることができるはずである。

(3) データベース間の連携

上述の実例では、あらかじめ用意されていたデータを用いて解析が行われたが、実際にはまずデータ自体を何らかの方法で探し出す必要に迫られる。そのためには複数のデータソースから効率よくデータを探し出すしくみ、つまりクリアリングハウスやデータベース間での連携が必要になってくる。また、時空間情報に基づく解析を行うのであれば、データが対象とする時間的、空間的な範囲を絞り込む必要もある。それぞれのデータベースがどのようなメタデータを利用し、そこにどのような形で情報が収容されているかを整理し、連携させてゆく必要がある。

3. 分野連携による時空間データ構築へ向けて

3.1 基準となる時空間情報

(1) 場所

場所を特定する方法は大きく分けて、緯度経度などの座標による空間参照と、地名や郵便番号などの地理識別子による空間参照とが考えられる。座標による空間参照は一定の条件下では普遍的であり、座標同士の換算、変換も明確であるので、空間情報の基準としてよく利用される。空間座標を正確に表現するには原子 (Datum) や座標系を座標値と共に示す必要がある。標準という観点では本来の意味での世界測地系 (ITRF94 座標系、GRS80 楕円体; 国内では JGD2000 が準拠) が適当とも思われるが、実際には GPS が用いている WGS84 が利用されることも多い。

一方、実際の多くの文書や資料では空間情報は地名として登場する。最近でこそ携

帯電話やデジタルカメラに GPS が内蔵され、自分自身がいる場所や写真の撮影場所の緯度経度を容易に知ることができるようになった。ところが、ひと昔前は GPS や六分儀を気軽に持ち歩くような状況には無かったので、空間情報はやはり地名が主役である。地名は、それ自体が変化すること、また指し示す場所が時代と共に変わることがあるなどの問題もあり、空間情報の基準としては必ずしも適切ではない。地名による空間情報を活用するためには、地名と緯度経度の対応や地名同士の階層構造を収録した地名辞典が必要になる。海外では TGN[6]や ADL Gazetteer[7]のような大規模な地名辞典が存在する。国内では、GT-Map/Time を作成した H-GIS 研究会や人間文化研究機構が中心となって、歴史地名の地名辞典を構築する動きなどがある。

(2) 時間

時間情報で問題なのは暦 (暦法) であろう。これは、空間情報における原子や座標系の違いに相当する。本邦においても和暦と西暦が併用され、二種類の年号が同時に存在している。海外に目を移せば、それぞれの国や地方で使われている様々な暦が存在するし、歴史的な資料であれば当時の暦による日付や時刻が用いられている。異なる暦に基づく資料の比較に用いる共通の軸となる暦には、連続性や変換の容易さが重視される。多くのケースでいわゆる西暦が用いられるが、西暦には連続性に関する問題がある。つまり、ユリウス暦からグレゴリオ暦に改暦された際に 1582 年 10 月 4 日の翌日が 10 月 15 日になっており、10 日間の不連続がある。さらにこの改暦の日が国によってまちまちであり、この時期の時間情報の扱いに注意を要する事態となっている。

時間情報の共通の軸としては、ユリウス通日が用いられるようになってきている。ユリウス通日は紀元前 4713 年 1 月 1 日正午からの日数である。ユリウス通日はその連続性や単独の数値であることから取り扱いが容易であり、それぞれの暦からユリウス通日への変換方法が提唱されている。これにより、ユリウス通日を介することで暦同士の変換を容易に行うことができる。地理情報の国際規格である ISO191XX シリーズでも時間のデータ型の 1 つとしてユリウス通日を導入しており、それぞれの暦による表記をユリウス通日に対応付けることを求めている。ユリウス通日も厳密にはうるう秒の扱いに問題があることや小数以下を切り捨てたユリウス日数 (JDN) などの幾つかのバリエーションが存在するので取り扱いに多少の注意は要する。

時間情報についても地名と同じように特定の名称で示されることがしばしばある。たとえば、戦後、バブル期、終戦の日などであり、これらを具体的な日時と対応付ける「時間名辞典」が必要である。一部に Time Period Directory[8]などの構想があるものの、その必要性があまり認知されてはいない。本格的に利用可能なものの開発が待たれるところである。

3.2 時空間データの表記

日付を表現するのに“03/02/01”といった表記をしばしば目にする．日本国内であれば，年月日の順番で並ぶことが慣習になっているので，3年2月1日と読めるが，海外の場合は定かではない．日付の表現は，日常生活に密着しているだけに，その表現方法も様々である．

表1 時間の表示形式に関する標準規格の例．JIS X 0301 は，ISO 8601 に元号による日付表現を追加するなどの修正を加えたものであり，JIS X 7100 シリーズ（地理情報）をはじめとする国内規格は JIS X 0301 を参照している．RFC 1123 は電子メールだけでなく，RFC 2616（HTTP）でも利用されている．RFC 2616 では，互換性を確保するために RFC 860 や ANSI C も解釈できることを要求している．ISO/IEC 8825-1 は RFC 3161（公開鍵暗号基盤のタイムスタンプ）などで採用されている．

規格	表記例	説明
基本的な規格		
ISO 8601（基本形式）	20090725T14:45:00+0900	
ISO 8601（拡張形式）	2009-07-25T14:45:00+09:00	
JIS X 0301（元号による日付）	H21.07.25	ISO 8601 を翻訳・修正したもの（MOD）
RFC 1123	Sat, 25 Jul 2009 14:45:00 +0900	SMTP 用の時間表記（RFC822 形式）の暦年を4桁にしたもの
RFC 850	Saturday, 25-Jul-09 14:45:00 JST	USENET 用の時間表記として策定されたもの
ANSI C	Sat Jul 25 05:45:00 2009	ctime 関数の出力形式
ISO/IEC 8825-1	20090725054500Z	
派生の規格		
RFC 3339	2009-07-25T14:45:00+09:00	ISO 8601 の一部を利用
W3C-DTF	2009-07-25T14:45:00+09:00	ISO 8601 の一部を利用
ITU-T X.690	20090725054500Z	ISO/IEC 8825-1

計算機上で扱われる時間の表記については幾つかの標準規格が存在する（表1）．この中で，現在最も広く用いられているのは ISO 8601（拡張形式）であり，RFC3339 や W3C-DTF など名前を変えて用いられることも多い．

この ISO 8601 表記については，紀元前の年号をどのように扱うか，つまり，紀元前1年1月1日は-0001-01-01 か，0000-01-01 かということがしばしば問題になる．規格の中ではグレゴリオ暦以前（1582年10月4日以前）の暦の表記をこの規格で扱う場合は情報交換当事者間で合意しておくこととしているものの，グレゴリオ暦をそのまま過去に拡張した場合，[0000]は閏年になると記されている．従って規格の中では0000で表される年が存在し，それは紀元1年の前の年である紀元前1年であること想定していることが伺える（実際のユリウス暦では，この頃の閏年の周期が厳密に決まっていなかったため，紀元前1年は閏年でなかった可能性がある）．

一方，緯度経度の表記は，大きい単位から順番に書く方法で一定しており，日付のように年月日の順番が入れ替わるような表記の混乱はみられない．また，度分秒まで書くか（DMS 型），分以下を分の小数点で書くか（DMM 型），度以下を度の小数点で書くか（DEG 型）といった違いはるものの，相互に変換することは容易である．但し，度分秒の区切りをどのように書くかについては統一した決まりが無い．混乱を避けるのであれば，常にどのような書式で記されているかを併記する必要があるであろう．

3.3 データベース間の連携

（1）メタデータ

様々な学問分野において標準化されたメタデータ規格が用いられており，分野の中での情報交換を促進するために用いられている．GT-Map/Time などの時空間解析を行うアプリケーション用のデータをこれらのメタデータを使って探そうとすれば，メタデータが対象とする資源の時空間範囲は重大な関心事であり，どのような形で時空間情報が収容されているかを把握する必要が生じる．表2は3つのメタデータについて，時空間情報を収容する項目と入力フォーマットを示したものである．

DC（Dublin Core）は周知の通り最も汎用的なメタデータとして既に広く用いられている．この中で，時空間データは date と coverage（spatial, temporal）にある．MODS（Metadata Object Description Schema）は米国議会図書館が管理している規格で書誌データとして用いられてきた MARC 21 の一部を XML に置き換えたメタデータ規格である．originalInfo や subject の中に時空間の記述があり，それぞれ DC の date や coverage に対応している．EML（Ecological Metadata Language）は，生態学関連のデータを交換するためのメタデータ規格である．もともとは長期に渡る生態学研究のためのプログラム（LTER：Long Term Ecological Research）のための規格であり，気象，水文，水質，地質などの物理化学的な情報から，生物の種類や量，個体群内の個体の大きさの分布などの生物学的な情報まで幅広い分野の情報を扱うことを前提としている．

eml-coverage に主題の対象範囲を示す項目があり、観測や採集が行われた日時や場所などの時空間情報が含まれる。また、EML では時空間情報だけでなく、対象となる生物種の範囲も eml-coverage に記入する（バラ科、スズキ目など）。この点は、特定の目的に特化したメタデータの例として興味深い。

（２）Metadata Suite

各種のメタデータを階層的に連携させる方法として、原（2009）は Metadata Suite という考え方を提唱している[9]。Metadata Suite は Database Layer, Standard Layer, Sharing Layer, Application Layer の４層に分かれている。Database Layer におかれた各データベースは、まず分野ごとに標準化されたメタデータ規格により Standard Layer で統合される。そして、Standard Layer の各分野のメタデータを Sharing Layer で汎用的なメタデータに統合することにより、全データベースが統合されてゆく。この Metadata Suite の利点は、分野ごとに標準化されたメタデータを Standard Layer で用いることにより、特定の分野の中ではより緻密なデータベース間の連携を行い、一方で Sharing Layer により精度は粗くなるものの全体を包括する連携を実現するという異なる連携を同時に実現することにある。また、いきなり全体を包括する連携を実現するのではなく、特定の分野に特化したメタデータをはさむことで、効率良くマッピングを進めることも利点であろう。これらの統合されたデータベースのデータは、Application Layer にあるクリアリングハウスや API (Application Program Interface) を介して GT-Map/Time などのアプリケーションで利用される。

（３）メタデータに基づくデータベースの接続実験

この Metadata Suite の考え方に沿う形で、総合地球環境学研究所（地球研）と京都大学地域研究統合情報センター（京大地域研）の間でデータベースの接続実験が行われた。地球研を中心に、「地域環境情報ネットワーク」事業が計画されている。この事業は、地域研究と環境研究に関連する様々なデータを連携させながら、新たな「地域の知」を見出してゆこうとするものであり地域研究と環境科学双方の情報を連携させることが求められる。事業の準備段階として、平成 20 年に地域研が保有する地域研究に関するデータベースと環境科学のデータの代表として EML によるデータベースとの接続実験が行われた。

EML のデータは Morpho と呼ばれる専用のエディタにより生成され、Metacat と呼ばれる EML 専用のデータベースに蓄積・管理される。今回は、この Metacat に収容されている情報と、地域研の各データベースを接続して横断検索を行う実験を行った。また、先行している人間文化研究機構（人文機構）の研究資源共有化システム（統合検索システム）についても試験的に接続し、横断検索の対象とした。

表 2 メタデータ規格の中での時空間情報（抜粋）

	項目	入力	説明
Dublin Core	Date	W3C-DTF	資源に関する出来事（出版等）が起こった日
	coverage.spatial	DCMI Point, ISO 3166, DCMI BOX, TGN	資源に関する空間的特性
	coverage.temporal	DCMI Period, W3C-DTF	資源に関する時間的特性
EML	dataset/pubDate	ISO 8601	資源がリリースされた日
	eml-coverage/geographicCoverage	geographicDescription 要素に地名等を入力 boundingCoordinates 要素に東西南北端の緯経度（DEG 型）を入力 datasetGPolygon 要素にポリゴンの各点の緯経度を入力	資源の地理上の対象範囲
	eml-coverage/temporalCoverage	clandarDate 要素, time 要素に ISO 8601 で入力 timeScaleName 要素と timeScale AgeEstimet 要素の組合せで地質年代等を入力	資源の時間上の対象範囲
MODS	originInfo/dateIssued	W3C-DTF, ISO 8601, MARC 21 (008/07-14)	資源がリリースまたは出版された日
	originInfo/dateCreated	W3C-DTF, ISO 8601, MARC 21 (008/07-14)	資源が作成された日
	originInfo/dateCaptured	W3C-DTF, ISO 8601, MARC 21 (008/07-14)	資源がデジタル化された日
	originInfo/dateOther	W3C-DTF, ISO 8601, MARC 21 (008/07-14)	資源に関するその他の重要な日
	subject/geographic	地名等を入力	地理的な主題に関する地名等（階層的な地名ではないもの）
	subject/hierarchicalGeographic	country, province, city 等の階層的に整理された要素に地名を入力	地理的な主題に関する階層的に整理された地名
	subject/cartographic	coordinates 要素に座標（緯経度等）を入力	地理的な主題に関する地図上のデータ
	subject/temporal	W3C-DTF, ISO 8601, MARC 21 (008/07-14)	時間的な主題に関する日付等

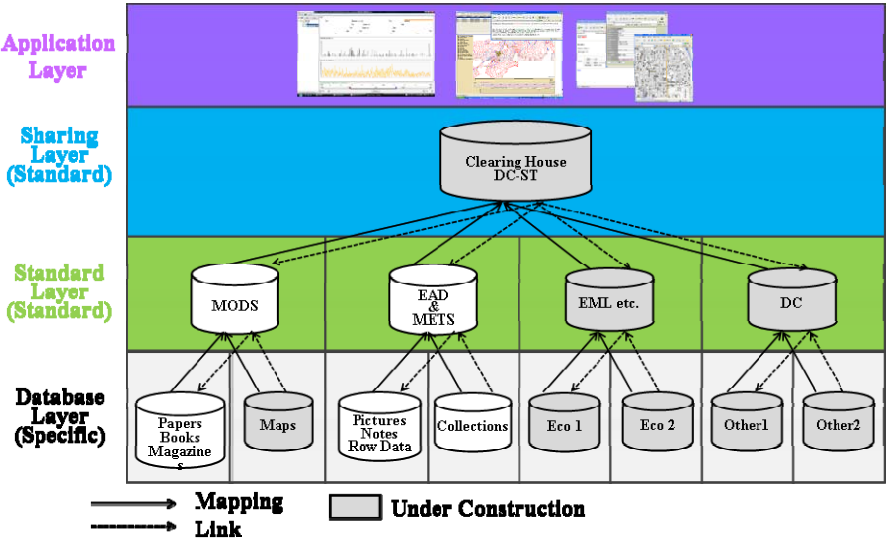


図3 Metadata Suite (原 2009)

MetaCat 内の EML のデータはフロントエンドシステム内でインデックス化され、EML と Dublin Core の間のマッピング (表 3) に基づいて京大地域研および人文機構の研究資源共有化システムと SRW および Z39.50 で横断検索が実現された。仕組みとしては、EML データ用の Metacat が入っている以外は基本的に人文機構の研究資源共有化システムと同じである。実験は 2009 年 3 月末に行われ、EML データと京大地域研、及び人文機構の研究資源共有化システムのデータが横断的に検索できることが確認された (図 4)。今回は Metadata Suite の Sharing Layer に相当する部分のメタデータとして Dublin Core が用いられたが、対象となる分野レベルのメタデータが増えてきた場合、Sharing Layer のメタデータをどのような構造にするのかは今後検討が必要である。

表 3 Dublin Core から EML, MODS へのマッピング例

Dublin Core	EML	MODS
title	dataset/title	titleInfo/title
creator	dataset/creator	name/namePart
subject	dataset/keywordSet	subject/topic classification
description	dataset/abstract	abstract note tableOfContents
publisher	dataset/publisher	originInfo/publisher
contributor	dataset/associatedParty	name/namePart
date	dataset/pubDate	originInfo/dateIssued originInfo/dateCreated originInfo/dateCaptured originInfo/dateOther
type	dataset citation protocol software	type/OfResource genre
format	dataset/dataTable/physical	physicalDescription/internetMediaType physicalDescription/extent physicalDescription/form
identifier	dataset/@packageId	Identifier location/uri
source	dataset/methods/methodStep/ dataSource	relatedItem[@type='original']/ titleInfo/title relatedItem[@type='original']/ location/uri
language		Language
relation	dataset/methods/methodStep/citation	relatedItem/titleInfo/title relatedItem/location/uri
coverage		
(spatial)	dataset/coverage/geographicCoverage	subject/geographic subject/hierarchicalGeographic subject/cartographics
(temporal)	dataset/coverage/temporalCoverage	subject/temporal
rights	dataset/intellectualRights	accessCondition

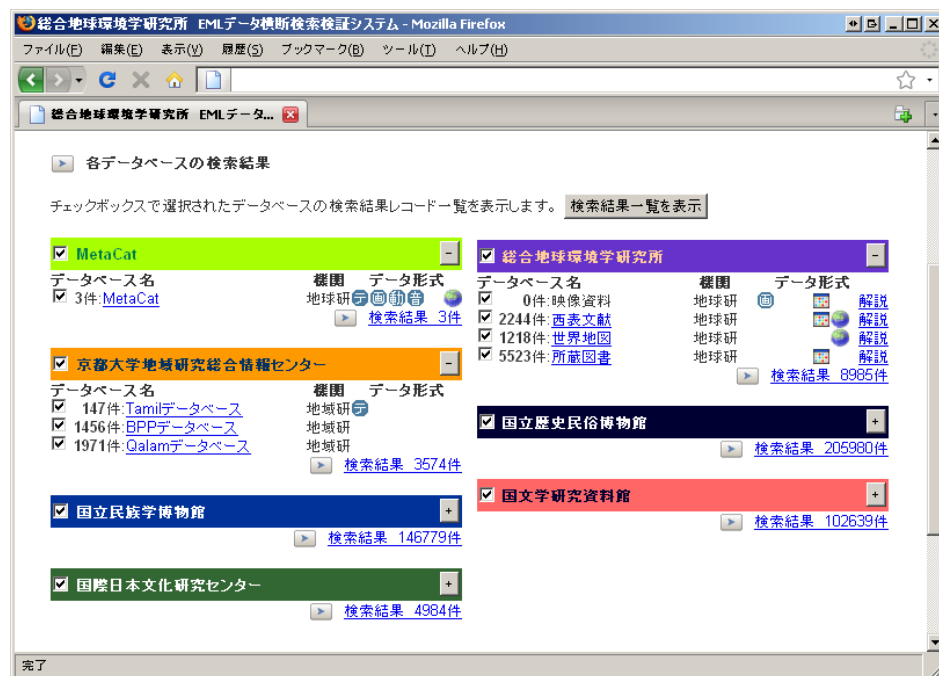


図4 EML (Metacat), 京都大学地域研究統合情報センター, 人間文化研究機構研究資源共有化事業の統合検索システム間の接続実験の様子。

4. 今後の展開

データベース間の連携により利用可能なデータソースが増えれば、異質なデータを時空間情報に基づいてつき合わせてみる機会は増えるであろう。一方で、分野間でのデータの交換が盛んになれば、分野間での用語や概念の違いが利用の妨げとなってくることも予想される。同じ語彙であっても分野ごとの意味の違うケースがある。また、時空間情報であってもそれが何を意味する時間や場所なのか分からないというケースも考えられる。結果として、データはダウンロードできたのだけれども利用できない、という事態も当然予想される。この点では、オントロジー技術の導入により、各分野の用語や概念意味づけが行われ、それらの間の関係が記述されてゆくことで解決がはかれることが期待される。

また、レコード単位やコンテンツのレベルでの時空間情報をどのように使うのかも今後の課題の1つである。今回はMODS, EMLなどのメタデータの規格についてふれてきたが、レコード単位やコンテンツのレベルのデータを記述するための規格も数多く存在する。KMLやGMLのように地理情報を中心に設計されたもの、TEIのように文書のコンテンツの記述用として設計されたものもある。また、スケジュール情報を交換するためのiCalendarやデジタルカメラの画像に埋め込まれたExifの撮影日時や撮影場所のGPSデータも時空間情報を使った解析に用いることが考えられる。

参考文献

- 1) 関野 樹: 時空間処理ツールと情報処理. 平成 20 年度京都大学地域研究統合情報センター共同研究ワークショップ, 2009 年 4 月 25 日, 稲盛財団記念館, 京都.
- 2) 気象庁: 気象統計情報. <http://www.data.jma.go.jp/obd/stats/etrn/index.php>
- 3) 滋賀県: 滋賀県災害誌第三部 (1990).
- 4) 滋賀県: 滋賀県災害誌第四部 (2000).
- 5) 滋賀県琵琶湖研究所: 琵琶湖年表. 滋賀県琵琶湖研究所所報 1-21 (1982-2002).
- 6) Getty Thesaurus of Geographic Names Online. http://www.getty.edu/research/conducting_research/vocabularies/tgn/
- 7) ADL Gazetteer. Alexandria Digital Library Project. <http://it-drupal.library.ucsb.edu/adl-gazetteer>
- 8) Vivien Petras, Ray R. Larson & Michael Buckl: Time period directories: A metadata infrastructure for placing events in temporal and geographic context. 6th ACM/IEEE-CS Joint Conference on Digital Libraries; June 11 - 15, 2006. Chapel Hill, NC, USA.
- 9) 原 正一郎: 地域情報学の道程. 平成 20 年度京都大学地域研究統合情報センター共同研究ワークショップ, 2009 年 4 月 25 日, 稲盛財団記念館, 京都.