

# オントロジを利用したメタデータ構築に基づく 地球観測データ統合検索フレームワークの研究

高橋 慧<sup>†</sup> 絹谷 弘子<sup>††</sup> 吉川 正俊<sup>†</sup>

現在様々な機関において地球観測データの作成が行われているが、観測期間によって利用できるデータの観測対象や観測頻度、データの集約方法や観測地点が異なる場合が多い。これらを統合し解析を行う場合、その差異を吸収したシステムを構築する必要がある。本稿ではオントロジを利用してデータ間の差異を明確にし、統合利用できるデータ間の横断的検索フレームワーク構築について述べる。地球観測データの横断的な検索のための概念スキーマを提案し、概念スキーマ上で横断的検索を実現するための関係代数演算の定義を行う。また本スキーマを利用した、データ取得可能性を明示しデータ検索効率を高めるためのユーザインタフェースについて説明する。

## Research for an Integrated Earth Observation Data Retrieval Framework Based on Metadata Development with Ontologies

AKIRA TAKAHASHI<sup>†</sup> HIROKO KINUTANI<sup>††</sup>  
and MASATOSHI YOSHIKAWA<sup>†</sup>

The earth observation data comes from various sources in various forms. They differ in many ways, e.g. subject, frequency, means of aggregation, and locations. Those difference must be clearly described when integrating and using earth observational data from different sources. In this paper, we describe how we build metadata of earth observational data for integrated use with help of ontologies. We also present an efficient data retrieval interface by showing the availability of data using the framework.

### 1. はじめに

近年の観測技術やリモートセンシングの技術向上に伴い、さまざまな観測機関により地球観測データの取得が行われており、一部は web 上で広く公開されている。このような地球観測データは、それを収集し、永続的な蓄積、統合、解析を行うことによって、地球科学研究や社会に有益な情報を導くことができる。しかし収集した観測データを統合利用、解析利用するためには観測機関によって異なる観測項目や期間、地点などの差異を統一形式に変換して明確化し、統合利用可能なデータを示す必要がある。例えば、気象予測をより精密に行うためには、対象地域で観測されている質の高いデータを優先的に使う要求がある。観測データは様々な要因により欠損値が生じることがあり、統合利用可能なデータは、データ項目のカatalogだけで

は判断できない。実際にデータが取得可能かどうかを提示することにより、効率的な利用が可能となる。このように地球観測データの記述を統一し、統合利用を行うための研究は盛んに行われている。

しかし、既存の研究<sup>8),10),11)</sup>ではデータが統合利用できるかどうかは利用者がデータの性質を見て判断する必要がある場合が多い。そこで本研究ではオントロジを利用したメタデータ設計を行い明確な枠組みを作ることによって、観測データの持つ意味をシステムが判定・利用できるようにし、将来多様な、そしてより有用な観測データが投入されてもオントロジの拡張により柔軟に対応するフレームワークを考える。オントロジを利用してデータの統合を行う研究はすでいくつか存在するが<sup>7),9)</sup>、本研究ではさらに検索用概念スキーマ設計によりデータ取得可能性やデータの統合可能性、データの質などを考慮した検索インタフェースをしている。

本稿では以下、2章で構築するメタデータの基本的な設計及び利用するオントロジの説明を行い3章でメタデータの変換可能性とその利用について説明する。

<sup>†</sup> 京都大学情報学研究所  
Graduate School of Informatics, Kyoto University

<sup>††</sup> 東京大学 生産技術研究所  
Institute of Industrial Science, the University of Tokyo

4章でフレームワークの実装, およびフレームワークを利用したユーザインタフェースについての説明を行い, 最後に5章でまとめと今後の課題について記述する.

## 2. オントロジを利用したメタデータ設計

本章では我々の設計した地球観測データのメタデータについて説明する. まず統合のためにデータの異種性を吸収する統一概念スキーマを示し, 次にスキーマにおけるオントロジの利用について述べる.

### 2.1 検索用概念スキーマ

異種データ集合をボトムアップ的に統合利用する場合は, 各データ集合の異種性を吸収した統一的な概念スキーマが利用者に提供されることが望ましいと考えられる. 本研究では, 利用可能なデータはすべて概念スキーマ上の Availability という関係表に格納されているものとする. この表には, データ観測を行ったがプロジェクト名, 観測地点, 観測項目, 観測時刻などのデータを持つ. この表の概要を作成することにより, 複数のデータ集合の統合利用可能な部分を容易に知ることができる.

Availability 表に格納されるデータのうち, 複数の観測地点は緯度, 経度により容易に比較可能である. また, 観測期間もデータ集合によらず minute, hour, day, month など共通の単位を持つ. それに対し, 観測項目は, 観測データ集合によって多様性が高い. そこで, 観測項目はオントロジを用いて詳細に記述し, 異なる観測項目の統合利用可能性を厳密に判断できるようにする.

本研究では地球観測データ検索に必要なメタデータを一つの概念スキーマに統合して表現している. 本研究で用いる地球観測データ検索のための概念スキーマは以下の通り.

Availability (Project, Latitude, Longitude, Name, Item, Time, Value, Quality)

以下にスキーマの各項目の説明を行う. Project はプロジェクトを一意的に識別するための識別子, Latitude, Longitude は観測地点の緯度経度情報. Name はプロジェクト内で観測地点を一意的に識別する地名. Item は観測項目であり, Time は観測された時刻. Value は観測された値である. Quality はデータの質を表すフラグが格納される. データの質的チェックが行われていない, データが異常値であったり, 値を導く過程で桁落ちが発生したり, 欠損値の存在により厳密な値が算出されていない等の場合にそれを表すフラグを格納する. データは基本的にその値毎にスキーマ

内で組を保持する. 表1に本スキーマを利用して記述した地球観測データの例を示す.

一般に地球観測データにおける観測項目 (Item) は観測対象や量などの多様な性質を持っていると考えられる. そこで, 本スキーマを利用する際にはこの観測項目の持つ性質をオントロジを利用して記述する事を考えている. 具体的な記述方法については次節以降で説明を行う.

また本章の議論はこの概念スキーマに対して行われるが, 実際の実装では拡張性を考慮してスキーマの分解を行っている. 実装に用いるスキーマについては4章で述べる.

### 2.2 観測項目のオントロジ記述

地球観測データにおいて観測項目を決定付ける要素は様々であるが, 主に以下の5つの要素が必要であると考えている.

観測対象: 大気や土壌, 降雨など, 観測の対象となる物質や現象.

観測周期: 観測が行われる周期や, 観測を集計する期間.

観測量: 温度や湿度, 質量, 速度など, 観測対象の持つ性質.

観測単位: 摂氏や華氏, メートルなどの, 記録される観測データの単位.

集約方法: 平均や最大・最小など, 観測項目がある一定の期間のデータとして与えられる場合にどの様に集約・算出されているか. データが観測時の瞬間的な値を示す場合はこの要素は不要である. 本研究ではこれらをオントロジを用いて観測項目毎に記述し, その差異を機械にも人間にも明確にすることを考えている.

#### 2.2.1 SWEET オントロジを利用した観測項目の表現

観測項目の持つ多様な性質の次元を表現するには, 次元毎に語彙をもった強力なオントロジが必要である. しかし, このようなオントロジを自ら構築するには非常にコストがかかるため, 本研究ではNASAのジェット推進研究所によって整備されており, すでにいくつかの地球科学プロジェクトでの利用実績<sup>2),7),9)</sup>があるSWEET<sup>6)</sup> オントロジを利用する. SWEETは地球環境データに関する概念を記述するための語彙を持っている. SWEETでは, 例えば“Air Temperature”のような複合語をそのまま定義するのではなく, 属性である“Temperature”と対象となる“Air”を個別に定義することでその概念を明確に定義している, このように観測項目が持つ性質の次元それぞれに個別のオン

データ集合 A

Project	Latitude	Longitude	Name	Item	Time	Value	Quality
a	35.00	135.75	Kyoto	ex:max_temp_monthly	2007-06	28.5	-
a	35.00	135.75	Kyoto	ex:max_temp_monthly	2007-07	35.5	-
a	35.00	135.75	Kyoto	ex:max_temp_monthly	2007-08	38.0	-
:	:	:	:	:	:	:	:

データ集合 B

Project	Latitude	Longitude	Name	Item	Time	Value	Quality
b	34.70	135.50	Osaka	ex:max_temp_daily	2007-06-01	24.5	-
b	34.70	135.50	Osaka	ex:max_temp_daily	2007-06-02	25.2	-
b	34.70	135.50	Osaka	ex:max_temp_daily	2007-06-03	24.8	-
:	:	:	:	:	:	:	:

表 1 Availability データの例

トログを設けることによって高い拡張性を有しており、拡張性を要する本研究に用いるのに適切であると考えた。SWEET はいくつかのオントロジからなるオントロジ群である。以下に本研究で主に利用する SWEET オントロジの説明を行う。

**property:** 物理的属性を記述するためのオントロジ。Temperature, Pressure, Speed などの語彙が含まれる。

**substance:** 物質など記述するオントロジ。原子や代表的な分子, Air や Soil などの語彙が含まれる。

**phenomena:** 物理的現象を記述するためのオントロジ。Rain や Wind, ElNino などの語彙が含まれる。

**time:** 時間を記述するためのオントロジ。Second や Minute に加え, 季節や地質時代などの語彙が含まれる。

**space:** 空間的な情報を記述するためのオントロジ。方角などの記述に用いることができる。

**unit:** 単位を記述するためのオントロジ。meter や degreeC に加え, mega などの接頭辞も定義されている。

**numerics:** 数字に関するオントロジ。interval や point, 0 に加え maximum, minimum, summation, average や normalization など演算に関する語彙も含まれている。

本研究では観測項目はこれらの語彙を用い, RDF で記述を行う。また SWEET では記述しきれない部分は SWEET を拡張する語彙を作成して対応する。観測項目の各要素とオントロジの対応を表 2 に示す。

表 2 の Property のうち, ex:aggregatedBy 以外は必須である。

例として, あるプロジェクトにおいてある日の最高気温を華氏で集計している場合と, 10 分毎に気温

を摂氏で測定している場合の観測項目をそれぞれ, ex:max\_temp\_daily, ex:air\_temp\_10min とするとき, それぞれを RDF グラフで表現すると図 1, 図 2 のようになる。図表中の接頭辞は主に SWEET オントロジの名前空間を表し, ex は本研究で新たに定義する語彙の名前空間を表す。

### 3. メタデータの変換と利用

本章では, 2 章で述べたメタデータ概念スキーマを利用することによって, 統合利用に有用なデータの変換について考えていく。まずデータの単位変換をオントロジを利用してどう行うか述べ, 次に時系列に関するデータの変換について考察していく。

#### 3.1 観測単位の変換可能性

観測項目は前節で示したようにオントロジを利用してその差異が明確化されるが, 一部の観測項目には変換を加えることによって同じ観測項目と見なすことができる。例えば units が華氏である気温と摂氏である気温は, どちらかのデータに適当な変換を加えることによって比較が可能となる。このように, 単位の変換によって観測単位が変換可能である場合は, SWEET オントロジの枠組み内で厳密に定義できる。すなわち, ある観測単位  $x$  から  $y$  にデータの変換が可能な場合,  $x$  と  $y$  を構成する次元それぞれにおいて, derivedFromUnit プロパティを辿って至るリソースが一致する, または一方の derivedFromUnit プロパティを辿ると他方に到達する。例として, 図 3 に示す SWEET オントロジ内の華氏と摂氏の記述を見る。摂氏, 華氏は “degreeC”, “degreeF” としてそれぞれ図 3 の 1~6 行目, 15~22 行目, において定義されている 16~18 行目より “degreeF” の derivedFromUnit プロパティが示すリソースは “degreeCby1.8\_interimInstance”, 10 行目より “degreeCby1.8\_interimInstance” の derived-

観測項目の要素	Property	domain ontology
観測対象	phenomena:hasAssociatedPhenomena substance:hasAssociatedSubstance	phenomena substance など
観測周期	ex:observationInterval	time
観測量	property:hasAssociatedQuantity	property など
観測単位	numerics:hasUnits	units
集約方法	ex:aggregatedBy	numerics

表 2 SWEET オントロジを用いた観測項目の記述

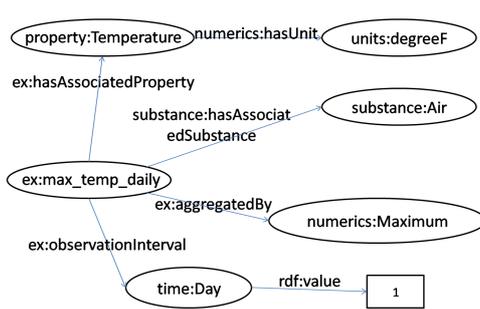


図 1 ex:max\_temp\_daily の RDF 表現

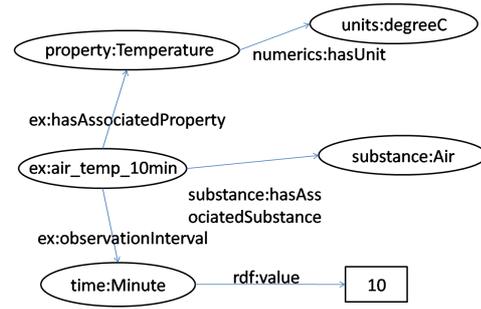


図 2 ex:air\_temp\_10min の RDF 表現

```

1: <units:UnitDerivedByShifting rdf:about="#degreeC">
2:   <units:derivedFromUnit rdf:resource="#kelvin"/>
3:   <units:hasShiftingNumber rdf:datatype="&xsd;double">
4:     -273
5:   </units:hasShiftingNumber>
6: </units:UnitDerivedByShifting>
7: <units:UnitDerivedByScaling
8:   rdf:about="#degreeCby1.8_interimInstance"
9: >
10:  <units:derivedFromUnit rdf:resource="#degreeC"/>
11:  <units:hasScalingNumber rdf:datatype="&xsd;double">
12:    1.8
13:  </units:hasScalingNumber>
14: </units:UnitDerivedByScaling>
15: <units:UnitDerivedByShifting rdf:about="#degreeF">
16:   <units:derivedFromUnit
17:     rdf:resource="#degreeCby1.8_interimInstance"
18:   />
19:   <units:hasShiftingNumber rdf:datatype="&xsd;double">
20:     32
21:   </units:hasShiftingNumber>
22: </units:UnitDerivedByShifting>

```

図 3 SWEET における摂氏、華氏の定義 (SWEET Units オントロジより引用)

FromUnit プロパティが示すリソースは “degreeC” であり, “degreeC”, “degreeF” は前述の条件を満たすので互いに単位が変換可能である.

### 3.2 観測項目の統一可能性

観測項目は単位の変更で比較可能となる場合に加え, 時間軸において適切に値の集約を行い観測周期を一致させることによって比較可能になる場合がある. 例えば, 日の最高気温と月の最高気温という二つの観測項

目において, 日の最高気温を一ヶ月毎に集約することによって月の最高気温と比較可能となる. 観測周期を一致させる変換を本研究では観測項目の統一と呼ぶ. 以下の場合, 観測項目  $\alpha$  から  $\beta$  への統一が行えると定義する.

- (1)  $\alpha$  と  $\beta$  の観測対象, 観測量が同じ.
- (2)  $\alpha$  の観測単位が  $\beta$  の観測単位に変換可能.
- (3)  $\alpha$  と  $\beta$  の集約方法が同じ, 若しくは,  $\alpha$  の集約が行われていない.
- (4)  $\alpha$  の観測周期が  $\beta$  の観測周期と同じか短い.

この条件は, 観測項目の RDF 表現において, 以下の条件とそれぞれ等価である.

- (1)  $\alpha$  と  $\beta$  において hasAssociatedQuantity, hasAssociatedSubstance (または hasAssociatedPhenomena) で示されるノード要素がそれぞれ等しい
- (2)  $\alpha$  の観測単位を表すリソースと  $\beta$  の観測単位を表すリソースの derivedFromUnit プロパティを辿って至るリソースが一致する. (または一方の derivedFromUnit プロパティを辿ると他方に到達する.)
- (3)  $\alpha$  と  $\beta$  において aggregatedBy で示されるリソースが等しい, または  $\alpha$  に aggregatedBy プロパティが存在しない.
- (4)  $\alpha$  の observationInterval で示される値が  $\beta$  のそれ以下である.

$\alpha$  から  $\beta$  へ統一が可能であるとき  $\alpha \Rightarrow \beta$

と表記するまたこのとき  $\alpha$  は  $\beta$  への変換可能性があるという.  $\alpha$  と  $\beta$  は  $\gamma$  に統一可能であると言う. 例として図 1, 図 2 で表される観測項目 ex:max\_temp\_day と ex:air\_temp\_10min を見ると, ex:air\_temp\_10min は ex:max\_temp\_day への変換が可能であり, ex:max\_temp\_day と ex:air\_temp\_10min は ex:max\_temp\_day, 及び ex:max\_temp\_day より観測周期の長い観測項目に統一可能である.

実際の観測データには値の欠損や異常値が含まれる可能性があり, 厳密には, そのような値を含む項目は変換されないが, 利便性を考えるとこのような変換も許容されるべきである. このような変換の際には, 変換後の観測項目を持つデータの, Quality 属性には, 欠損値を含む変換がなされたことを表すフラグを立てる.

### 3.3 観測データ時系列の統一

概念スキーマを用いてデータを保存している場合, Project, Latitude, Longitude, Name, そして Item の等しいデータの集合を導くことができる. この集合のある地点のある観測項目に関する, (Time, Value, Quality) の系列とみなすことができる. この系列を以降, 観測データ時系列と呼ぶ. 観測項目  $\alpha$  に関する観測データ時系列を以下の様に表記する.

$$\alpha\{(t_1, v_1, q_1), (t_2, v_2, q_2), \dots, (t_n, v_n, q_n)\}$$

観測データ時系列の変換方法は, データの集約方法, つまり観測項目の RDF 表現における aggregatedBy の示すノードに依存する. 本研究では現在 aggregatedBy の値域として SWEET の numerics オントロジに含まれる Maximum, Minimum, Summation, Average, NumberOfSamples を考えている. データの集約には他にもサンプリングなどが考えられるが現段階ではそのような変換を行ったデータを入手していないため, 本研究では基本的な集約演算のみを扱うことにしている.

以下にそれぞれについてどのように観測データ時系列の変換が行われるかを述べる.

#### 3.3.1 Maximum, Minimum

観測項目  $\alpha, \beta$  が  $\gamma$  に統一可能で,  $\gamma$  の aggregatedBy プロパティの示すリソースが Maximum である場合を考える.  $\gamma$  の観測周期を  $\tau_\beta$  とし, ある地点において観測項目を  $\alpha$  から  $\gamma$  へ変換した場合, 観測データ時系列  $\alpha\{(t_1, v_1, q_1), (t_2, v_2, q_2), \dots, (t_n, v_n, q_n)\}$  (以降  $A$  と表記する) は以下のように変換される.

$$\beta\{(T_1, \max\{v_i | i = 1, 2, \dots, m_1\}, Q_1), \dots, (T_1 + (k-1)\tau_\beta, \max\{v_i | i = m_{k-1} + 1, m_{k-1} + 2, \dots, m_k\}, Q_k)\}$$

ただし  $m_i$  は  $t_m < T_1 + (i-1)\tau_\gamma$  を満たす最大の  $m$  であり  $m_k \leq n$  である. 以後簡単のために,  $\{v_i | i = m_{k-1} + 1, m_{k-1} + 2, \dots, m_k\}$  を  $V_k$  と表記する.

なお, 変換の際  $Q_i$  の値には  $V_i$  に異常値が含まれる場合や, 値の欠損がある場合にはそれぞれを含んだ集約を行ったことを示すフラグを立てる.  $\beta$  に関する変換も同様に行われる.

$\beta$  の aggregatedBy の示すリソースが Minimum の場合も同様,

$$\gamma\{(T_1, \min\{v | v \in V_1\}, Q_1), \dots, (T_1 + (k-1)\tau_\gamma, \min\{v | v \in V_k\}, Q_k)\}$$

に変換される.

#### 3.3.2 Summation, Average

観測項目  $\alpha, \beta$  が  $\gamma$  に統一可能で,  $\gamma$  の aggregatedBy プロパティの示すリソースが Summation ある場合を考える. ある地点において観測項目を  $\alpha$  から  $\gamma$  へ変換した場合, 観測データ時系列  $A$  は以下のように変換される.

$$\gamma\{(T_1, \sum_{v_i \in V_1} v_i, Q_1), \dots, (T_1 + (k-1)\tau_\gamma, \sum_{v_i \in V_k} v_i, Q_k)\}$$

Average の場合も同様に,

$$\gamma\{(T_1, \frac{\sum_{v_i \in V_1} v_i}{|V_1|}, Q_1), \dots, (T_1 + (k-1)\tau_\gamma, \frac{\sum_{v_i \in V_k} v_i}{|V_k|}, Q_k)\}$$

に変換される.

#### 3.3.3 NumberOfSamples

観測項目  $\alpha, \beta$  が  $\gamma$  に統一可能で,  $\alpha$  の aggregatedBy プロパティが無く,  $\beta$  の aggregatedBy プロパティの示すリソースが NumberOfSamples である場合を考える. ある地点において観測項目を  $\alpha$  から  $\gamma$  へ変換した場合, 観測データ時系列  $A$  は以下のように変換される

$$\gamma\{(T_1, |V_1|, Q_1), \dots, (T_1 + (k-1)\tau_\gamma, |V_k|, Q_k)\}$$

次に,  $\alpha, \gamma$  の aggregatedBy プロパティの示すリソースが共に NumberOfSamples である場合を考える. このとき観測データ時系列  $A$  は以下のように変換される

$$\gamma\{(T_1, \sum_{v_i \in V_1} v_i, Q_1), \dots, (T_1 + (k-1)\tau_\gamma, \sum_{v_i \in V_k} v_i, Q_k)\}$$

$\beta$  に関する変換も同様に行われる.

### 3.4 メタデータを利用したデータ検索

地球観測データの検索にはその性質上観測項目情報, 時間情報, 位置情報の3方向からの検索が重要であるといえる<sup>10)</sup>. 本節では前節までに述べてきたメタデータとその変換手法を用いて, 観測項目情報, 時間情報,

位置情報に関してどのような検索を実現できるかを述べていく。

#### 3.4.1 観測項目を指定した検索

本研究では観測項目の統一可能性を利用することによって、指定した観測項目以外にも、変換して利用できる観測項目全てに関して検索することができる。

概念スキーマで記述されたデータ集合  $A$  に対して以下の選択演算が行われた場合、 $A$  中のデータのうち、観測項目が  $\alpha$  であるデータに加え、 $\alpha$  に統一可能な観測項目はデータを  $\alpha$  に変換した上で返す。

$$\sigma_{A.item \Rightarrow \alpha} A$$

例として、表 1 のデータに対して

$$\sigma_{(A \cup B).item \Rightarrow ex:max\_temp\_monthly} A \cup B$$

のような演算を行った場合を考える。この選択演算ではまず観測項目  $ex:max\_temp\_monthly$  への統一が可能である観測項目をもつ組が選択され、観測項目ごとにまとめられる。統合のために観測項目の統一が必要なデータに関してはそれぞれ変換を行い、最後に全体の和集合を取り、結果として表 3 のようなデータが得られる。

3

また、同じくデータ集合  $A, B$  に対して以下の結合演算が行われた場合、 $A.item$  と  $B.item$  を観測項目  $\gamma$  に統一できるデータに関して統一を行い結合する。

$$A \bowtie_{A.item \equiv B.item} B$$

例えば表 1 で与えられる入力データに対して

$A \bowtie_{A.Item \equiv ex:max\_temp\_monthly} B.Item, A.Time = B.Time$   $B$  を行うと表 4 のような結果を返す。

#### 3.4.2 時間を指定した検索

時間を指定した検索においては、観測項目の  $observationInterval$  プロパティを利用して、利用できるデータを自動的に絞り込むことができる。選択条件式や結合条件式の中で時間に対して加算、減算などが行われた場合加える数、引く数の絶対値より大きい周期をもった観測データは自動的に選択しない。

例として、概念スキーマ中のデータ集合  $A$  に対して、 $\sigma_{t-24h} A$  のような演算は

$$\sigma_{(t-24h) \wedge (A.item_{observationInterval} \leq 24h)} A$$

に自動的に書き換えることができる。ここで、 $item_{observationInterval}$  は観測項目の  $observationInterval$  によって示される観測周期を示す。

#### 3.4.3 位置を指定した検索

本節では概念スキーマ上で緯度経度を指定した検索について議論する。緯度経度情報を利用した検索は地球観測データを検索するにあたって非常に重要な検索方法であると言え、どのような検索需要があるかは

データの用途によって異なる。ある緯度経度を持つ点が入力として与えられた場合、以下の様な検索の指定を実現する。

- (1) 入力点を指定した矩形範囲の指定。
- (2) 最も近傍の観測点
- (3) 指定された地点の属する市町村などの行政区域内に存在する観測点
- (4) 指定された地点の属する河川流域に存在する観測点

これらの検索を実現するため、ジオコーディングサービスの利用や、地図を利用したインタフェースが有用であるといえる。またこれまでに議論してきた観測データ時系列の変換と同様の変換をデータの空間系列についても今後考慮していく必要があると言える。

## 4. 統合検索フレームワークの実装

本研究で設計したフレームワークの概念スキーマを通じてデータ検索・解析を行うクライアントアプリケーションは容易にデータに到達できるようになる。本章では、フレームワークの概念スキーマの実装方法の要旨および、概念スキーマに対して検索を行うアプリケーションの例としての統合メタデータ検索インタフェースについて説明を行う。

なお試験的にシステムに投入する地球観測データとして、中央農業総合研究センター<sup>4)</sup>が公開している FieldServer<sup>1)</sup> 観測データ、および NOAA<sup>5)</sup> が公開を行っている WMO の観測データを用いた。それぞれのデータの特徴として、FieldServer 観測データは比較的最近のデータが主であり、観測点によって観測周期や観測項目に違いがある。対して、WMO の観測データには数十年前のデータも存在し、また観測周期は 1 日単位であり、観測点によって観測項目にも大きな差はない。

### 4.1 メタデータスキーマ

2 章で概念スキーマの提示を行ったが、実際にはスキーマを用いてデータ値を全て保存するのではなく、メタデータ部では値は保存しない。また概念スキーマにおいて各データ毎に組を持つとデータ量が膨大になってしまうため、実装ではスキーマの分解を適切に行い、拡張性を確保する。この実装スキーマと実データを組み合わせ、データ問い合わせ変換を適切に行うことにより、概念スキーマのもつ性質を再現している。図 4 に概念スキーマと実装スキーマの関係を示す。実装に用いるスキーマを以下に定義する。

MonthlyAvailability

(ProjectID, Latitude, Longitude, Name,

$$\sigma(A \cup B).item \Rightarrow ex:max\_temp\_monthly^{A \cup B}$$

Project	Latitude	Longitude	Name	Item	Time	Value	Quality
a	35.00	135.75	Kyoto	ex:max_temp_monthly	2007-06	28.5	-
a	35.00	135.75	Kyoto	ex:max_temp_monthly	2007-07	35.5	-
a	35.00	135.75	Kyoto	ex:max_temp_monthly	2007-08	38.0	-
:	:	:	:	:	:	:	:
b	34.70	135.50	Osaka	ex:max_temp_monthly	2007-06	27.4	-
b	34.70	135.50	Osaka	ex:max_temp_monthly	2007-07	36.0	-
b	34.70	135.50	Osaka	ex:max_temp_monthly	2007-08	37.5	-
:	:	:	:	:	:	:	:

表 3 変換可能性を利用した選択演算の例

$$A \bowtie A.Item \equiv ex:max\_temp\_monthly^{B.Item, A.Time=B.Time} B$$

A.Project	A.Latitude	A.Longitude	A.Name	B.Project	B.Latitude	B.Longitude	B.Name	Item	Time	A.Value	B.Value
a	35.00	135.75	Kyoto	b	34.70	135.50	Osaka	ex:max_temp_monthly	2007-06	28.5	27.4
a	35.00	135.75	Kyoto	b	34.70	135.50	Osaka	ex:max_temp_monthly	2007-07	35.5	36.0
a	35.00	135.75	Kyoto	b	34.70	135.50	Osaka	ex:max_temp_monthly	2007-08	38.0	37.5
:	:	:	:	:	:	:	:	:	:	:	:

表 4 統一可能性を利用した結合演算の例

ItemID, DA\_Month, DA\_Quality, URI)

#### Missing\_Value

(ProjectID, Latitude, Longitude, Name, ItemID, MV\_Date, MV\_Quality)

MonthlyAvailability には観測項目を月単位に集約したデータを格納する。観測周期が1ヶ月より大きい場合、観測項目の集約は行わないが、同様に月単位でデータを保存する。集約の際に異常値や欠損値が含まれる場合、DA\_Quality にそのフラグを立てる。

Missing\_Value には異常値や欠損値を含む日についての情報が格納され、同様に MV\_Quality に異常値または欠損値のフラグを立てる。

このようにスキーマを分割することによって、大まかな観測期間の比較や欠損値が存在しない月等のデータ取得可能性は MonthlyAvailability へのアクセスだけで判定でき、日単位より細かいデータ取得可能性が必要な場合にのみ Missing\_Value にアクセスする。

また観測項目は2章に示したように RDF で与えられるが、2章で提示した RDF グラフの形状は観測項目間で大きな差異はないため、管理を容易にするために各属性を抜き出し RDBMS で管理する。観測項目を保存するスキーマを以下に示す。

#### Item

(ItemID, ItemName, Property, Target, Unit, Interval, IntervalValue)

### 4.2 統合メタデータ検索インタフェース

本節では今回実装した、フレームワークを利用したデータ検索インタフェースの説明を行う。本インタフェースの設計方針としては、観測項目、地域、時間など様々な条件を同時に指定する複合検索の実現、お

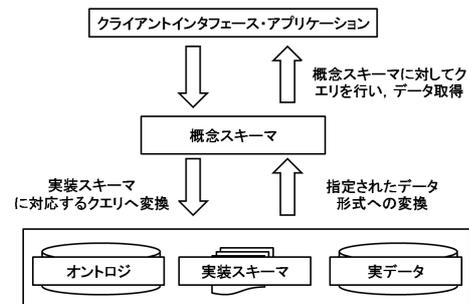


図 4 概念スキーマと実装スキーマの関係

よび地図インタフェースによる利用可能データの視覚的把握の実現を目指している。

図 5 に実装した検索画面全体を示す。画面左部に検索条件指定部、画面中央部に GoogleMapsAPI<sup>3)</sup> を利用した地図インタフェース、画面右部に指定条件での検索結果のサマリが表示される。

#### 4.2.1 3方向からの検索とプリフェッチ機能

検索条件指定部では、観測データの検索条件として、観測項目、観測地域、観測期間の3つのパラメータを指定することができる。観測項目の指定は、観測項目を観測対象 (target) や観測量 (property)、観測周期 (interval) の上限等をそれぞれ指定することによって行われる。観測項目を複数指定することも可能で、複数の観測項目を必ず含む (AND 検索) か、何れかを含む (OR 検索) などの指定も可能である。また観測項目からの検索では指定された観測項目に統一可能である、3.2 節で示した条件を満たす他の観測項目も合わせて検索を行う。

観測地域を指定した検索では緯度経度それぞれ上限

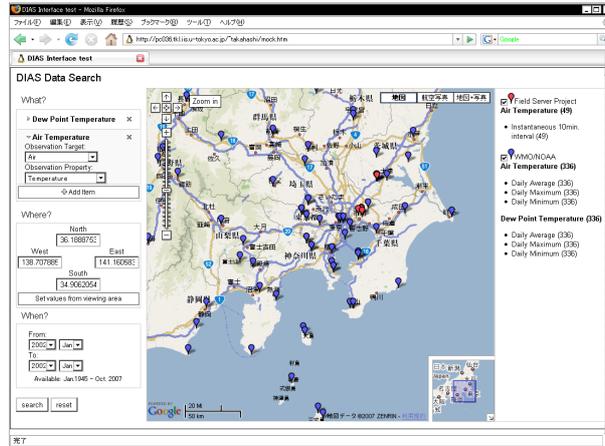


図5 統合メタデータ検索インターフェース

と下限を指定することにより得られる矩形範囲内の観測地点の検索が行える。3つのパラメータのうち何れかが指定されると、その値に応じて他のパラメータの指定可能な範囲が制限される。例えば、観測地域の指定を行うとその地域内で観測されているデータが存在する観測項目のみが、観測項目の検索パラメータとして指定できるよう、観測対象や観測量の選択肢が自動的に減らされる。また地域内で観測されたデータの存在する期間が提示される。各パラメータの指定・変更が行われると検索条件を元に、利用可能なデータの問合せを行い、結果を地図インターフェース上にプロットする。また検索結果サマリの表示部では、データベースやプロジェクト毎にデータが利用可能な観測サイトの数が表示される。この様に検索条件の制限したり、地図インターフェースでデータの有無を確認できるようにすることで、実際のデータにアクセスする前にそのデータの利用可能性が容易に分かり、ユーザの効率的な検索行動を実現することができる。

#### 4.2.2 観測項目の統一利用

検索結果サマリ部では各サイトで利用できる観測項目から、3.2節で提示した統一可能性を満たす観測項目の候補を自動的に導き提示を行う。利用者がインターフェース上で観測項目を指定してデータをダウンロードする際には概念スキーマに対して3.4.1節で示した検索を行い、観測項目を統一してデータをダウンロードすることができる。例として、観測項目が図1、図2で表されるデータを、月単位で集約して利用する際の画面を図6に示す。データはXMLやNetCDF、CSVなどの一般的なフォーマットを指定してダウンロードすることができる。

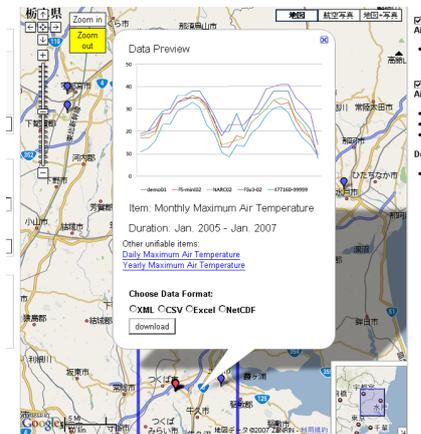


図6 データ取得インターフェース

## 5. まとめと今後の課題

本稿では地球観測データの検索のためのフレームワークとして、オントロジを利用した観測データメタデータの設計を示し、オントロジにより意味を明確にしたことで統合利用可能な観測データをシステムが効率的に認識できる枠組みを提示した。またこの枠組みを利用し、データ検索時にデータの取得可能性を予め提示することによって効率的な検索を実現するインターフェースの紹介を行った。現状では観測点のデータのみを投入・運用しているが、今後は位置情報スキーマの改良を行い衛星観測データ、モデル出力データなど、観測データ位置情報が点ではなく座標系で与えられるグリッドデータの投入への対応を行う。また、観測データの時系列の変換と同様に、位置系列での集約・変換の手法や有用性についても考察していく。将来的には現在の実データ利用の枠組みに加え、検索結果が

ら統合利用のために加工されたデータの履歴や出典等のメタデータを記録しシステムに再投入・活用する枠組みの構築を行うための考察を行っていく。

## 謝 辞

本研究は、文部科学省委託業務研究費国家基幹技術「データ統合・解析システム」の支援を受けており、ここに記して謝意を表します。

## 参 考 文 献

- 1) Fieldserver. <http://model.job.affrc.go.jp/FieldServer/>.
- 2) The geosciences network (geon) project. <http://www.geongrid.org/>.
- 3) Google maps api. <http://www.google.com/apis/maps/>.
- 4) National agricultural research center. <http://narc.naro.affrc.go.jp/>.
- 5) National oceanic & atmospheric administration. <http://www.noaa.gov/>.
- 6) Semantic web for earth and environmental terminology (sweet). <http://sweet.jpl.nasa.gov/ontology/>.
- 7) Semantically-enabled science data integration (sesdi). <http://sesdi.hao.ucar.edu/>.
- 8) Stars web solar-terrestrial data analysis and reference system. <http://www.infonet.cite.ehime-u.ac.jp/STARS/>.
- 9) Virtual solar-terrestrial observatory (vsto). <http://vsto.hao.ucar.edu/>.
- 10) 生駒栄司, 沖大幹, 喜連川優. 内容・空間・時間に基づいた地球環境データ検索インターフェースシステムの構築. 電子情報通信学会データ工学研究会 (福岡) 電子情報通信学会技術研究報告, Vol. Vol.100 No.31, pp. pp.121-128, 2000.
- 11) 根本利弘, 小池俊雄, 喜連川優. 地球水循環データアーカイブシステムにおける異種データ相互解析機能の実装. 電子情報通信学会第 18 回データ工学ワークショップ第 5 回日本データベース学会年次大会 (DEWS2007).