大規模な科学データベースの構築と効率的なデータ検索配信システムの開発

メタデータ	言語: Japanese 出版者: 公開日: 2017-10-05 キーワード (Ja): キーワード (En): 作成者: 高田, 良宏, 笠原, 禎也, 田中, 祥平, 大林, 信, Takata, Yoshihiro, Yoshihiro, Kasahara, Yoshiya, Tanaka, Shouhei, Oobayashi, Makoto メールアドレス: 所属:
URL	https://doi.org/10.24517/00028511

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 International License.

大規模な科学データベースの構築と効率的なデータ検索配信システムの開発

Construction of an Enormous Science Database and Development of an Efficient Data Retrieval and Distribution System

高田　良宏[†]，笠原　禎也[†]，田中　祥平[‡]，大林　信[‡]
Yoshihiro TAKATA, Yoshiya KASAHARA, Shouhei TANAKA, Makoto OBAYASHI

yosihiro@kenroku.kanazawa-u.ac.jp, kasahara@is.t.kanazawa-u.ac.jp
tanaka@cie.ec.t.kanazawa-u.ac.jp, obayashi@cie.is.t.kanazawa-u.ac.jp

† 金沢大学総合メディア基盤センター
Information Media Center of Kanazawa University

† 〒920-1192 石川県金沢市角間町
Kakuma-machi, Kanazawa, Ishikawa 920-1192, Japan

‡ 金沢大学工学部
Faculty of Engineering, Kanazawa University

‡ 〒920-8667 石川県金沢市小立野 2-40-20
2-40-20, Kodatsuno, Kanazawa, Ishikawa, 920-8667, Japan

概要

情報化が進む今日，大学は貴重な情報を蓄積するだけではなく，世界に向けて発信することが求められている．筆者らは，大学内の学術情報を一元的に管理し，情報を提供する総合実験データベースを構築中であるが，その際に問題となるのが，自然科学系の多種多様で，またテラバイト（TByte）にもおよぶ膨大な量のデータの取り扱いである．そこで，TByte オーダーの自然科学データを公開する際のデータベース構築およびデータ検索配信の指針となるべく，本学で蓄積管理している「あけぼの衛星」で計測された膨大な量の電磁波計測データを用い，実際に大規模データベースを構築し，さらにデータの検索配信システムを開発した．その際，データベースの構築から，検索，ネットワークを通した配信までを一つの流れとして捕らえて総合的に検討し，蓄積されている計測データが有用かつ効率的に利用できるシステムとすることを目指した． 本稿では，開発したシステムの紹介を行い，その評価について述べる．

キーワード：大規模データベース，データ配信，SOAP，セキュリティ，宇宙環境計測，あけぼの衛星

Abstract

Along with the development of information society, universities are required not only to accumulate precious academic information but also to open them to the whole world. We are developing a general-purpose database system which manages and provides various kinds of information accumulated in our university. One of the most important points in the design of the system is management of the enormous amount of data, which comes up to several tera-bytes in the field of natural science in some cases. In the present study, we constructed a database system on the geospace radio environment obtained by the Akebono satellite. This is a model case of enormous database systems especially in natural science because there are huge and variety kinds of data on the space environment measurement. In order to improve the system performance, we synthetically examined the flow of functions from construction of the database, data retrieval, and data distribution through the computer networks. In the present paper, we introduce the system and evaluate its performance.

Keyword：Enormous database, Data distribution, SOAP, Security, Space environment measurement, Akebono satellite

1　序論

大学には，世界的にみても価値の高い学術情報が数多く蓄積されている．さらに，大学の研究室では日々実験・計測が行われており，大量のデータが生み出されている．これらの情報をデータベース化し，自らが情報を有効に活用すると同時に，世界に向けて情報発信することは，大学が研究・教育拠点として発展するとともに，その成果を地域貢献・国際貢献につなげるために，大学が早急に進めるべき課題である．

このような背景のもと，筆者らは，大学内の学術情報を一元的に管理し，情報を提供する総合実験データベース[1,2]を構築中であるが，その際に問題となるのが，自然科学系の TByte にもおよぶ膨大な量のデータの取り扱いである．

そこで，本研究では，TByte オーダーの自然科学データを公開する際のデータベース構築およびデータ検索配信の指針となるべく，本学で蓄積・管理している科学衛星「あけぼの」による宇宙電波環境計測に関する膨大な量のデータを用いて，実際に大規模データベースを構築し，さらに効率的なデータの検索配信システムについて検討を行った．研究にあたっては，データベースの構築から，検索，ネットワークを通した配信までを一つの流れとして捕らえて総合的に検討し，蓄積されている大容量データが有用かつ効率的に利用できる手法を開発し，大容量の自然科学データに対し汎用的に利用できることを目指した．

今回データを利用した「あけぼの衛星」とその計測データについて簡単に触れておく．「あけぼの衛星」（EXOS-D）はオーロラおよびオーロラ粒子加速域の精密計測を通して地球周辺宇宙プラズマ環境の理解を主目的として 1989 年に打ち上げられた．「あけぼの衛星」には 8 種類の計測装置が搭載されており，その中でも金沢大学では VLF（低周波波動）計測装置で計測された地球周辺電波に関するデータを取り扱っている．運用開始からすでに 15 年が経過しており，その間に蓄積されたデータはデジタルデータが 1.5TByte（1/3 が VLF データ）以上，アナログデータは DAT テープに音声情報として記録されており，その数は 20,000 本を越えている．これをデジタルデータに変換した場合は約 15TByte 以上となる [3,4].

本論文の構成は以下の通りである．2 章では，本研究で主題とした宇宙環境計測データの管理の現状と問題点を述べ，それを踏まえたシステムの開発条件と研究概要について述べる．3 章では，データベース構築についての詳細と，データの登録方法，参照方法について述べる．続いて 4 章では，今回開発したデータ提供用のメソッドおよび配信システムを紹介する．最後に 5 章で結論をまとめる．

2　宇宙環境計測データの管理

2.1　管理の現状と問題点

宇宙環境の計測手段は，地上に測器を設置して行う地上計測からロケットや科学衛星に計測器を搭載する方法まであり，さらに，非常に多種多様の計測手段を用いるため，データの特性やフォーマット，計測の分解能もそれぞれ異なる．近年では計測技術やデータ伝送技術の進歩により，より詳細な計測データを得られるようになり，1 つのプロジェクトで得られるデータ量が十数 TByte を超えることは珍しいことではない．また，これら多種多様なデータは，計測器を開発した各研究機関が個別に保有・蓄積するのが一般的で，その結果，国内外に分散管理されている．

一方，一般の学術情報データの電子化は広く行われており，種々のデータがオンラインデータベース化されつつある．これにより単一種のデータのみに着目した研究から，各所に分散した異種データの相互比較による研究への移行が進んでいる．しかし，宇宙環境計測分野では，上記のような多様性が原因で，データの提供，相互利用が十分でないのが現状である．

内閣府総合科学技術会議では「地球環境情報の世界ネットワークに関する報告書」の中で「地球環境情報の世界ネットワークに関する現在の問題点とその解決への提案」と題して過去に蓄積された膨大なデータを有効活用するとともに，異なる機関に属する複数の観測装置で計測された膨大なデータの高度利用を効率的に推進できるシステム構築の必要性と，その実現への方向性について議論している[5].

各研究機関が構築したデータベースを扱い，データの参照・解析を目的としたシステムには愛媛大学総合情報メディアセンターが開発中の Solar-Terrestrial data Analysis and Reference System (STARS)あるが，STARS はメタデータベースであり，各研究機関において，公開できる形で保管・管理されている FTP,HTTP でダウンロードできるデータを対象としている[6,7,8].

世界的にみると米国では，米国航空宇宙局(NASA: National Aeronautics and Space Administration)の国立宇宙科学データセンター(NSSDC: National Space Science Data Center)，米国海洋大気庁(NOAA: National Oceanic and Atmospheric Administration)などにおいて，計測データを集中管理している．日本国内では，このような大規模データセンターは，宇宙航空研究開発機構(JAXA:Japan Aerospace eXploration Agency)の地球観測推進センター(EORC: Earth Observation Research and

Application Center)があり，地球環境計測のデータを蓄積・公開している[9]．しかし，宇宙環境計測分野の学術的な科学衛星の計測データは，同じく，JAXA の宇宙科学情報解析センターの DARTS(Data Archive and Transmission System)[10] で蓄積・公開しているが，全科学衛星のデータをカバーするにいたっておらず，多くのデータは，計測機器の開発や，計測に関わった研究室に分散して保管されている．これらの計測データの蓄積管理の責任は研究室にあり，管理を担当する研究者は本来のデータ解析業務以外に，データを安全に管理するために様々な労力を割く必要がある．

2.2 開発環境

2.2.1 サーバ環境

先述の通り，宇宙環境計測データを蓄積管理しているのは，計測機器の開発や計測に関り，データの解析を行っている研究室であることが多いことから，本研究では，計測データを実際に蓄積管理している研究室レベルでのデータベース・サーバ構築を想定した．

一方，公開サーバは研究室レベルの小規模でも運用可能なレベルで，かつ他の多様データとの流通性を上げるため，他のデータの検索システムと共通仕様とする．さらに，1 台の公開サーバが複数のデータベース・サーバからのデータを参照できる設計にし，総合メディア基盤センター（以降センターとする）での集中管理や，他関連機関との連携も容易に行える設計とする．

また，宇宙環境計測データは二度と再現できないデータであるので，データの改竄や悪意に基づくデータ消去など，情報が失われることに対するセキュリティ対策を最重要事項とした．これらを踏まえて，表 1 に示すような開発条件を決定した．

表 1　開発条件

サーバ環境
①データベース・サーバは研究室で管理
・データベース専用マシンなどは導入しない
・データベース環境は PostgreSQL[11,12]などのオープンソースを利用する
②データベース・サーバと公開サーバを切り離す
・公開サーバは共通仕様で作成
・公開サーバは複数のデータベース・サーバから接続可能
・センターでの集中管理も可能であること
③データに対するセキュリティ対策を重点化
配信環境
④SOAP[13,14]を利用
検索速度
⑤一般の検索エンジンでの検索時間と同程度

今回，データベース・サーバとして使用したマシンのスペックは，CPU が Intel® Xeon™ 2.80GHz デュアルプロセッサー，メモリが 2GByte，ハードディスクは内蔵 80GByte，外付け 3.5TByte である．

2.2.2 配信環境

自然科学分野におけるデータの検索システムは，HTML を用いたものが多い．HTML はおもに人間にとっての見栄えを意識したものであり，結果をブラウザに表示するためには有効であるが，データの構造化が不十分であり，データの再利用が困難である．データの解析を行うには検索結果を閲覧するだけではなく，その結果を容易に利用できなければ，検索システムを使用するメリットが低い．そこで今回，データの配信には次に述べる SOAP を利用した．

SOAP(Simple Object Access Protocol)はネットワークに接続したプログラム（オブジェクト）同士が XML のメッセージを交換するための規約である．SOAP は企業間（BtoB），企業顧客間（BtoC）などにおける電子商取引の世界では Web Service を実現するための手段としてかなり以前から利用されている．1999 年に規格化されてから，XML とともに急速に各分野に浸透している．今後は自然科学分野での利用も進むと考える．

SOAP を利用することの利点を次に示す．

- XML を利用
  データが構造的に表現されるため，データの検証，データの再利用が容易である．
- 通信プロトコルに HTTP，SMTP などを利用
  従来のネットワーク環境を変更せずに開発を行うことができる．通常は HTTP を用いるためサーバは，外部から HTTP:80 番ポートに POST アクセス可能ならば設定の変更不要．
- マルチプラットフォーム
  ネットワークを流れるデータ構造だけを定義しているので，OS の種類や解釈するためのプログラムが限定されない．
- アプリケーション間の穏やかな連携
  SOAP による連携は，メッセージを交換するための簡単な仕組みであり，サーバのアプリケーションの仕様変更や機能拡張による影響を最小化できる．

2.3 研究概要

本研究では，「あけぼの衛星」に搭載された VLF（低周波波動）計測装置で計測された数 TByte にも及ぶデータを利用する．今回扱うデータは 14 年分で，衛星の軌道情報においては約 1,600 万件，計測機器の状態に関する情報においては約 9 億件に達する．これらをす

べて公開できることを前提とし，データを登録する際に，分解能を落としたり，データを間引くなどの処理は行わず，すべての情報を登録する．大容量データを全登録しても実用に耐えるよう，作成するデータベースの容量をできるだけ小さく抑える工夫を加える．

次に，検索参照部の開発では，システムをモジュール化し，他のシステムにも再利用できる設計とする．

データの配信では，SOAP を用いてデータの再利用性を重視したデータ配信システム（SOAP サーバ）を作成する．受信用プログラム（SOAP クライアント）も作成し評価を行う．

3　データベース構築と検索参照システム

3.1　システムの概要

本研究で作成したデータベース検索参照システムの概要を図 1 に示す．システムの構成はごく一般的なものとなっている．ユーザインターフェイス及びデータベースとの連携を行う部分には PHP4.3.4[15,16]を採用し開発を進めた．これは CGI などに比べると開発が容易であり，また PostgreSQL との相性が良いためである．データベース・サーバは OS が Linux，RDBMS には PostgreSQL-7.3.4 を用いている．公開サーバは WWW サーバとして Apache2.0.47 を採用した．

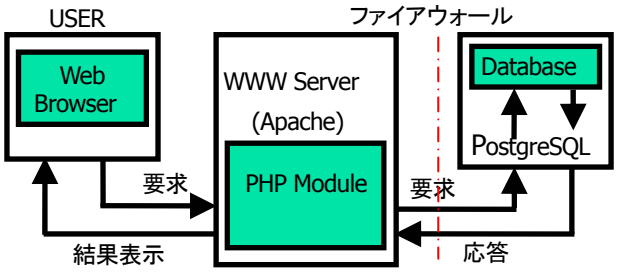

図 1　検索参照システム

サーバは，図 2 に示すように，データベース・サーバと公開サーバは分離して管理する．①データベース・サーバはオリジナルデータを蓄積している研究室で管理，公開サーバはセンターに委託，②両方を研究室で管理，③両方をセンターに委託する，3 つのパターンが考えられる． 通常，データは内容を熟知し，新規登録などを責任もって行える研究室で管理されるべきであるので①と②の構成について実験を行った．最終的には①を推奨構成として実現する予定である．もちろん技術的には③の構成も可能で，自前でのデータ管理が困難な研究室へのサービス提供を想定している．

今回，データベース・サーバを設置する研究室は，他キャンパスの工学部内の研究室である．

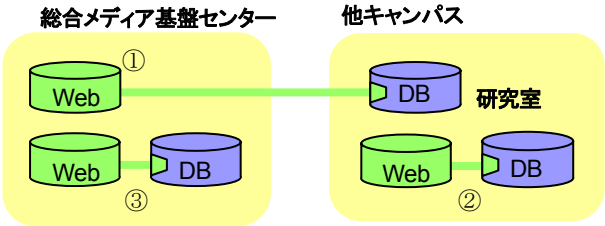

図 2　サーバの分離

3.2　テーブル作成

作成したテーブルは図 3 に示す，衛星の軌道情報，搭載機器の ON/OFF 状態や動作モードなどの情報を表す計測機器の状態(以後ステータスとする)，衛星の周回番号の 3 種類である．今回，1989 年から 2002 年までのデータを登録したが，全てを 1 つのテーブルとする場合と，1 年毎に分割する場合の双方について検討した．その結果，1 つのテーブルとした場合はステータスの件数が 9 億件を超え，検索時間が極めて長くなり，実用上問題が発生するため，1 年毎に分割してテーブル化することとした．

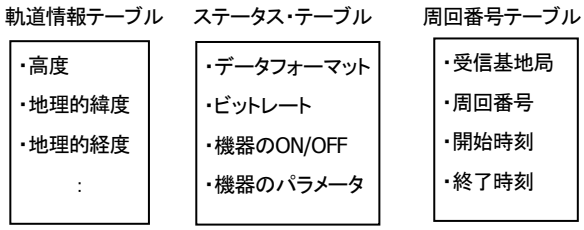

図 3　作成テーブルの概要

軌道情報テーブルは「あけぼの衛星」の軌道情報に関するデータで，30 秒毎に 1 データが入っており，1 年分の件数は約 105 万件となる．定義した列項目は，時間情報，高度，地理的緯度，地理的経度，磁気的緯度，磁気地方時間など全 11 項目である．

ステータス・テーブルは「あけぼの衛星」のステータスに関するデータで，0.5 秒毎に 1 データが入っており，1 年分の件数は約 6,300 万件となる．データは伝送されてきたデータのビットレート，各搭載機器の ON/OFF 状態や動作モードなどが記録されている．

周回番号テーブルは計測データ受信基地局，衛星の周回番号，計測モード，計測開始時刻と計測終了時刻を定義した[17]．

3.3　データの縮小化

先述した通り，データ量が膨大なため，データをそのままデータベースに登録すると，データ量が大きくなり，検索速度の低下，さらにはハードディスクの容量不足などの様々な問題が発生する．そこで，データを登録する際に，データを縮小してから登録する．

軌道情報データでは，緯度や経度などは 4Byte 実数型で記録されているが，値が取りうる範囲，精度から 2Byte 整数に変換できる．具体的には，2Byte 整数の範囲である-32,760～32,760 の値に変換し，2Byte の整数型として登録した．これによって，1 データあたり 14Byte 削減することができた．

ステータスでは, 機器の状態が on/off で表現されるか, 値のとり得る範囲が限定的であるため，0 と 1 のビット列で表現した．実際のデータでは 101Bit で占有長は 17Byte となった．

これらの方法は，自然科学系のデータの縮小に応用できる場合が多々あり，有効な方法と言える．

3.4　検索の高速化

ステータスは, 3.3 節で述べたようにビット列で表現されているため，データ検索ではビット列の演算（論理積）を行う．ビット列の演算は非常に高速であるが，条件にマッチするデータを抽出するのに，毎回 6,300 万件のデータを検索すると，検索時間が非常に長くなる．この問題を解決するため以下のような改善を行った．ステータスは 0.5 秒分解能であるが，状態変化はそれほど速くなく，さらに同じパターンが繰り返されることが多いため，出現するビット列のパターンは限られることに着目し，検索時間を短縮するための新たなテーブルを作成した．このテーブルは図 4 に示すように，各年毎の 1 月 1 日からの通年の日付と，その日に存在するすべてのビット列を登録したものである．新しく作成したテーブルは平均で 20 万件となった．これによって，はじめから 6,300 万件のデータを全検索するのではなく，先ず 20 万件の検索を行い，該当するデータが存在する日付の一覧を取得することにした．検索時間は表 2 に示すように大幅に短縮することができた．

date	Status
3	000100000……0001000101
3	000100000……1001010001
3	000100000……0010001001
:	:

図 4　高速化用のビット列パターンテーブル

表 2　高速化用のテーブルの効果

改善前	改善後
207,363.568ms	383.233ms

さらに検索システム実装時の高速化のために，PHP のコンパイルキャッシュの使用，PostgreSQL のメモリ割り当てパラメータ変更，検索結果の表示方法の工夫などを行った．軌道情報，ステータス，軌道情報とステータスの 2 種類の同時検索で標準的な検索においての検索時間を比較した結果を表 3 に示す．軌道情報とステータスではそれぞれ約 1.5 秒と約 0.7 秒と実用的な値となった．ただし，同時検索においては約 1 分と多少長い時間がかかっている．これは，結合処理が発生するため検索コストが高くなるためであり，テーブル設計や検索方法のさらなる最適化が必要である．

表 3　平均検索時間

	軌道情報	ステータス	同時検索
検索時間	1,544.825ms	744.153ms	65,144.305m
表示件数	124 件	358 件	172 件

3.5　検索画面

作成した検索参照画面のうち，ステータスの検索画面と検索結果の表示例を図 5，図 6 に示す．この画面では, 衛星から伝送されてきたデータのフォーマット，ビットレート，各ステータスを入力し，その組み合わせに一致した日付の一覧が表示される（図 5 の下半分）．次に，一覧より日付を選択すると，その日における該当する時間帯の一覧が表示される．

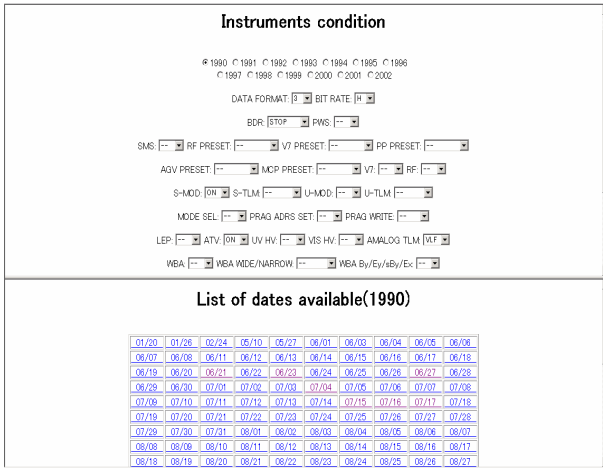

図 5　ステータス検索画面

List of time spans available(1990/06/11)

start	finish
02:03:16.0	02:03:16.0
03:00:23.5	03:00:23.5
06:33:27.0	06:33:27.0
09:21:51.5	09:21:51.5
09:55:58.5	09:55:58.5
16:00:58.5	16:00:58.5
16:07:53.5	16:07:53.5
16:21:36.5	16:21:36.5
19:30:42.0	19:30:42.0
19:48:58.5	19:48:58.5
21:04:37.5	21:04:37.5
21:09:16.5	21:09:16.5
23:00:10.5	23:00:10.5
23:36:33.5	23:36:33.5

図 6　ステータスの検索結果例

3.6　セキュリティ

2.2 節でも述べたが，宇宙環境計測データは二度と再現できないデータであるので，外部からのデータベース・サーバへの不正アクセスを防止し，データの改竄や悪意に基づくデータ消去は絶対に防がなくてはならない．このため，次のようなポリシーでセキュリティの設計を行った．データベース・サーバはファイアウォール内部に設置し，ファイアウォールは内から外への接続要求は通すが，外から内への接続要求はすべて拒否する一方通行の設定とする．このため，外部からの接続要求によるデータベース・サーバとの交信は行えず，データベース・サーバが接続要求を行った相手だけと交信可能状態になる．

実装時にはセキュリティの確保のため 3 種類の対策を実施している．その概要を図 7 に示す[18].

まず入り口部のユーザ認証については，ユーザ名とパスワードによる認証方法を利用する（図 7 の①）．この方法はパスワードが盗聴される危険があるため，HTTPS(HyperText Transfer Protocol Security)と併用する．

次に，Web サーバとデータベース・サーバ間のファイアウォール越しのネットワーク経路のセキュリティ確保については，ポリシーに基づき，SSH(Secure SHell)を用いたポートフォアディングを利用した（図 7 の②）．この機能を利用しデータベース・サーバ側の特定のポートと Web サーバ側の特定のポート間に経路を確立し，その経路上を流れるデータを暗号化している．経路を確立する際は，接続要求はデータベース・サーバ側からのみ行い，公開サーバは接続を許可するだけである．データベース・サーバは一切の接続要求を拒否する設定となっている．Web サーバ上の検索システムにはデータベース･サーバを特定する情報を持たない．

最後は PostgreSQL が持つユーザ認証機能を利用する（図 7 の③）．この認証ではハッシュ関数である MD5 を用いてパスワードを暗号化している．ここでデータベース・サーバにアクセスできるのは②の経路でアクセスするユーザのみである．データベース内で実行できるコマンド（閲覧，挿入，更新，削除）のうち閲覧のみを許可している．

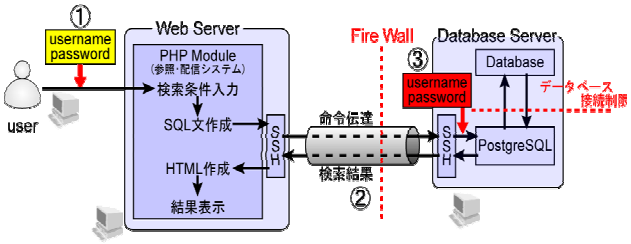

図 7　セキュリティの概要

3.7　考察

大容量のデータを間引かずにデータベースに登録し，さらに検索システムを作成した．

データ量の縮小に関しては，それぞれのデータの特性を生かした形でデータの縮小を行うことができた．この方法は，他の自然科学系のデータへの応用が可能である．

検索時間については，高速化の対策を行った結果，明らかに検索時間が短縮され改善の効果がみられた．その結果，軌道情報の検索，測機器の状態の検索とも実用的な値を得ることができた．このため，4 章で述べる配信システムにも問題なく利用することができる．

4　配信システム

4.1　SOAP による配信システム

前章で構築したデータベース検索システムでは，データを効率的に参照できるようになった．本章では前章で構築したデータベースからの検索結果を参照するだけではなく，データを効率的に配信し，さらにそのデータが有効利用できるように，SOAP を用いた配信システムを開発する．

4.2　データ配信システムの概要

本研究で作成した SOAP を用いた配信システムの概要を図 8 に示す．SOAP サーバはセンターに設置し，データベース・サーバは，３章で使用したものと同一であり，他キャンパスにある工学部研究室のファイアウォール内で管理されている．

クライアントは Web クライアント版のものと，Windows Form 版を作成した．

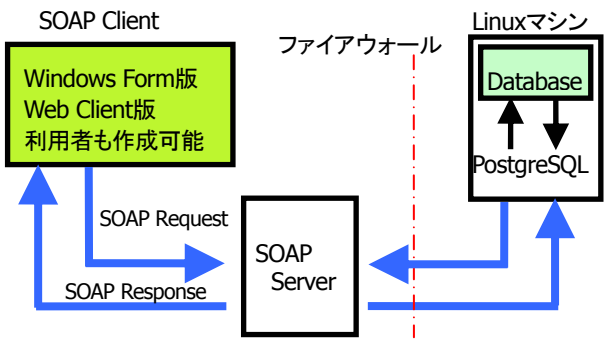

図 8　配信システム

4.3　配信システムの実装

4.3.1　SOAP サーバ

SOAP サーバは図 9 に示すように，2 種類のものを準備してテストを行った．一つは OS が Linux，Web サーバは Apache2.0.47，システムの構築には PHP4.3.4 と PHPSOAP を用いたサーバで，これは図 1 の WWW サーバと同一のものを利用している．もう一つは OS が Microsoft Windows Server 2003，Web サーバは Microsoft Internet Information Services (IIS) 6.0，システムの構築には Microsoft Visual Studio.NET（使用言語は Visual C#.NET）を用いたサーバである[19,20].

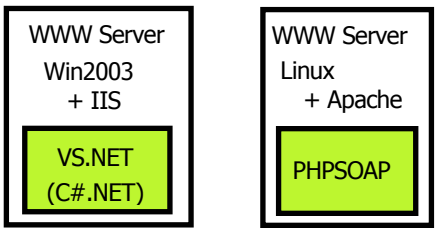

図 9　SOAP サーバ

4.3.2　配信サービスとメソッドの種類

SOAP サーバで動作する２つのサービス（プログラム）を作成した．軌道情報に関するサービスを OrbitService，ステータスに関するサービスを StatusService とした．各サービスには，データの検索・配信のためのメソッドが実装されている．

データベースからデータの検索を行う部分は 3 章で述べた検索参照システムで作成したものと共通であり，これらはモジュール化されている．このため，メソッドは検索用のモジュールを単独またはそれらの組み合わせで呼び出す仕組みを記述すればよく，比較的容易に作成できる．表 4，表 5 に今回作成した OrbitService および StatusService のメソッドを示す．また，作成された WSDL(Web Services Description Language)の一部を図 10 に示す.これは TGetOrbitCount の仕様を記述した部分である．

表 4　OrbitService のメソッド

メソッド名	機　　能
TGetOrbitCount	日付より該当する軌道情報の件数を取得
TGetOrbit	日付より該当する軌道情報を取得
OGetDate	軌道情報より該当する日付を取得
OGetTime	軌道情報と日付より該当する時間を取得．string 形式(HH:MM:SS)
OGetTime2	軌道情報と日付より該当する時間を取得．int 形式(1/10 秒単位)
CheckCredential	認証のみを行う

表 5　StatusService のメソッド

メソッド名	機　　能
TGetStatusCount	日付より該当する機器情報の件数を取得
TGetStatus	日付より該当する機器情報を取得
IGetDate	機器情報より該当する日付を取得
IGetTime	機器情報と日付より該当する時間を取得．string 形式(HH:MM:SS)
IGetTime2	機器情報と日付より該当する時間を取得．int 形式(1/10 秒単位)
CheckCredential	認証のみを行う

```
<s:element name="TGetOrbitCount">
  <s:complexType>
    <s:sequence>
      <s:element name="year" type="s:string" />
      <s:element name="month" type="s:string" />
      <s:element name="day" type="s:string" />
      <s:element name="hour" type="s:string" />
      <s:element name="minute" type="s:string" />
    </s:sequence>
  </s:complexType>
</s:element>
<s:element name="TGetOrbitCountResponse">
  <s:complexType>
    <s:sequence>
      <s:element name="TGetOrbitCountResult" type="s:int" />
    </s:sequence>
  </s:complexType>
</s:element>
```

図 10　WSDL の一部

4.3.3　SOAP メッセージの例

メソッドに対してデータの要求（リクエスト）メッセージを送信した場合の例を図 11 に示す．また，メソッドが受信したデータの要求に対する応答（レスポン

ス）メッセージの例を図 12 に示す．これらの例では，TGetOrbit メソッドに対して，1990/04/08 の 12:00:00 から 12:59:59 までの軌道情報を要求している．TGetOrbit メソッドからは，要求に該当する軌道情報が返される．

```
<?xml version="1.0" encoding="utf-8"?>
<soap:Envelope
      ... 省略 ...
  </soap:Header>
  <soap:Body>
   <TGetOrbit xmlns="http://websrv.gipc.kanazawa-u.ac.jp/
      <year>1990</year>
      <month>04</month>
      <day>08</day>
      <hour>12</hour>
      <minute></minute>
      <limit></limit>
    </TGetOrbit>
  </soap:Body>
</soap:Envelope>
```

図 11　リクエスト SOAP メッセージ

```
<?xml version="1.0" encoding="utf-8" ?>
     ... 省略 ...
    <xs:element name="id" type="xs:int" minOccurs="0" />
    <xs:element name="alt" type="xs:decimal" minOccurs="0" />
    <xs:element name="glat" type="xs:decimal" minOccurs="0" />
    <xs:element name="glon" type="xs:decimal" minOccurs="0" />
    <xs:element name="mlat" type="xs:decimal" minOccurs="0" />
    <xs:element name="mlt" type="xs:decimal" minOccurs="0" />
    <xs:element name="flat" type="xs:decimal" minOccurs="0" />
    <xs:element name="ilat" type="xs:decimal" minOccurs="0" />
    <xs:element name="x" type="xs:float" minOccurs="0" />
    <xs:element name="y" type="xs:float" minOccurs="0" />
    <xs:element name="z" type="xs:float" minOccurs="0" />
      ... 省略 ...
  <Table diffgr:id="Table1" msdata:rowOrder="0">
    <id>280801</id>
    <alt>5169.963</alt>
    <glat>28.497</glat>
    <glon>67.984</glon>
    <mlat>20.022</mlat>
    <mlt>16.600</mlt>
    <flat>41.632</flat>
    <ilat>45.533</ilat>
    <x>978.084</x>
    <y>10106</y>
    <z>5491.73</z>
  </Table>
  <Table diffgr:id="Table2" msdata:rowOrder="1">
     <id>280802</id>
     <alt>5095.788</alt>
     <glat>27.734</glat>
     <glon>68.121</glon>
     <mlat>19.250</mlat>
     <mlt>16.609</mlt>
   ... 省略 ...
```

図 12　レスポンス SOAP メッセージ

4.4　クライアントの実装

クライアントはデータの管理者が作成し配布する場合も考えられるが，本来は利用者がその利用形態に応じて最も利用しやすい形で作成するものである．しかし，今回は配信システムの構築とその評価を目的としているため， Web クライアント版のものと，Windows Form 版を作成した．作成したクライアントの例として，Windows Form 版の軌道情報用画面を図 13，図 14 に，ステータス用の画面を図 15，図 16 に示す．

図 13 および図 15 では計測日付と時間を入力することにより，衛星の軌道情報やステータスを得られる．図 14 および図 16 では軌道情報やステータスを入力することにより，該当する日付と時間帯を得られる．

今回作成したクライアントでは，データを取得し，そのデータを表示およびテキストデータ（CSV 形式）として保存するところまでを実装した．また，研究者が従来から使用している解析プログラムを組み込むことは容易であり，研究者の利用形態に応じたクライアントを作成することができる．

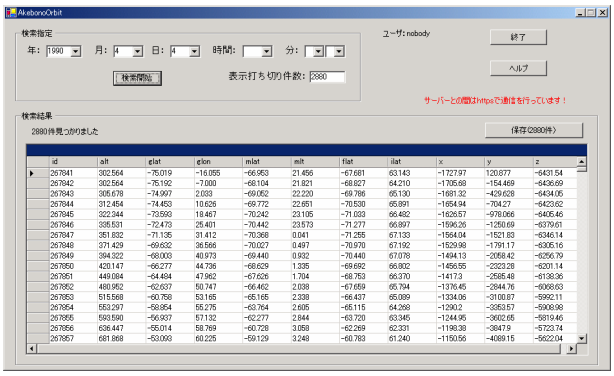

図 13　日付情報から軌道情報を取得

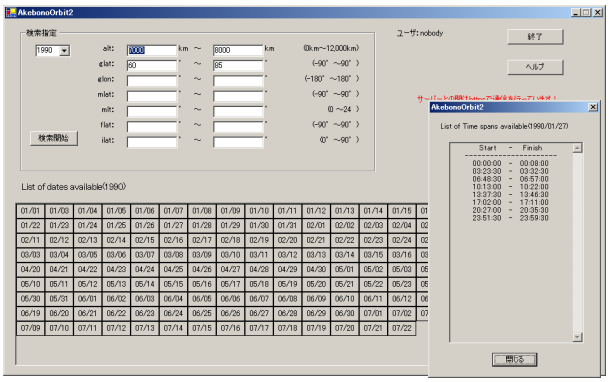

図 14　軌道情報から日付情報を取得

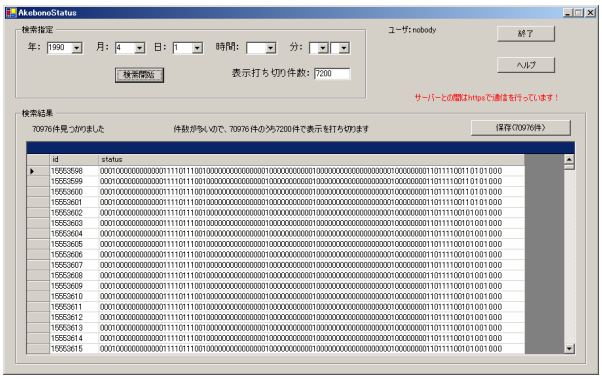

図 15　日付情報からステータスを取得

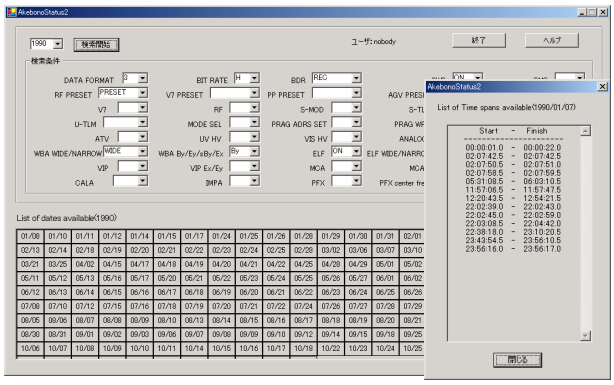

図 16　ステータスから日付情報を取得

4.5　セキュリティ

貴重なデータを取り扱うため，データの配信を行う場合も，セキュリティの確保が重要である．データベース・サーバ，SOAP サーバ間とデータベースのセキュリティは3.6節で述べた方法がそのまま適用される．ここでは，SOAP サーバと SOAP クライアント間のセキュリティの確保について述べる．

SOAP はフォームなどのユーザインターフェイスを持たないため，SOAP ヘッダーを用いたユーザ名とパスワードによる認証方法を利用する．そのときの SOAP メッセージの例を図 17 に示す．この例は認証のみを行う場合のものであり，これを用いた認証用のダイアログボックスを図 18 に示す．　ただし，この方法はパスワードが盗聴される危険があるため，HTTPS と併用した．

認証のためのリクエストメッセージ例

```
<?xml version="1.0" encoding="utf-8"?>
.... 省略 ...
<soap:Header>
  <Authentication
        xmlns="http://websrv.gipc.kanazawa-u.ac.jp/OrbitService/">
    <User>takata</User>
    <Password>Mao1192S1</Password>
  </Authentication>
</soap:Header>
<soap:Body>
<CheckCredential
```

認証のためのレスポンスメッセージ例

```
<?xml version="1.0" encoding="utf-8"?>
.... 省略 ...
<soap:Body>
  <CheckCredentialResponse
        xmlns="http://websrv.gipc.kanazawa-u.ac.jp/OrbitService/">
    <CheckCredentialResult>true</CheckCredentialResult>
  </CheckCredentialResponse>
</soap:Body>
```

図 17　SOAP ヘッダーを用いた認証メッセージ

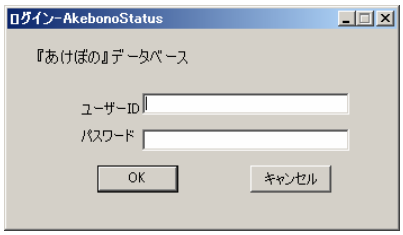

図 18　認証用ダイアログボックス

4.6　考察

SOAP サーバおよびクライアントを作成し，データの配信を行うことが可能となった．今回，データ取得のためのメソッドを十数種類作成したが，メソッドの作成は検索用のモジュールを単独またはそれらの組み合わせで呼び出す仕組みを記述すればよく，共同研究者からの新たな要望があれば，容易にメソッドを追加することができる．また，作成されたメソッドの仕様を公開すれば，共同研究者が容易にクライアントを作成することができ，これにより，データの要求，取得，解析，グラフ化といった一連の作業を準リアルタイムで行うことが可能になる．

今回作成した配信サービスでは個々の計測データを表す XML の要素名やメソッドの命名法が「あけぼの衛星」の独自仕様であり，今後，XML の利点を活かして，地上計測データを含む多様性に対応できる標準化を進める必要がある．

データは XML で配信されるため，付加されるタグ分のメッセージ量が増加し，通信量は配信する計測データの 2 倍程度となった．しかし，現在，ほとんどの大学や研究所間は 100Mbps 以上の速度が確保されているため大きなインパクトはないと考えられる．

セキュリティ面では SOAP ヘッダーを利用した認証と HTTPS を併用することにより，安全にデータを配信することができた．

5　結論

5.1 まとめ

本研究では，膨大な量の宇宙環境計測データから大規模データベースを構築し，さらに各地に点在している共同研究者に対して効率的にデータの配信を行うための検索・配信システムを開発した．

データをデータベースに登録する際に，データの特性を生かし，出来るだけデータを縮小することにより，データを間引かずにすべてのデータを登録することができた．登録の際に用いたデータ縮小の方法は，他の

自然科学系のデータへの応用が可能である．大規模データベースから，高速かつ効率的にデータを検索するために，検索プログラムのコンパイルキャッシュの使用，メモリの割り当て変更，検索結果の表示方法の改善，さらに高速化用テーブルの作成などを行った．この結果，実用に耐える十分な検索速度を得ることができた．

大容量のデータを効率的に配信する方法について検討し，SOAP を用いた配信システムを開発した．十数種類のメソッドを作成し，軌道情報，ステータスのデータを効率よく配信できるようになった．さらに検索用のプログラムをモジュール化したことにより，共同研究者の要望に対して容易にメソッドを追加できるようになった．メソッドの仕様を公開すれば，共同研究者が容易にクライアントを作成できるため，よりいっそうデータを有効に利用できると考える．

システムの開発を進める際に，常にセキュリティの確保を重視し，ユーザ認証，SQL コマンドの使用制限，SSH ポートフォアディングによるサーバ間の通信の暗号化，HTTPS を用いたサーバとクライアント間の暗号化などにより，一貫したセキュリティ対策を行った．

今回開発したシステムは，十分実用的に機能するものであり，また，研究室レベルの環境での公開という当初の目標も達成している．さらに，「あけぼの衛星」のデータ特有の特性をあまり使っていない汎用的な仕様であるため，多様なデータに対応可能であり，更に検討を行わなければならないが，他種の TByte オーダーの計測データにも応用できると期待できる．

5.2　今後の課題

宇宙環境計測データの配信をより効率の良いものとするために，各機関と共同で宇宙環境計測データ用の XML 文書定義の標準化やデータ取得のためのメソッドの共通化を進めることを目標にする．また，「あけぼの衛星」の十数 TByte におよぶ計測データを用いて，さらに大規模なデータに適用するときの問題点の洗い出しと，その解決法を検討する必要がある．

さらに，本研究で開発したシステムが，一般の自然科学データに対し汎用性が高いことから，宇宙環境計測分野と同様の問題を持っている異種の分野のデータへの適用を試みたい．

参考文献

[1] 笠原 禎也, 金沢大学における実験データベースの構築, 国立情報学研究所 学術情報ネットワーク(スーパーSINET/SINET) 成果報告集, pp.221-228, 2004.

[2] 金沢大学総合メディア基盤センター, COM.CLUB, Vol.27, 2003.

[3] 山本 正幸, 木村 磐根, 長野 勇, 橋本 弘藏, 岡田 敏美, 林 幹治, 澤田 晃, 伊藤 嘉彦, 軒内 栄一, 岸 洋介, 笠原 禎也, あけぼの(EXOS-D)搭載 VLF 装置による初期観測結果, 宇宙科学研究報告, 第 25 号, 1990.

[4] I.Kimura, K.Hashimoto, I.Nagano, T.Okada, M.Yamamoto, T.Yoshino, H.Matsumoto, M.Ejiri and K.Hayashi, VLF Observations by the Akebono (EXOS-D) Satellite, 42(4), pp.459-478, J.Geomag.Geoelectr, 1990.

[5] 内閣府総合科学技術会議,地球環境情報の世界ネットワーク構築に関する報告書, 2003.

[6] 村田 健史, 岡田 雅樹, 阿部 文雄, 荒木 徹, 松本 鉱, 太陽地球系物理観測の分散メタデータベースの設計と評価, 情報処理学会論文誌:データベース, pp.115-130, Vol.43, No.SIG 12(TOM16), 2002.

[7] 矢原 大幹, 村田 健史, 松本 鉱, 人工衛星・地上観測データ解析参照システムの構築, 電子情報通信学会論文誌:通信技法,pp.9-16, Vol.100, No.38, 2000.

[8] 愛媛大学 STARS(Solar-Terrestrial data Analysis and Reference System) HOME PAGE, http://www.infonet.cs.ehime-u.ac.jp/stars/

[9] 宇宙航空研究開発機構(JAXA) 地球観測推進センター(EORC) HOME PAGE, http://www.eorc.jaxa.jp/

[10] 宇宙航空研究開発機構(JAXA) DARTS(Data Archive and Transmission System) HOME PAGE, http://www.darts.isas.ac.jp/

[11] Richard Stones, Neil Matthew, 株式会社スリー・エー・システムズ監訳, PostgreSQL 活用テクニック, 株式会社インプレス, 2002.

[12] 石井 達夫, PostgreSQL 完全攻略ガイド, 株式会社技術評論社, 2003.

[13] 日本アイ・ビー・エム 株式会社 jStart チーム著, 最新 Web サービスがわかる,株式会社技術評論社, 2002.

[14] ケナード・スクリブナー, マーク・C・スタイバー著, 株式会社スリー・エー・システムズ訳, SOAP 技術入門, 株式会社ピアソン・エデュケーション, 2001.

[15] 堀田 倫英, 石井 達夫, 廣川 類, PHP4 徹底攻略 改訂版, ソフトバンク パブリッシング株式会社, 2002.

[16] 廣川 類, 桑村 潤, 小山 哲志, PHP4 徹底攻略 実線編, ソフトバンク パブリッシング株式会社, 2002.

[17] 大林 誠, 高田 良宏, 笠原 禎也, 後藤 由貴, 大規模な科学データベースからの情報検索・配信の高効率化, 地球惑星関連学会合同大会予稿集, http://www-jm.eps.s.u-tokyo.ac.jp/2004cd-rom/pdf/j031/j031-002.pdf, 2004.

[18] 田中 祥平, 笠原 禎也, ネットワークを利用したあけぼの衛星 VLF 波動データ参照・配信システム, 地球惑星関連学会合同大会予稿集, http://www-jm.eps.s.u-tokyo.ac.jp/2003cd-rom/pdf/j031/j031-p005.pdf, 2003.

[19] G.Andrew, Duthie 著, (株)クリープ訳, マイクロ公式解説書 ASP.NET 実践講座, 日経 PB ソフトプレス, 2002.

[20] arton 著, Visual C#.NET による Web プログラミング入門, 株式会社アスキー, 2002.