

カタログサーバの仕組みと CSIS における運用

白石陽・高橋昭子（東京大）・山下重紀郎（酪農学園大）・

佐藤 英人（東京大）・大橋智美（筑波大・院）

1. はじめに

東京大学空間情報科学研究センター（以下 CSIS）では、これまでクリアリングハウスを運用してきたが、いくつかの問題点が指摘されている(高橋ほか, 2004)。そこで、CSIS では、文部科学省科学研究費補助金基盤研究（A）「大学間連携分散自律型・地理データ基盤システムの開発研究」（岡部ほか, 2005）の成果として、新しいメタデータ公開サーバである「カタログサーバ」を開発した（白石ほか, 2005）。カタログサーバは、従来のクリアリングハウスとは異なり、複雑な検索パラメータの選択を行うことなく、キーワード検索および地図検索というシンプルな操作方法を提供することにより、ユーザにとって使いやすい検索インタフェースを提供している。また、メタデータ作成者の負担を軽減するため、カタログサーバに登録するメタデータは最低限のものとして、カタログサーバから各提供機関がそれぞれ別途作成した詳細情報ページへのリンクを張ることにより、分散型の空間データ基盤を構成する。本稿では、カタログサーバの概要と CSIS における運用について報告する。

2. カタログサーバの仕組み

2.1 カタログシステムにおけるカタログサーバの位置付け

カタログサーバは、カタログシステム（高橋ほか, 2006）の中心的なシステムであり、国際的な記述標準の一つである FGDC 形式のメタデータを管理し、ユーザの入力に応じてメタデータの検索を行う。

各データ提供機関は、カタログサーバが提供する最低限のメタデータを説明する詳細情報（本稿では、「カタログ情報」と呼ぶ）を作成・公開、および、空間データの公開・提供を担う（図 1）。前者の機能を持つサーバをカタログ情報公開サーバ、後者を空間データ提供サーバと呼ぶが、これらのサーバは、データ提供機関ごとに設置されるものとする。CSIS では、空間データ共有システムが空間データ提供サーバに相当する。

カタログサーバは、各データ提供機関が提供するそれらの空間データに関する最低限のメタデータのみを集中的に管理し、データ閲覧者に提供するものとする。利用者は、カタログサーバの検索機能を利用して、空間データに関する情報を取得する。

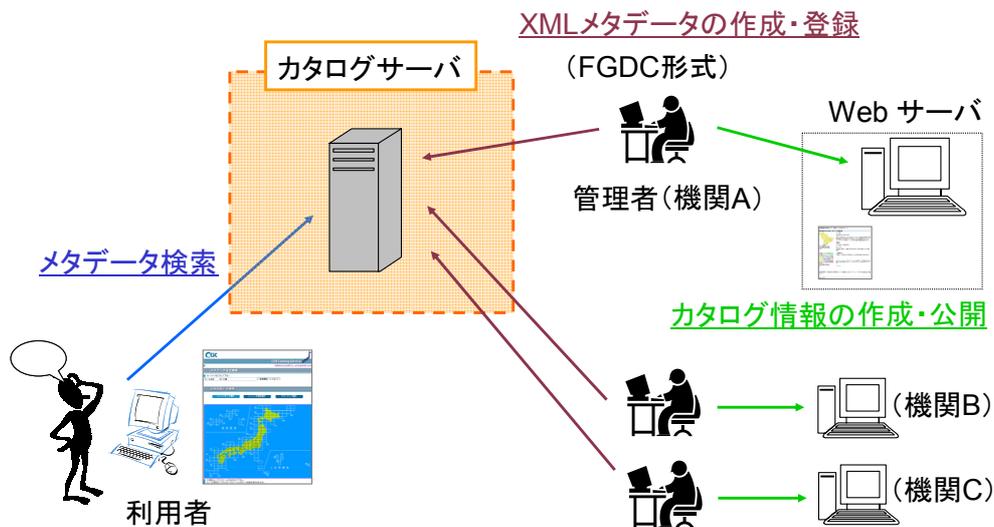


図 1 カタログサーバの構成

2.2 メタデータ項目とデータベース格納情報

カタログサーバでは、空間データ提供機関ごとにディレクトリを作成し、各ディレクトリの中で、各機関の提供する空間データのメタデータを管理する。カタログサーバは、指定したディレクトリ以下に保存されている FGDC 形式のすべてのメタデータを読み込み、データベースにテーブルとして格納する。データベースのテーブルの属性は、FGDC の項目 (XML タグ) に対応し、項目ごとに、抽出された XML タグの値が登録される。テーブルの属性および対応する FGDC の項目としては、例えば、以下のものが挙げられる。

- ・ 項目名「データ名称」
空間データの名称を表す。
FGDC : /metadata/idinfo/citation/citeinfo/title
- ・ 項目名「問い合わせ先」
空間データを提供する機関などの問い合わせ先を表す。
FGDC : /metadata/metainfo/metc/continfo/contorgp/contorg
- ・ 項目名:「詳細情報ページの URL」
空間データのメタデータの詳細情報ページ (カタログ情報) の URL を示す。
FGDC : /metadata/idinfo/citation/citeinfo/onlink
- ・ 項目名:「サムネイル画像の URL」
空間データのサムネイル画像の URL を示す。
FGDC : /metadata/idinfo/browse/brosen

- ・ 項目名「空間データの空間範囲」

空間データの空間的な範囲を示し、東西南北の端点の経緯度の組で表される。

FGDC : /metadata/idinfo/spdom/bounding/{eastbc,westbc,southbc,northbc}

- ・ 項目名「メタデータ全文テキスト」

FGDC 形式のメタデータのうち、XML タグを除いたテキストデータである。キーワード指定によるメタデータ全文検索のためにデータベースに登録される。

2.3 メタデータの登録手順

以下の手順にしたがって、カタログサーバにメタデータを登録する。

① XML メタデータの作成

FGDC 形式のメタデータを作成する。

② XML メタデータのアップロード

FGDC 形式のメタデータをカタログサーバの所定のディレクトリにアップロードする。

③ XML メタデータの登録

所定のディレクトリのすべてのメタデータをデータベースに登録する。

3 カタログサーバの検索機能

まず、図 2 にカタログサーバの検索画面ページを示す。

カタログサーバは、次の 2 つの検索機能を提供する。

- メタデータ全文検索

指定したキーワードを用いて、メタデータの全文検索を行う。

- 地図インタフェースを利用した検索

地図上に表示された標準地域メッシュ（一次メッシュ）をクリックすることで、関連するメタデータを探し出す。

3.1 メタデータ全文検索

3.1.1 検索キーワードの入力

ユーザのキーワード入力に対して、カタログサーバは、SQL を用いてデータベースに問合せを行い、データベース中の「メタデータ全文テキスト」項目の情報と入力されたキーワードを比較して、該当するメタデータを検索する。例えば、図 3 に示すように、検索画面中のテキストボックスにキーワード「ZmapTown」を入力し、「検索開始」ボタンを押すと、その結果として、カタログサーバに登録されている ZmapTown II のメタデータが検索結果一覧として表示される（図 4）。

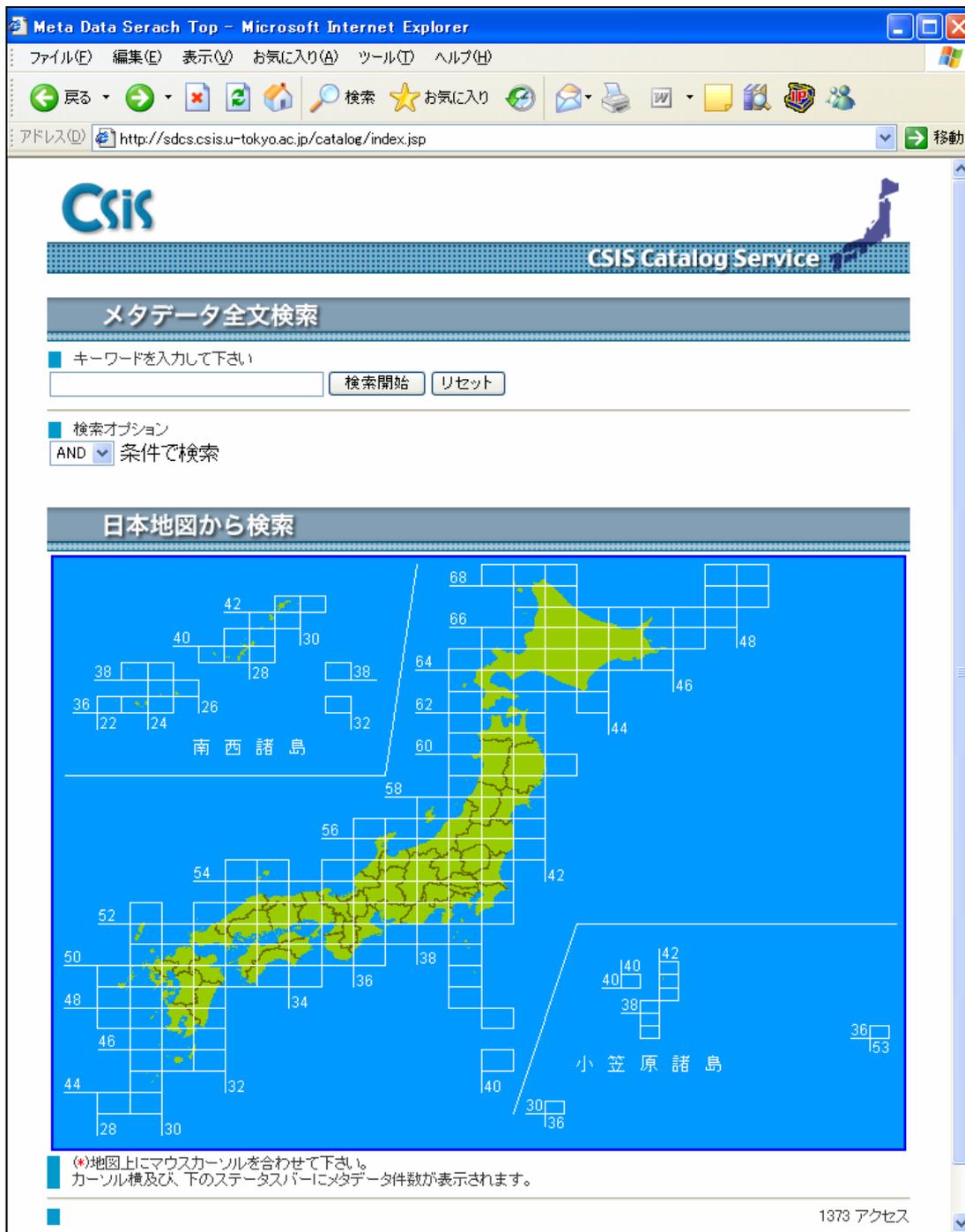


図 2 カタログサーバの検索画面

メタデータ全文検索

■ キーワードを入力して下さい

ZmapTown

■ 検索オプション

AND

図 3 キーワード入力による全文検索

メタデータ検索結果一覧

■ 検索結果件数設定

10

検索結果: 631件のデータを検出しました。10件ごと“AND”条件で表示します。

1.



データ名称: ZmapTownII(旭川市)
 問い合わせ先: 東京大学空間情報科学研究センター
[詳細情報のページへ](#)

2.



データ名称: ZmapTownII(石狩郡当別町)
 問い合わせ先: 東京大学空間情報科学研究センター
[詳細情報のページへ](#)

図 4 検索結果一覧の表示例 (キーワード「ZmapTown」)

3.1.2 検索結果一覧の表示

検索結果一覧のページでは、検索結果の総数と、各メタデータ（すなわち空間データ）に関する最低限の情報が提示される。すなわち、前述した FGDC 項目のうち、「データ名称」「データ提供機関」「詳細情報ページへのリンク」を表示している。これらの FGDC の項目がメタデータとして入力されていない場合には、検索結果一覧のページで表示されない

め、これらのメタデータ項目の入力は必須としている。さらに、データの範囲や概要を示すサムネイル画像を画面左側に表示している。図 4 では、キーワードに「ZmapTown」を指定した場合の検索結果一覧の例を示しているが、図 5、図 6 に、それぞれ、キーワードに「数値地図」、「国勢調査」を入力した時の検索結果一覧の例を示す。

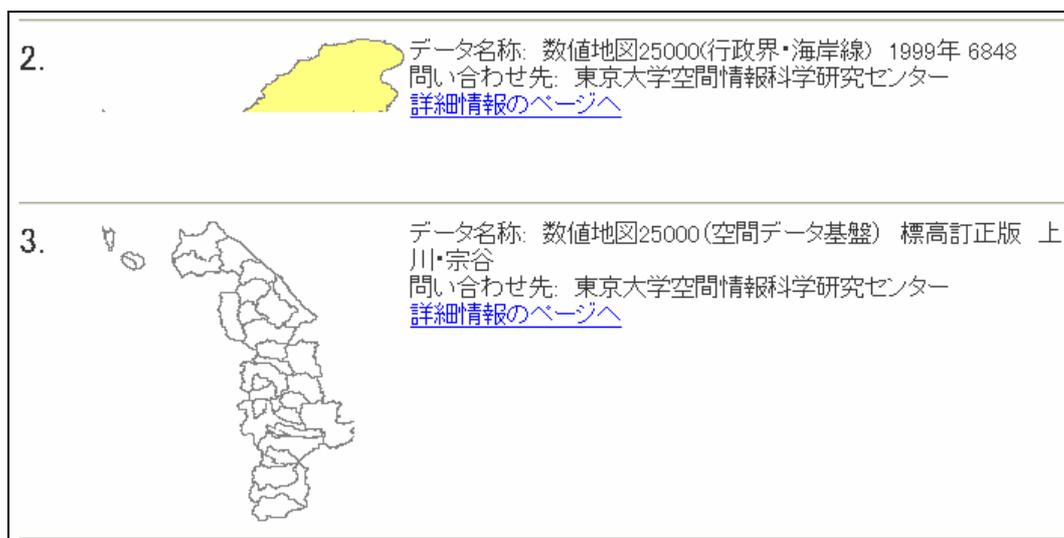


図 5 検索結果一覧の表示例（キーワード「数値地図」）



図 6 検索結果一覧の表示例（キーワード「国勢調査」）

3.1.3 詳細情報ページの表示

図4の検索結果一覧のページの「詳細情報のページへ」というリンクをクリックすると、そのメタデータの詳細な情報を閲覧することができる(図7)。詳細情報のページは、空間データ提供機関ごとに作成され、各機関のウェブサーバにて公開される。図7は、CSISの管理している空間データに関する詳細情報のページであり、データ名、座標系、キーワード、利用方法など、より詳細な情報が表示されている。さらに、ページ下部の「Download」をクリックすると、CSISの空間データ共有システムのページに進むことができる。



図7 詳細情報のページの例(データ名称「ZmapTown II(旭川市)」)

3.1.4 検索オプションの指定

テキストボックスにスペース区切りで複数のキーワードを入力した場合には、検索オプションによって、AND 検索、OR 検索のいずれかを選択できる。

- AND 検索

指定したすべてのキーワードを全文テキストに含むメタデータを探し出す。

- OR 検索

指定したキーワードのいずれかが全文テキストに含まれるメタデータを探し出す。

例えば、テキストボックスに「千葉」と「空間データ基盤」を入力し、AND 検索を行うと、図 8 のように、これらのキーワードをともに含むメタデータが検索結果一覧として表示される。最初の検索結果のデータ名称には、「空間データ基盤」は含まれていないが、FGDC の別項目として入力されているため、全文検索の結果として表示されている。

検索結果: 5件のデータを検出しました。10件ごと“AND”条件で表示します。	
1.	 <p>データ名称: 数値地図2500(茨城・千葉1) 問い合わせ先: 東京大学空間情報科学研究センター 詳細情報のページへ</p>
2.	 <p>データ名称: 数値地図2500(空間データ基盤) 世界測地系版 千葉 問い合わせ先: 東京大学空間情報科学研究センター 詳細情報のページへ</p>
3.	 <p>データ名称: 数値地図2500(空間データ基盤) 標高訂正版 千葉 問い合わせ先: 東京大学空間情報科学研究センター 詳細情報のページへ</p>

図 8 AND 検索の結果

3.2 地図インタフェースを利用した検索

検索画面の下部では、日本列島の地図画像に対して、一次メッシュが重ね合わせ表示されている（図 9）。図中で、メッシュの端に表示されている数字は、メッシュ番号を表している。マウスカursorをメッシュ上に移動させると、そのメッシュと空間範囲が重なっている空間データの件数が表示される。図 9 を見ると、マウスカursorが「5339」のメッシュを指している時に、そのメッシュと空間範囲が重なっている空間データのメタデータが 291 件登録されていることを表しているのがわかる。

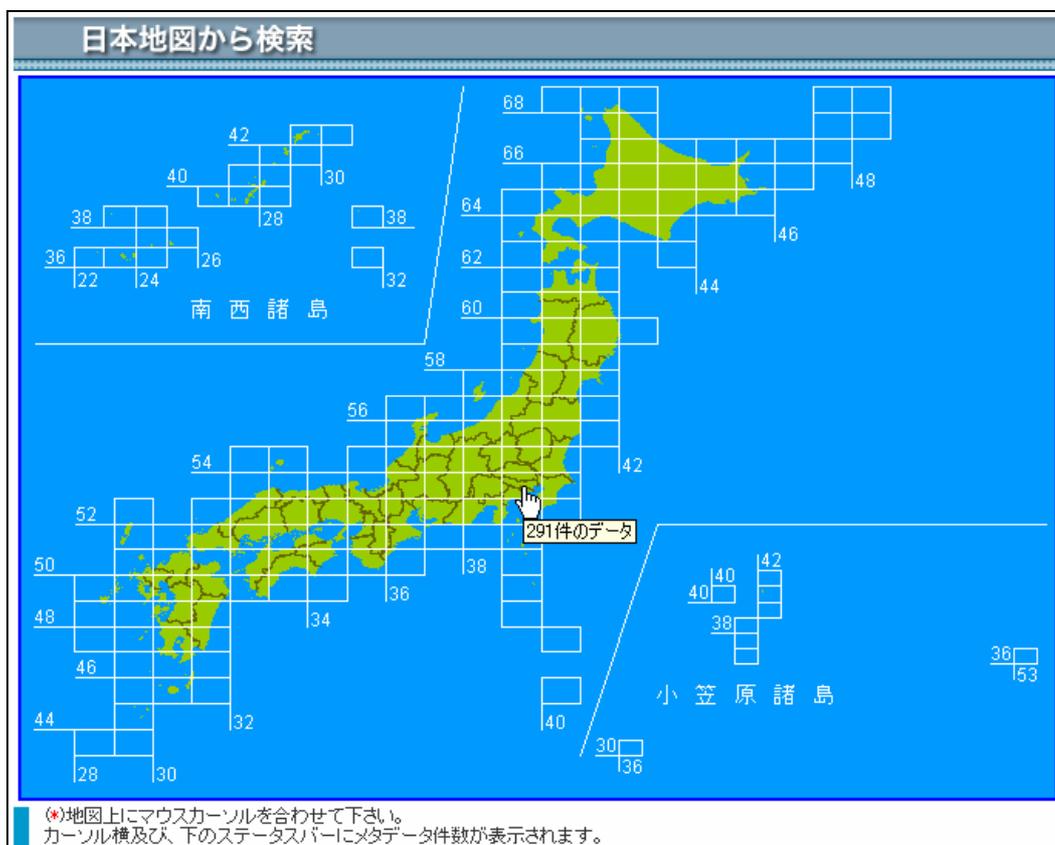


図 9 地図インタフェースを用いた検索

地図上でマウスカursorの指しているメッシュをクリックすると、それらの空間データが検索結果一覧として表示される。図 10 は、その検索結果一覧の一部であり、「5339」のメッシュと重なる空間データに関するメタデータが表示されていることがわかる。

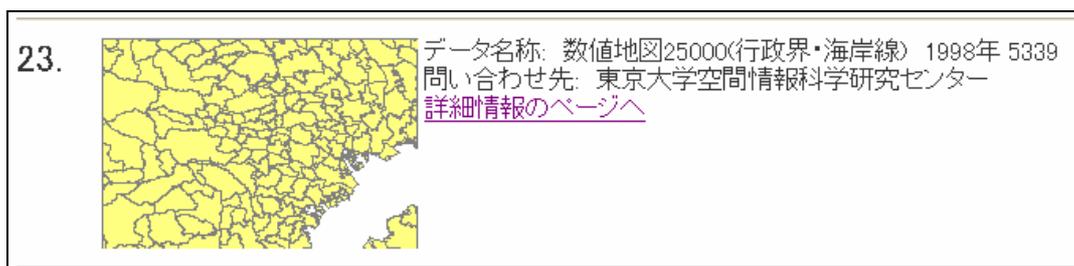


図 10 地図インタフェースを用いた検索の結果表示

4. CSIS におけるカタログサーバの運用

4.1 CSIS カタログサーバの公開

CSIS では、2005 年 5 月より、下記 URL にて、カタログサーバを公開している。

http://www.csis.u-tokyo.ac.jp/japanese/service/cs_top.html

「空間データ利用を伴う共同研究」で利用できる空間データに関するメタデータを登録する作業を進めて来たが、2005 年 5 月の時点で、数値地図シリーズ、国勢調査シリーズ、ZmapTownII シリーズに含まれる利用頻度の高い空間データ 2044 件についてカタログサーバへの登録作業を完了している。その後も、新たな空間データに関するメタデータ登録作業を進めた結果、2006 年 5 月での登録データ件数は 13981 件となっている。利用者は、カタログサーバの検索結果として最低限のメタデータを閲覧できるだけでなく、その説明情報であるカタログ情報を閲覧することができる。

4.2 動作環境とフォルダ構成

CSIS カタログサーバの動作環境を表 1 に示す。

表 1 CSIS カタログサーバの動作環境

プログラム言語および処理系	Java (j2sdk1.4.2_05)
ウェブサーバ	Apache 2.0.40
データベース	PostgreSQL 7.3.4
OS	Red Hat Linux 2.4.20-8

カタログサーバでは、空間データ提供機関ごとにディレクトリを作成し、各ディレクトリの中で、各機関の提供する空間データのメタデータを管理する。CSIS カタログサーバでは、CSIS で管理している大量の空間データを対象としているため、そこで、CSIS 用のメタデータ保存ディレクトリ（例えば、`catalogsrv/public_xml.src/csis/`）を作成した。さらに、

JMP2.0 形式のデータを FGDC 形式に変換する作業が必要になるため、次の 2 つのディレクトリを作成した。

- ・ JMP2.0 保存用ディレクトリ (catalogsrv/public_xml.src/csis/jmp2.0)
アップロードされた JMP2.0 メタデータを階層的に管理する。
- ・ FGDC 保存用ディレクトリ (catalogsrv/public_xml.src/csis/fgdc)
JMP2.0 より変換された FGDC メタデータを階層的に管理する。

これらのディレクトリは、CSIS の提供する基幹サービスの一つである空間データ共有システムにおけるデータ管理構造と同じ階層構造を持つようにしている。さらに、カタログ情報を示す HTML ファイルを保存するディレクトリも同じ構造を持つようにしている。なお、上記のディレクトリパスは、相対パス表示である。

4.3 メタデータの登録手順

CSIS で管理しているメタデータについて、国土地理院のメタデータエディタなどのツールを用いて、JMP2.0 形式のデータとして編集作業を進めている。したがって、JMP2.0 形式のメタデータを FGDC 形式に変換する必要がある。その変換手順については、後述する。なお、個別のメタデータの作成および登録に関する詳細については、山下ほか (2006)、佐藤ほか (2006)、高橋ほか (2006) を参照されたい。

CSIS で管理している大量の空間データについては、まず JMP2.0 形式のメタデータを作成した後、FGDC 形式のメタデータに変換する。CSIS においては、この変換作業をカタログサーバで行うものとして、JMP2.0 形式の XML メタデータをカタログサーバにアップロードするものとする。したがって、2.3 で述べた登録手順とは、若干異なる手順が必要となる。具体的には、以下の手順で、メタデータの登録およびカタログ情報の登録を行う。

- ① XML メタデータ (JMP2.0 形式) のアップロード
- ② カタログ情報のアップロード
- ③ XML メタデータの変換
- ④ XML メタデータ (FGDC 形式) のデータベースへの登録

4.3.1 メタデータのアップロード

JMP2.0 形式のメタデータをカタログサーバにアップロードする。CSIS では、JMP2.0 メタデータエディタを利用したメタデータ作成作業を Windows 環境で行っているため、WinSCP などのファイル転送ツールを利用してカタログサーバマシンへのアップロードを行う。

4.3.2 カタログ情報のアップロード

カタログ情報公開サーバに対して、詳細情報ページの HTML ファイルとサムネイル画像のファイルのアップロードを行う。現在の運用では、便宜上、カタログ情報公開サーバを、カタログサーバと同一のマシンとしている。

4.3.3 メタデータの変換

カタログサーバマシン上で、XSL 変換 (XSLT: XSL Transformations) を用いて、JMP2.0 形式から FGDC 形式への XML データの変換を行う。JMP2.0 と FGDC の各タグの対応関係 (表 2) を変換ルールとして記述し、XML 処理系に入力として与えることで、メタデータの変換を行うことができる。具体的には、国土地理院が配布しているクリアリングハウスノードサーバ構築キットに含まれている XSL ファイル (jmp20_2_fgdc.xsl) を変換ルールとして利用した。さらに、指定したディレクトリ中のすべての JMP2.0 メタデータを FGDC 形式に変換するための一括変換スクリプトを作成し、変換作業の効率化を図った。

表 2 FGDC タグと JMP2.0 タグの対応表

データ名称	
FGDC	/metadata/idinfo/citation/citeinfo/title
JMP2.0	/MD_Metadata/identificationInfo/MD_DataIdentification/citation/title
問い合わせ先	
FGDC	/metadata/metainfo/metac/cntinfo/cntorgp/cntorg
JMP2.0	/MD_Metadata/contact/organisationName
カタログ情報ページの URL	
FGDC	/metadata/idinfo/citation/citeinfo/onlink
JMP2.0	/MD_Metadata/identificationInfo/MD_DataIdentification/ pointOfContact/contactInfo/onlineResource/linkage
サムネイル画像の URL	
FGDC	/metadata/idinfo/browse/browsem
JMP2.0	/MD_Metadata/identificationInfo/MD_DataIdentification/ graphicOverview/MD_BrowseGraphic/fileName
空間データの空間範囲	
FGDC	/metadata/idinfo/spdom/bounding/{eastbc,westbc,southbc,northbc}
JMP2.0	/MD_Metadata/identificationInfo/MD_DataIdentification/extent/ geographicElement/EX_GeographicBoundingBox/ {eastBoundLongitude,westBoundLongitude, southBoundLatitude,northBoundLatitude}

JMP2.0では、経緯度座標の場合は、geographicElementのEX_GeographicBoundingBoxタグを利用して空間範囲が記述されているが、公共座標系の場合は、EX_CoordinateBoundingBox内で、westBoundingCoordinate, eastBoundingCoordinate, southBoundingCoordinate, northBoundingCoordinateを利用して端点の座標が記述されている。国土地理院のメタデータエディタでは、空間範囲が公共座標系で記述されている場合でも、経緯度座標の空間範囲に変換して保存することができる。

4.3.4 データベースへの格納

JMP2.0メタデータをアップロードし、XSL変換を適用することで、所定のディレクトリ（例えば、catalogsrv/public_xml/csis/fgdc）以下に、FGDC形式のファイルが生成される。その上で、カタログサーバマシンで、コマンドラインから登録プログラムを実行することによって、catalogsrv/public_xml/csis/fgdc以下のすべてのFGDC形式のメタデータがデータベースに登録される。

5. まとめ

カタログサーバの仕組みとCSISにおける運用について説明を行った。カタログサーバは、空間データに関する最低限の情報を管理し、データ管理者の負担を軽減するだけでなく、キーワード検索機能と地図検索機能によってデータ利用者にとっても直感的に使いやすいものとなっている。今後は、CSISで管理している空間データに関するメタデータの登録作業を進めることはもちろん、他のデータ提供機関と連携して、登録メタデータを充実させることも重要である。また、運用中のCSISカタログサーバは、検索の面でいくつかの問題点があるが、現在、検索機能の改善を進めており、修正版のカタログサーバの公開に向けて準備中である。

6. 謝辞

空間情報科学研究センターのメンバーの方々には様々な形でご支援頂きました。厚く御礼申し上げます。また、株式会社ジャスミンソフトには、システム開発にご協力頂き、運用面でも有用なアドバイスを頂きました。ここに記して感謝します。

参考文献

高橋昭子, 生駒栄司, 伊藤香織, 浅見泰司, 有川正俊, 小口高, 岡部篤行 (2004). 研究機関用空間データ基盤システムの検討. 地理情報システム学会講演論文集, Vol.13.

高橋昭子, 山下亜紀郎, 白石陽, 佐藤英人 (2006). 空間データ基盤システムの検討とカタログシステムの提案. 東京大学空間情報科学研究センター ディスカッションペーパー No.71.

高橋昭子, 山下亜紀郎, 佐藤英人, 白石陽, 大橋智美 (2006). CSIS カタログシステムにおけるメタデータおよびカタログ情報の作成. 東京大学空間情報科学研究センター ディスカッションペーパー No.73.

山下亜紀郎, 大橋智美, 高橋昭子, 白石陽, 佐藤英人 (2006). CSIS カタログシステムにおける数値地図シリーズのメタデータおよびカタログ情報. 東京大学空間情報科学研究センター ディスカッションペーパー No.74.

佐藤英人, 高橋昭子, 山下亜紀郎, 白石陽, 大橋智美 (2006). CSIS カタログシステムにおける国勢調査データセットのメタデータおよびカタログ情報. 東京大学空間情報科学研究センター ディスカッションペーパー No.75.

高橋昭子, 山下亜紀郎, 佐藤英人, 白石陽, 大橋智美 (2006). CSIS カタログシステムにおける ZmapTownII シリーズのメタデータおよびカタログ情報の作成手順. 東京大学空間情報科学研究センター ディスカッションペーパー No.76.

岡部篤行ほか (2005). 大学間連携分散自律型・地理データ基盤システムの開発研究. 平成 15 年度～平成 16 年度科学研究費補助金 (基盤研究 (A) (1), 課題番号: 15200060) 研究成果報告書.

白石陽, 高橋昭子, 山下亜紀郎, 佐藤英人 (2005). CSIS カタログサーバの設計と開発に関する研究. 第 8 回東京大学空間情報科学研究センターシンポジウム (CSIS DAYS 2005), 全国共同利用研究発表大会 研究アブストラクト集

参考サイト

<http://zgate.gsi.go.jp/ch/jmp20/jmp20.html>, 国土交通省国土地理院, クリアリングハウスとメタデータ(JMP2.0 メタデータエディタ, クリアリングハウスノードサーバ構築キット).

<http://www.fgdc.gov/>, Federal Geographic Data Committee (FGDC)