

東京大学空間情報科学研究センターにおける 空間データ基盤システムの概要

貞広 幸雄, 相良 毅, 杉盛 啓明

1. はじめに

東京大学空間情報科学研究センター(以下、CSIS)は、1998年4月、空間情報科学の研究を行うために東京大学に設置された学内共同研究機関である。研究部門は空間情報システム部門、空間情報解析部門、時空間社会経済システム部門の3つから成り、それぞれ出身分野の異なる研究者が空間情報科学という共通の研究課題に取り組んでいる。

CSISは、自身の研究活動はもちろんのこと、他の組織に属する研究者に対する研究支援も積極的に行っている。例を挙げると、

- ・空間データクリアリングハウスの公開
- ・空間データ共有による共同研究の実施
- ・空間情報科学に関する情報交換の場の提供

などがある。このうち、空間データクリアリングハウスと空間データ共有システムが本稿の題名にある空間データ基盤システムを構成しており、以下、その内容について説明する。空間情報科学に関する情報交換の場とは、空間データを扱う際に直面する様々な問題を解決するために、データ変換ツールを提供したり、空間データ利用者がお互いに情報交換するための場である。こちらについては、現在CSISで開発を行っている最中であり、今年度末には一般公開できる予定である。

2. 空間データ基盤システムの概要

空間データ基盤システムは、大きく分けて、空間データクリアリングハウスと空間データ共有システムの2つのサブシステムから成り立っている。それぞれの詳細は後述することにして、ここではシステム開発の背景や目的を述べる。

現在、空間情報科学の研究に携わっている多くの研究者が、最も支援を必要としているのは、それぞれ異なる形式の空間データを同時に組み合わせる多目的に利用する技術である。例えば、様々な地理的現象の解明や、地球環境問題のような空間的社会問題の解決策策定において、国勢調査の町丁・大字別集計データ、国土数値情報土地利用メッシュデータ、LANDSAT-TM衛星画像データなど形式の異なる空間データを同時に組み合わせる研究を行いたいという要請は高い。

しかしながら現状においては、このような異種データを同時に使用するまでに、実に多くの実際的な障害が存在する。即ち、

(課題1) 様々な空間データの形式・所在を調べるための仕組みが整っておらず、研究者の求める空間データの所在が分からない。

(課題2) 必要な空間データが存在しても、データ形式や投影法、空間集計単位などがデータによって異なっており、変換に大きな手間がかかる。

といった問題が挙げられる。これらはいずれも、技術的あるいは研究的に困難な問題ではないが、実際的には研究者にとって大きな障害であり、その結果、研究者の空間情報科学研究へ取り組む意欲を削いでしまっている。空間情報科学発展には、このような問題を取り除くことが重要である。

このような問題意識のもと、CSISでは空間データ基盤システムの開発を行ってきた。前述の通り、このシステムは空間データクリアリングハウスと空間データ共有システムの2つのサブシステムから成り立っている。空間データクリアリングハウスとは、空間データに関するデータ、即ち、空間メタデータを備えた検索システムであり、空間データの形式や所在、内容などの情報をWebを通じて提供することができる。例えば、このシステムに対して研究者が自分の必要とする空間データに関する情報(地域、内容など)を入力すると、それに適合する空間データの一覧が表示される。即ち、空間データクリアリングハウスとは、前

さだひろ ゆきお, さがら たけし, すぎもり ひろあき
東京大学空間情報科学研究センター
〒153-8904 東京都目黒区駒場4-6-1

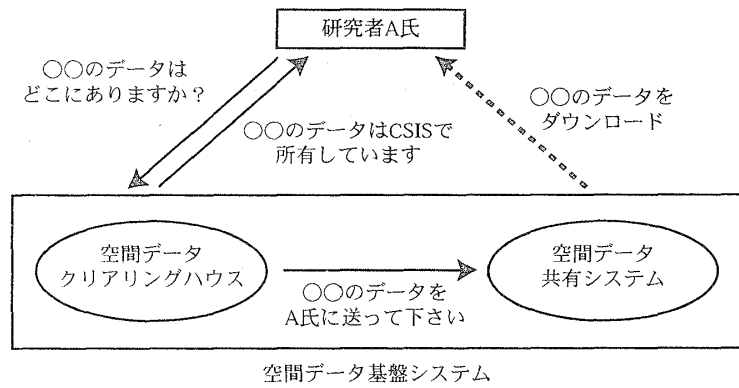


図1 空間データ基盤システムのイメージ

述の課題1を解決するためのシステムである。一方、空間データ共有システムとは、課題2の解決を目指して開発されているシステムであり、ここでは、様々な空間データが統一的な形式で管理されている。研究者はある一定の手続きを踏むことで、これらのデータをダウンロードし、自分の研究に利用することが可能である。二つのシステムのイメージは図1のようなものになる(但し、図1はシステム完成時の様子を示しており、現時点ではクリアリングハウスと空間データ共有システム間の連携は未完成である)。

なお、これらシステムの開発は科学技術振興調整費知的基盤整備促進制度「空間情報科学の確立のための空間情報データベース化に関する研究」の一環として行われている。研究期間の第一期(1997-99年度)には、空間データクリアリングハウスと空間データ共有システムの立ち上げ、及び、試験公開を実施した。現在、第二期(2000-01年度)の半ばであるが、試験公開の結果を踏まえたシステムの改良に取り組んでいる段階である。

3. 空間データクリアリングハウス

空間データクリアリングハウスとは、空間データに関する情報(空間メタデータと呼ばれる)を格納し、それらの検索を実現するシステムである。利用者は、自分の必要とする空間データを、様々な項目をキーとして検索することができる。

目的に適った空間データを検索するシステムとしては、米国において1994年にNSDI(The National Spatial Data Infrastructure: 国土空間データ基盤)の整備を指示した大統領令12906[1]の中に登場したクリアリングハウス(Clearinghouse, 厳密にはNational Geospatial Data Clearinghouse)が有名である。この大統領令の中でクリアリングハウスとは、

地理空間的なデータの製作者、販売者、利用者を電子的に結合する分散ネットワークのことであると定義している。これを受けて、米国ではFGDC(The Federal Geographic Data Committee)がクリアリングハウスシステムを作成、配布を開始し、2000年9月27日現在では世界中で216のサーバが稼働している[2]。日本からも国土院がこのネットワークに参加している[3]。

CSISの空間データクリアリングハウスは、これらの流れを踏まえた上で、特に研究者向けのクリアリングハウスを目指した開発を進めている。商業利用を目的とした場合、地図の主な用途は印刷して利用するための地形図であるため、主要な検索キーは地図が網羅している範囲と縮尺である。一方、研究利用目的ではより広範なデータが必要であり、どちらかといえば地形図上に乗るいわゆる属性情報が重要な事が多い。そのため、それぞれの地図に含まれている属性情報をキーワードで検索する機能を重視している。

研究第一期には、評価用クリアリングハウスはWWWアプリケーションとして実装し、Windows NT 4.0 Server上にWWWサーバとしてOracle Web Application Serverを、メタデータのデータベースとしてOracleを利用した。また、クライアント側はWWWブラウザを利用し、一部GUI部にJavaを使用した。この第1期クリアリングハウスでは、キーワードによる検索機能などのインターフェースや、データベースの管理手法などに不十分な点があった。

そこで第二期では、第一期の成果をもとに全文検索機能やソーラスを利用した連想検索などキーワードによる属性検索機能を強化し、メタデータの管理を容易にしたより実用的なクリアリングハウスシステムの構築と運用を行っており、実装途中ながら試験公開を行っている[4](図2~4)。現在のシステムは、Serv-

letを用いたWWWアプリケーションとして実装しており、Sorialis 2.6上にWWWサーバとしてApache[5]を、Servlet EngineとしてTomcat[6]を利用している。また、基本プログラム部分にはGEO Catalog[7]を利用しており、ほとんどの部分がフリーソフトから構成されている。第2期終了時にはCSISで

拡張した部分を含め、フリーソフトとして配布し、全国の研究機関でクリアリングハウスを構築する際に利用していただいて、分散クリアリングハウスシステムとして運用したいと考えている。

一方、CSISクリアリングハウスはFGDCのクリアリングハウスとプロトコルレベルでの互換性がないため、CSISからFGDCのサーバに検索を行ったり、その逆を行うことはできない。これはCSISクリアリングハウスが研究用であることが主な理由であるが、FGDCクリアリングハウスで利用されている通信プロトコルであるISO 23950がWWW普及以前の技術であり、日本語のサポートをはじめ、現在の技術に組み込むことが難しいという理由もある。しかし、国際的にはFGDCのクリアリングハウスが実質的な標準として利用されているので、将来的にはCSISクリアリングハウスとFGDCクリアリングハウスのプロトコルを交換するゲートウェイを開発するなど、何らかの方法でデータ交換を行える方向に進める予定である。

以上のように、空間データクリアリングハウスのシステム部分については、順次開発・改良を行っている段階である。一方、CSISではこの作業と並行して、クリアリングハウスに投入する空間メタデータの検討も行ってきた。以下、その内容について説明する。

空間メタデータとは、空間データの提供者や調査年月日、範囲、記載されている内容などを説明したデータであり、「空間データのデータ」と呼ばれることもある。空間データクリアリングハウスの検索は、このメタデータをもとに行われる。情報化社会の進展にもなって、「情報洪水」といわれるようなネガティブな側面も取り上げられ、多くのデータを前にした利用者が、どのデータを用いれば良いのかわからない、といった問題が生じている。空間メタデータは、このような問題を解消し、空間データが円滑に利用されることを目的に整備するもので、それぞれのデータが自分の利用目的に適しているかどうかの判断材料やデータの入手先等に関する情報を利用者に提供してくれるものである。

空間メタデータの記述方法については、国際的な標準化の取り組みが、ISOの専門委員会であるTC 211 (TC: Technical Committee) によって1994年以来進められており、それに同調する形で、国内においても、建設省官民連帯共同研究「GISの標準化に関する調査」(1996-98年度)のなかで国内標準に関する検討が進められてきた[8][9]。ISO/TC 211においては、

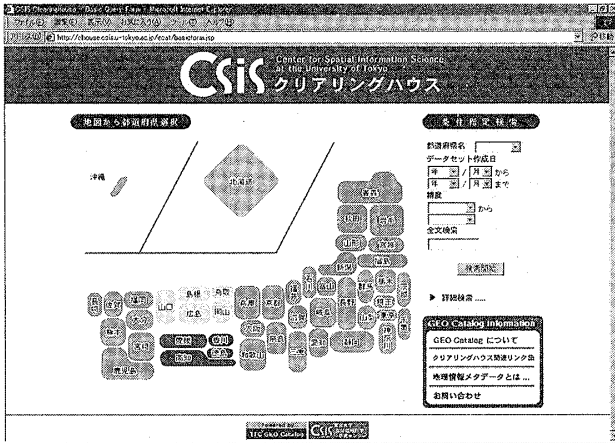


図2 CSISクリアリングハウスの起動画面 (暫定版のため、最終版では変更される可能性がある)



図3 検索条件の指定

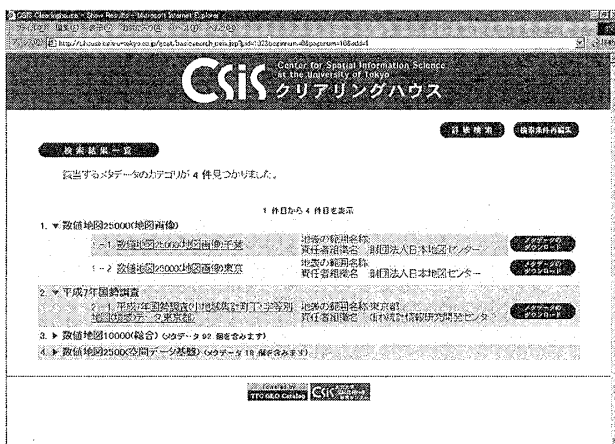


図4 検索結果の例

おもにクリアリングハウスでの検索に利用できるような、データのカatalog情報としてのメタデータ項目を「適合性レベル1」とし、より完全にデータの内容を記述するためのメタデータ項目を「適合性レベル2」として位置づけている。

そこでCSISでは研究第一期(1997-99年度)において、CSISが保有するデータの特性を考慮し、適合性レベル1全てとレベル2の一部のメタデータ項目について記述することでISO/TC 211の定める標準に準拠させるとともに、統計データに関する情報を「キーワード情報」というメタデータ項目に詳細に記述した(表1)。

必須項目のうち、重要なものは題名と分類、データセットの範囲の3つである。題名とは、文字通り空間データの一般的な名称を示す項目であり、「数値地図2500」「平成7年国勢調査第1次基本集計全国編」などと記入される。分類とは、空間データに含まれている内容を示す項目であり、予め定められた22の分類コード(1:地籍, 2:文化と人口統計学, 3:地形, ...)とキーワード(土地利用, 道路ネットワーク, 人口, ...)によって空間データの内容を説明する。データセットの範囲とは、その空間データの指し示す空間的(及び時間的)範囲を示す項目であり、南北の境

界を緯度, 東西の境界を経度で示す方法と, 自治体等の名称によって示す方法の2つがある。

これらメタデータの記述には, XML(eXtensible Markup Language)を用いている[10]。XMLは, おもにインターネット上におけるデータ交換のための標準的な記述言語として, 近年急速に普及しているものであり, HTMLと同様, タグによって構造化されている。HTMLとの大きな違いは, タグ自体を自由に定義できるという点であり, どのようなタグを用いるかについては, DTD(Document Type Definition)と呼ばれるファイルで定義を行う。図5にCSISで定義したDTDの一部を示す。また, メタデータの実例(数値地図2500(空間データ基盤)東京-5)を図6に示す。

表1 CSISにおけるメタデータ項目

	【必須項目】
ヘッダー	メタデータ作成に関する事項
題名	データセットの範囲
分類	データセット言語コード
	【オプション項目】
版・シリーズ名	ファイル識別子(ファイル名)
活動識別情報	参照日
責任者情報	解像度コード
要約・目的	データセットの文字符号集合
進捗状況コード	アクセスの制約, 使用の制約
系譜の説明	定性的な叙述報告
空間表現型コード	空間参照系型
配布データ識別子	適合性レベルのコード
識別情報	配布データのオンライン情報源のURL
フォーマット名	メタデータの言語コード, 文字コードセット
メタデータの日付	配布に使用するメディア
データ品質情報	空間データ表現情報
参照系情報	配布データフォーマット情報
フッター	

```
<?xml version="1.0" encoding="Shift_JIS"?>
<!ELEMENT STA メタデータ (題名, データセットの範囲+, データセット言語コード+, 分類, メタデータファイル識別子?, メタデータの親識別子?, 版?, シリーズ名?, 版の識別情報?, 活動識別情報*, 参照日?, 責任者情報*, 解像度コード*, データセットの文字符号集合?, 要約?, 目的?, 進捗状況コード?, アクセスの制約?, 使用の制約?, 系譜の説明?, 定性的な叙述報告?, 空間表現型コード*, 空間参照系型*, 配布データ識別子*, フォーマット名*, 配布に使用するメディア*, 配布データのオンライン情報源のURL*, 適合レベルのコード?, メタデータコードの言語?, メタデータ文字コードセット?, メタデータの日付?, 識別情報?, データ品質情報?, 空間データ表現情報?, 参照系情報?, 配布データフォーマット情報?)>
<!ELEMENT 題名 (#PCDATA)>
<!ELEMENT データセットの範囲 ((座標による地表の範囲: 地理的識別子による地表の範囲), 時間的範囲の日時?, 最小標高値?, 最大標高値?)>
<!ELEMENT 座標による地表の範囲 (西側境界座標, 東側境界座標, 北側境界座標, 南側境界座標)>
<!ELEMENT 西側境界座標 (#PCDATA)>
<!ELEMENT 東側境界座標 (#PCDATA)>
<!ELEMENT 北側境界座標 (#PCDATA)>
<!ELEMENT 南側境界座標 (#PCDATA)>
<!ELEMENT 地理的識別子による地表の範囲 (地表の範囲名称, 地表の範囲名称参照?)>
<!ELEMENT 地表の範囲名称 (#PCDATA)>
<!ELEMENT 地表の範囲名称参照 (#PCDATA)>
<!ELEMENT 時間的範囲の日時 (#PCDATA)>
(以下省略)
```

図5 DTD ファイルの一部

表2 CSISの所有する空間データ

```

<?xml version="1.0" encoding="Shift_JIS"?>
<!DOCTYPE STA メタデータ SYSTEM "stameta.
dtd">
<STA メタデータ>
  <題名>数値地図2500 (空間データ基盤) 東京-5</
  題名>
  <データセットの範囲>
    <地表の範囲座標>
      <西側境界座標>139.656412</西側境界座標>
      <東側境界座標>139.811255</東側境界座標>
      <北側境界座標>35.742960</北側境界座標>
      <南側境界座標>35.607868</南側境界座標>
    </地表の範囲座標>
    <時間的範囲の日時>1995</時間的範囲の日時>
  </データセットの範囲>
  <データセットの範囲>
    <地表の範囲名称>東京都_千代田区, 中央区, 港区,
    新宿区, 文京区, 渋谷区, 豊島区</地表の範囲名
    称>
    <時間的範囲の日時>1995</時間的範囲の日時>
  </データセットの範囲>

```

図6 メタデータ (XML ファイル) の一部

これらのメタデータ作成に際しては、メタデータ項目の意味の解釈や記述方法について、かなりの議論を行った。その成果を踏まえ、今後のメタデータ作成のための手引きを作成した[11]。

なお、研究第一期では、メタデータ項目 (および DTD) の決定に続き、実際のメタデータの作成と、評価用クリアリングハウスの構築・試験運用を行った。CSISが保有する約600の地図データ・統計データのメタデータをXML形式で整備完了している。これらのメタデータは全てCSISクリアリングハウスに格納されており、空間データの検索に利用されている。

4. 空間データ共有システム

次に、CSISのもう一つのシステムである、空間データ共有システムについて説明する。研究者の利用する空間データは極めて多岐に渡っているが、そのうちの一部は、多くの研究者に共通して利用されている。そのような空間データをCSISでは基盤空間データと呼び、その整備に力を注いでいる (表2)。

これらのデータは、全て学術研究目的で利用することを前提に、さまざまなデータ作成機関から特別な配慮を頂いて利用させて頂いているものである。

データの整備方針としては、国土地理院の数値地図

国勢調査 第1-3次基本集計都道府県編 (市区町村別集計) 基本単位別集計 町丁・字等別集計その1-3 町丁・字等別集計 従業地・通学地集計 従業地・通学地集計その1-3 抽出詳細集計都道府県編 地域メッシュ統計
事業所統計調査 事業所統計調査都道府県編 (市区町村別集計) 事業所統計調査町丁・字別集計 地域メッシュ統計
住宅・土地統計調査 住宅・土地統計調査 全国編, 大都市圏編, 都道府県編
全国物価統計調査 全国物価統計調査 大規模店舗 全国物価統計調査 小売, 卸売
サービス業基本調査 都道府県編 全国消費実態調査 就業構造基本調査 貯蓄動向調査 家計調査年報 工業統計表 市町村編
数値地図2500 数値地図10000 (総合) 数値地図25000 (行政界・海岸線, 地名・公共施設, 地図画像) 数値地図200000 (行政界・海岸線, 地図画像) 数値地図50mメッシュ (標高) 数値地図250mメッシュ (標高) 数値地図1kmメッシュ (標高, 平均標高)
細密数値情報 集約ファイル1 (首都圏, 中部圏, 近畿圏) 10mメッシュ土地利用 (首都圏, 中部圏, 近畿圏)
国土数値情報 昭文社地図 LIFE MAPPLE 1/200000 BITMAP 全国 1/2500 ベクター 東京都23区
ゼンリン住宅地図 ZMAP-TOWNII (1/2500 shape file) アドレスマッチング用データ (都道府県別)
NTTタウンページデータ 東京都, 神奈川県, 埼玉県, 千葉県
住友電工道路ネットワークデータ 全国道路データ

や国土庁の国土数値情報のように、日本全国を網羅し、自然環境から社会経済まで幅広い分野で利用できる汎用的なデータに加え、大都市圏の産業構造や居住生活についての詳細なデータにも着目した。また、国勢調査・事業所統計調査（平成8年度からは、事業所企業統計調査）をはじめとする統計データが充実していることも大きな特徴である。

これらの統計データをGIS上で地図として可視化するためには、各レコードに格納されている、地理的位置を特定できるようなIDを利用する必要がある。すなわち、地図データと統計データが共通のIDによって関連づけられることで、地図上の構成要素（例えば市区町村の範囲を示すポリゴン）が、多くの統計情報（例えば人口、世帯数など）を属性に持つデータとしてGISで解析可能になるのである。

日本国内においては、「行政コード」や「標準地域メッシュコード」などが、地理的なIDとして広く用いられている。例えば、「東京都目黒区」の行政コードは「13110」であり、目黒区が含まれる1/25000地形図「東京西南部」の範囲の標準地域メッシュコードは、「533935」となる。これらのコード体系の詳細については、財団法人日本地図センター（1998）[12]等に説明があるので参照して頂きたい。

CSISでは、これらの空間データについて、共同研究という形でデータを利用したい研究者を集め、学術的に価値があると判断した場合にはデータ提供者に直接交渉し、データ利用の覚書を交わして複製の許可を得ている。共同研究者間では、データをより使いやすい形に変換したり、必要な部分だけを抜き出したものを交換することで、研究を促進することができる。しかし研究者間の物理的な距離が遠かったり、データを配布用メディアにコピーして配送する手間がかかってしまう場合などには、コンピュータネットワークを利用したデータ交換が必要となる。この問題を解決するためにCSISで開発を行っているのが、空間データ共有システムである。このシステムは、利用者認証技術と暗号化技術を利用してデータの漏洩を防ぎつつ、適切な利用者であればインターネット経由で空間データをダウンロードできる仕組みを提供するものである。

空間データ共有システム[13]はWWWアプリケーションとして実装されており、利用者はWWWブラウザを利用してアクセスする（図7）。これ以降の通信は全てSSL暗号化を利用して行われ、通信路上で傍受されても内容を解読することはできない。まず、

CSISで発行したユーザ名とパスワードを入力すると、そのユーザに許可されたデータ（およびデータセット）の一覧が表示される（図8）。単にデータをダウンロードするだけであればFTPのような仕組みも存在するが、ユーザごとに細かく利用できるデータを設定することは難しい。ユーザが一覧から利用したいデータを選択すると、ダウンロード画面に移る（図9）。

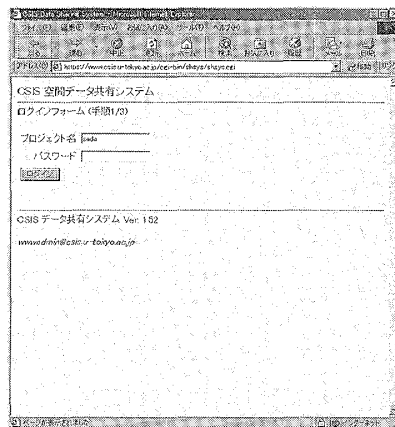


図7 空間データ共有システムの起動画面
(暫定版のため、最終版では変更される可能性がある)

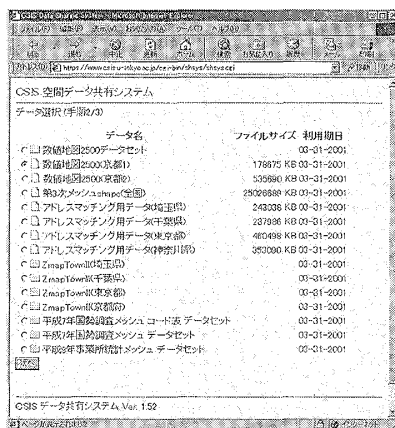


図8 利用可能な空間データの一覧

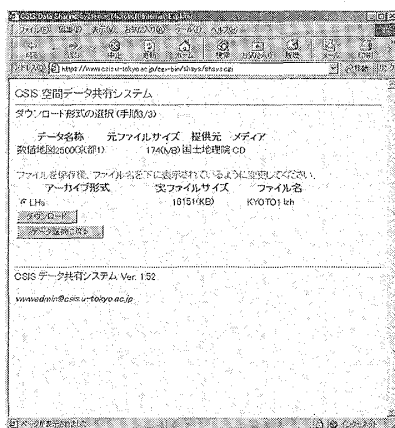


図9 空間データのダウンロード画面

ここでファイルのサイズを確認し、アーカイブ形式を選択してダウンロードを行う。サーバシステムはSolaris 2.6上でWWWサーバとしてApache+SSLパッチを利用しており、プログラムはPerl5で記述したCGIとして動作する。ユーザ管理やデータ管理はPostgreSQL[14]を利用して行っており、アクセスログも管理している。これらの管理を容易に行うために、ODBCに対応したクライアントソフト（MS Excelなど）を利用することもできる。

空間データ共有システムは2000年4月1日より運用を開始し、9月27日現在16のプロジェクトで利用している。3976ファイルの整備を完了し、データ提供者の許可が得られればいつでも利用できる。開発したソフトウェアは2001年3月に無料で配布を開始し、全国の研究機関で同様のデータ配布に利用していただく予定である。

5. おわりに

以上、CSISの空間データ基盤システムについて簡単に説明した。既に述べたとおり、このシステムは現在試運転中であり、どなたでも

<http://chouse.csis.u-tokyo.ac.jp/>

<https://www.csis.u-tokyo.ac.jp/cgi-bin/shsys/shsys.cgi>

からアクセスすることができる（但し、空間データのダウンロードはCSISの共同研究者に限られる）。システムの使い勝手等について、ご意見を頂ければ幸い

である。

参考文献

- [1] 大統領令 12906, US Government, <http://fgdc.er.usgs.gov/publications/documents/geninfo/execord.html>
- [2] Data Clearinghouse Gateway, <http://clearinghouse.esri.com/servlet/IntlServlet>
- [3] 国土地理院クリアリングハウス, <http://zgate.gsi-mc.go.jp/>
- [4] CSISクリアリングハウス, <http://chouse.csis.u-tokyo.ac.jp/>
- [5] Apache, <http://www.apache.org/>
- [6] Tomcat, <http://jakarta.apache.org/tomcat/index.html>
- [7] GEO Catalog, <http://gcat.ttc.co.jp/geo/index.html>
- [8] 空間データ標準化委員会(1999)『地理情報標準』。
- [9] 建設省国土地理院(2000)メタデータの記述・引用方法に関する研究作業報告書, 国土地理院技術資料, E・1_No. 258.
- [10] Extensible Markup Language(XML), WWWC, <http://www.oasis-open.org/cover/xml.html>
- [11] 貞広幸雄・奥貫圭一(1999)空間データクリアリングハウス構築のためのメタデータ記述法, CSIS Discussion Paper, 21.
- [12] 財団法人日本地図センター(1998): 数値地図ユーザーズガイド(第2版補訂版), 財団法人日本地図センター。
- [13] CSISデータ共有システム, <https://www.csis.u-tokyo.ac.jp/cgi-bin/shsys/shsys.cgi>
- [14] PostgreSQL, <http://www.postgresql.org/>