

◆ 特集：環境と GIS ◆

河川水辺の国勢調査データの GIS 化とデータセット

小川鶴蔵*

1. はじめに

国土交通省河川局は、平成2年度から全国規模で実施している河川内の生物調査である「河川水辺の国勢調査」について、平成11年度から電子データ化、GIS化、データベース化のための一連の「河川環境情報システム」構築を進めてきたが、平成15年度末でようやく完成した。これらのデータは、河川局の情報提供 Web サイトである水情報国土データ管理センター

<http://www.mlit.go.jp/river/IDC/database/databasetop.html> からデータを公開中である。筆者は、国土交通省から業務委託を受け、作業のマネージメントを担当する機会を得た。本報告は、その際得た知見と成果の中から、データの品質確保に関する部分を中心に報告する。

2. システム化の目的

システム化の目的は、データを入力段階から電子化し、位置情報を付与し、データベースに格納してデータを公開する一連の過程をシステム化するものであった。このデータ生成過程では、データの品質を確保するためのチェックシステムを組み込み、データの履歴管理を行い、データベースが恒久的に維持、更新できるシステムを構築したものである。この一連のシステム開発に至った動機は以下の通りであった。

2.1 データの公開不備

これまでは、全国の調査データの公開は CD 版として販売されていたが、植生図や確認種の生息環境情報が提供されていないなどで、調査データの公開率は4割に止まっていた。また、調査データは担当した河川事務所に分散、保管されており一元管理がなされていなかったことで、データの品質にばらつきがあり、種の同定を除き組織的な品質確保の対応はなされていなかった。

2.2 報告書（紙）タイプのデータは高コスト

従来はデータの再利用に当たっては個別業務ごとに電子化していた。したがって、データの分析などの利用は限定的で、かつコスト高の構造となっていた。また、広く国民がデータを利用する環境は構築されていなかった。

2.3 データの品質確保の仕組みに不備

生物種が適正に同定されることは調査の前提であるが、河川水辺国勢調査の特徴から、過去の調査との生物種の出現比較は欠かせない。生物学では、研究の進行によって種の分類が常に変化し、同定の基礎となる生物種目録が新しい知見を加味して変化する特徴があり、これまでのシステムでは、過去にさかのぼって、新しい生物種目録に従って同列でデータを比較する仕組みを持っていなかった。

また、データの作成規定とチェックシステムを持たなかったため、さまざまなフォーマットが存在し、また生物種のスペルの誤りなどが多く、データの生成に多大の時間を要していた。

3. 河川環境情報システムの概要

調査会社は、後述する「入出力システム」でデータを入力し、発注元の河川事務所に納品する。納品を受けた河川事務所は所要の検査を行い、事務所内のデータベースに格納する。このデータは、一次チェック済みのデータであることが記録される。次にこのデータが専門機関である財団法人リバーフロント整備センターに送付され、二次チェックを実施するとともに、データ公表に際して所在を明らかにできない特定種の生息場所の隠匿処理をして、データの真正化を完了し、これを河川事務所に送付するとともに、インターネットで公開するための「情報提供システム」を通して、河川局のホームページ「水情報国土データ管理センター」から提供されている。河川事務所では、真正化されたデータをデータベースに取り込み、一次チェック済みデータと入れ替えて、一連のデー

the Outline of the Database of the Results of National Censuses on River Environments with GIS

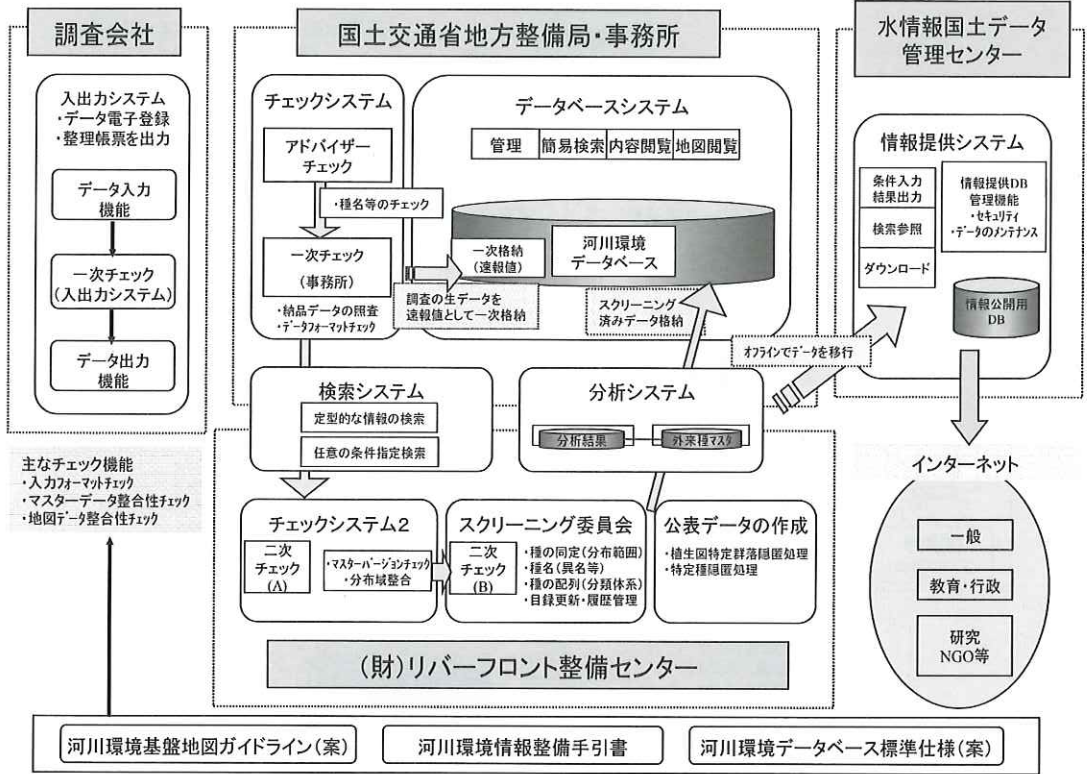


図-1 河川環境情報システムの構成

タ生成作業が完了することになる。(図-1、河川環境情報システム構成)

4. データの真正化 (品質確保) の仕組み

新しいシステムに取り入れるデータは、2種類存在していた。ひとつは、平成12年度から新しくデータを採取するもの、もうひとつは11年度以前の調査データで、これらは紙データとして存在していた(以下、過去データという)。この2種類のデータを、新しく開発した「河川環境情報システム」で運用が可能なデータへ変換した。

4.1 過去データの真正化

データは平成2年度から存在するが、システム化のための基本設計は9年度版調査マニュアルに準拠することとした。このためこのデータ作成規定と大きく異なるマニュアルで作成されたデータは利用価値が低いので、平成5年度以降の調査データを真正化の対象とした。

4.1.1 真正化の手順

報告書(紙データ)から電子データ化するため

には、短時間で大量のデータ処理を必要としたため、2段階の処理方式を採用した。

第一段階は、「河川水辺の国勢調査過去データ入力マニュアル」を作成し、河川水辺の国勢調査マニュアル(数次発刊され、フォーマットが異なる)で作成された報告書から、種名や調査地などの文字と数値を「エクセルデータ」、調査地点などを地図データとして示した「GISデータ」、「写真データ」に整理した。

第二段階は、第一段階で入力されたデータを、平成12年度以降の調査データと同様にシステム運用ができる形式のフォーマットに変換すること、GISデータが適正な位置を示していること、生物種が最新の生物種目録で変換することなどをチェックして、河川環境情報システムでの運用を可能(以下真正化という)にしたものである。

4.1.2 データの照査

第一段階のデータは、入札により異なる多数の作業会社が作成したが、さまざまな品質、かつ大量の第一次段階データが作成された。

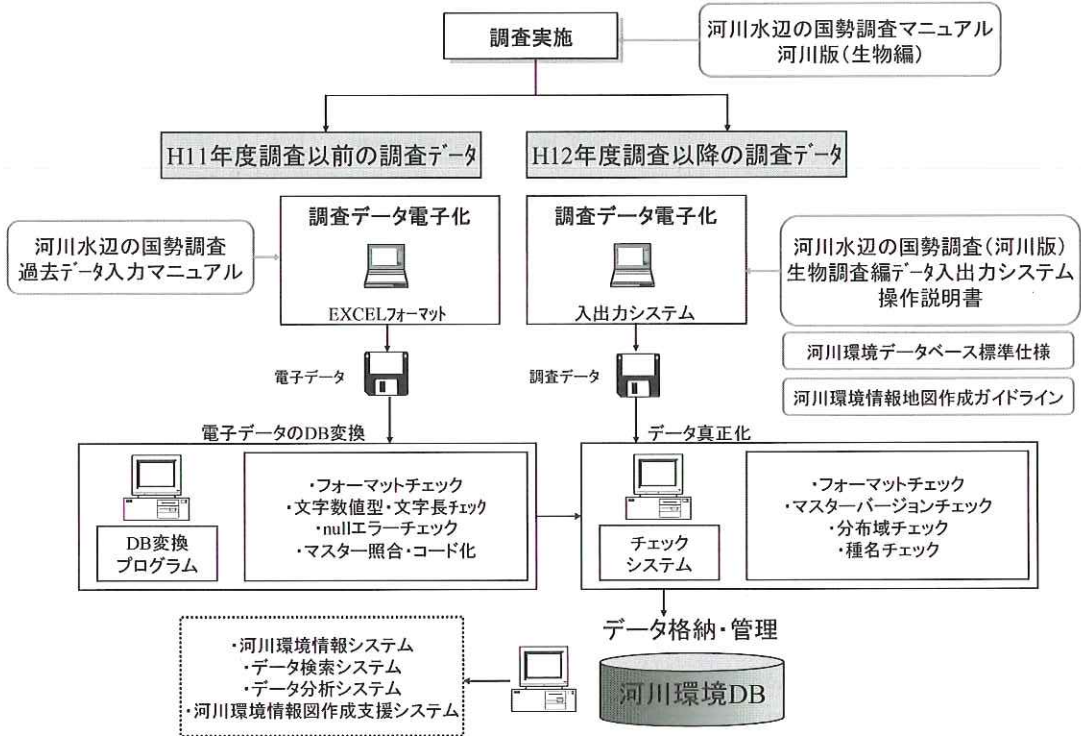


図-2 データ真正化の仕組み

このデータの変換作業を実施するにあたっては、変換が適正に行われているかの検証が必要であった。この当時は、一般的な照査方法が確立されていないが、おりから国土交通省の総プロで研究中の「GISを活用した次世代情報基盤の活用推

進に関する研究」においてGISデータの照査方法の検討が進められていた。

本作業では、これを参考に本作業の照査方法を提案し、実行した。

(1) 照査の流れ

作業は真正化前と真正化後の過去データを相互に比較することにより、適正にデータが変換され、

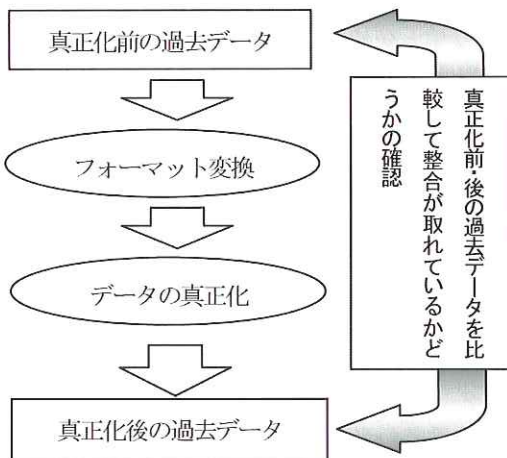


図-3 照査の流れ

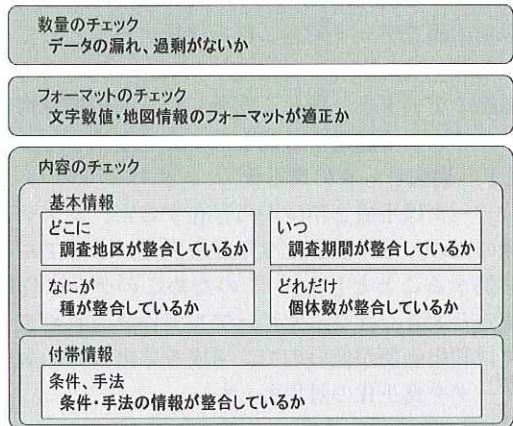


図-4 照査の視点

表-1 照査方法の設定

照査項目	照査方法	照査方法の設定理由
数量チェック	全数検査	調査データの数量は業務の基本事項であるため全数検査を行う。
フォーマットのチェック	全数検査	文字数値及び主題図のデータが遵守すべき関係やルールを満足しているかの基本的な検査であるため全数検査を行う。
内容のチェック	調査地区	地区情報、時間情報、種の分類や名称、個体数、調査環境等や手法は、生物調査の基本情報であることから、その内容について整合を確認するため抜取検査を行う。 <small>① 1調査業務毎に1サンプルを抽出 ② その上で、地方整備局内全数業務の地区数×回数(の総計)に 対する抽出サンプル数の比率が5%を下回らないようにさらにサンプルを抽出</small>
	調査時期	
	種名	
	個体数	
	調査環境等	
	調査手法	

整合がとれているかどうかのチェックを行った。(図-3)

(2) 照査の視点

照査の視点は、必ず必要な数量が漏れなく記入されているか、データフォーマットが適正か、内容が適正に変換移行できているかの3つの視点である。

(3) 具体的な照査方法

1) 照査の方法

調査数量及びデータフォーマットについては、調査データの基本事項であることから、全ての調査データを対象とした全数検査を行うこととした。また、調査地区、調査時期、種名、個体数、調査環境等、調査手法などのデータの内容に関する項目については、生物情報として基本情報であり、データ数が膨大な数であることから、その内容について整合を確認するため抜取検査を行うこととした。(表-1)

2) 照査の実行

データの照査は、前述の通り全てのデータについて照査を行う全数検査と、全データから検査対象を抜き取り抽出する抜取検査の2通りの照査を行った。(図-5)

照査の結果、データ変換プログラムに関するエラーは見られなかったが、手作業を加えた点につ

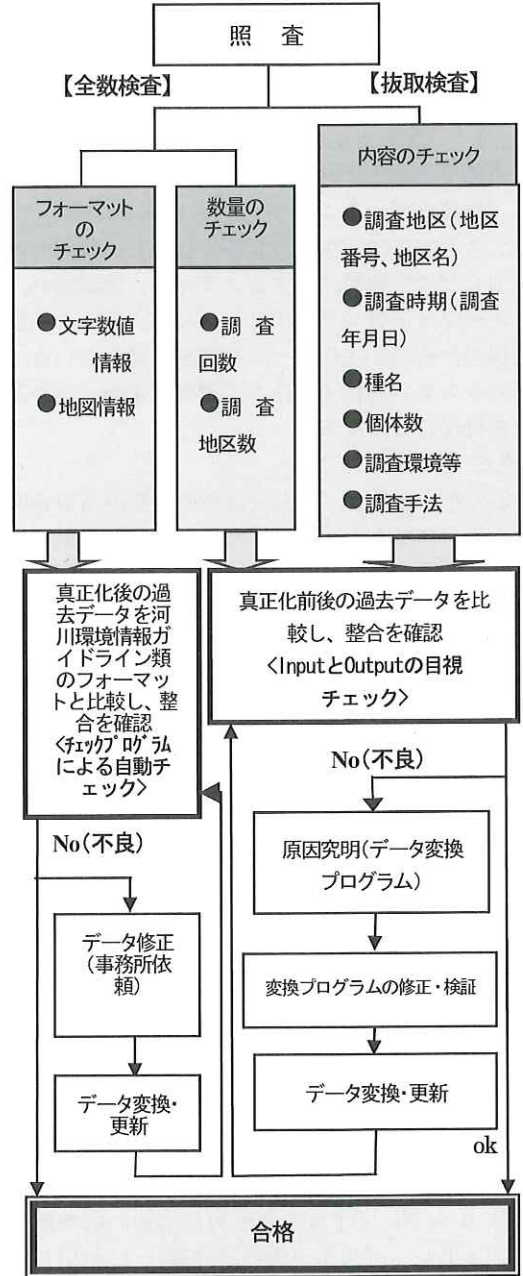


図-5 照査の作業フロー

いては若干の修正ミスを見つけた。このように不良が確認された場合は、その原因を究明した上で対応策を検討し、データの修正後に再チェックを行い真正化前後でデータに不整合がないことを確認した。

なお、抜き取り検査で確認された不良については、その原因が抜き取り検査対象外のデータにつ

いても影響を及ぼしていると考えられる場合は、それらデータについて修正を行った。

4.2 新しく採取するデータの真性化

12年度調査から、新システムに移行した。

4.2.1 一次チェック

現場で調査した生物データは、'入出力システム'を使ってパソコンで入力される。この入出力システムには、生物種目録を搭載し、種名の誤入力などの生物種に関するエラーや、記載漏れ、フォーマットのエラーチェックなど、データ品質確保のための1次チェック機能を搭載している。このシステムを終了すると、調査成果が一定の品質を得たことになる。

4.2.2 2次チェック

2次チェックは、国土交通省の調査担当事務所から送られてきたデータについて、目視チェック、マスター類のバージョンチェック、分布域が間違っていた場所を示していないかについてチェックするとともに、データの同定に誤りの可能性がないかを、スクリーニング委員会を組織して検証し、公表データに向けては、重要種の位置の隠匿処理を行ってデータの真性化を完了する。

5. データの構造

空間データ仕様は、「河川環境情報地図ガイドライン(案)」(平成12年1月、建設省河川局河川環境課)、データの定義、データベース項目規定等は「河川環境データベース標準仕様(案)」(平成12年1月建設省河川局河川環境課)で明らかにしている。これら環境情報データ規定の作成の基本方針は、前者は「河川基盤地図ガイドライン(案)」(平成10年12月 建設省河川局河川計画課)、後者は「河川基幹データ整備標準仕様(案)」(平成11年8月 建設省河川局河川計画課)、ならびに「平成9年度河川水辺の国勢調査マニュアル 河川版(生物調査編)」に準拠したデータ項目、フォーマットとした。

これらの規定を準拠した理由は、河川基盤地図が直接読み込み可能で、かつ基盤地図上に環境データを表現できることとしたこと、過去に調査したデータを活用するため、従来の河川水辺の国勢調査マニュアルに従ったことによるものである。

データ構造は、平成9年度河川水辺の国勢調査マニュアルに準拠したため、魚介類調査におけるデータの階層(図-6)に示すごとく、環境地図

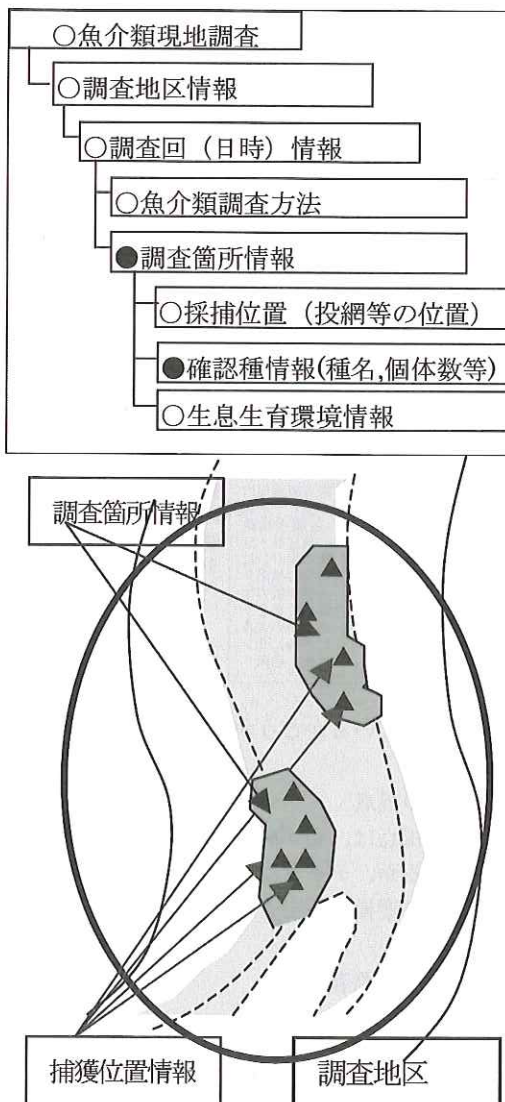


図-6 魚介類のデータ構造

データでは投網採捕位置は取得するが、確認種や採捕数などの属性情報との関連付けは、調査箇所レベルでまとめることにしている。

6. 次世代の環境データ構造

現在の河川水辺の国勢調査は、河川内の生物相(どんな生き物が、おおむねどのあたりにいたか)の把握を目的としている。前節で示したように、発見または捕獲した種が、ピンポイントで記録されず、地区、地先でまとめられているし、属性の種類も限られている。

河川法が改正され、環境が目的化されたいま、計画策定や実管理を行ううえで、どんな生物が、どんな場で生息しているかを、より精密に関連付けて知らなければならない。

ひとつひとつの発見種に座標位置や標高が記録され、また属性については河床材料の質や湿潤度、気温、水温などの場の属性や、同定者名などのデータの品質を明らかにする属性も、今以上に多く必要となろう。

このような場と生物の関係を表現する、新しいデータ構造を定義するためには、現在進行している国際標準規格である地理情報標準 ISO TC211 に従ったデータ構造を新たに提案することが、喫緊の課題と考える。各組織が保有する環境データの相互利用を実現し、流域の健全な生物環境を追求するためにも、急ぐべき課題のひとつなのである。

しかし、これが実現するためには解決すべき課題は多い。第一はデータ量が飛躍的に多くなることである。これには、コンピュータの性能アップとコストの低下が欠かせないが、最近のレベルアップの状況から見て、早晚解決されていくものと楽観している。

第二は、新しいデータベース構造とこれを検索するエンジンが必要なことである。データの品質や内容を示すメタデータが豊富に用意できて、このメタデータを自在に検索するためには、現在一般的に使われているオラクルなどのデータベースエンジンでは検索に時間がかかりすぎることと、データ相互間の関係を正しく関係付けることが技術的に困難となることである。

いま、地理情報標準第2版に基づくデータは、ようやく、試作され始めている。データが作られれば、それを扱うエンジンは、ベンダーによって発売され、実用に供されることは疑わない。

7. 作業を終えて

平成11年度末から開始した一連の作業は、4年を費やしてこのほどようやく実用の段階に達した。

しかし、まだ解決すべき課題はほかにもある。データ提供に関しては、一般向けにはCSVデータで公開したが、公開しているGISデータは、中心点座標しか公開できていない。これは、サーバにまだWebGISエンジンを搭載していないことで、複雑な図形とその属性が表現出来ていない。

また、データベースからのデータインポートや、その他操作に当たっては、より軽快でわかりやすいシステムへの改善が必要なことなどである。

現在この公開システムには、ヘルプデスクを併設しており、ここに利用者からの意見が、多く寄せられてきている。これらの意見は、吟味して、調査マニュアルの改定につなげ、より使いやすいシステムに改善してゆくことにしている。作業担当者としては、これらの意見に常に向き合うことが、作り終えたシステムが継続して運営されるうえで欠かせないものであることを改めて実感している。

参考文献

- 1) 松間 充、大石三之、小川鶴蔵：大規模データ群のデータベース化 —河川水辺の国勢調査の過去データの真正化— リバーフロント研究所報告第14号 2003年10月
- 2) 林 尚、小川鶴蔵：河川環境情報の効率な整備と有効活用について リバーフロント研究所報告第13号 2002年10月
- 3) 林 尚、小川鶴蔵、南城利勝、工藤容子：河川水辺の国勢調査のGIS化に関する検討 リバーフロント研究所報告第12号 2001年10月
- 4) 林 尚、小川鶴蔵、南城勝利、工藤容子：GISに対応した河川環境情報の標準化に関する研究 リバーフロント研究所報告第11号
- 5) 若林伸幸：河川環境情報の電子化、GIS化について 水環境学会誌 vol23 NO9 2000年9月
- 6) 森 吉尚：河川環境GISの開発および河川環境情報図の活用、土木技術 vol57 2002 NO8
- 7) 佐藤寿延：河川環境GISの導入 土木技術 vol57 2002 NO9

小川鶴蔵*



財団法人リバーフロント
整備センター審議役
Tsuruzo OGAWA