

国立公害研究所 正員 ○藤原 正弘
 同上 正員 大島 高志
 同上 松本 幸雄

はじめに

最近、環境公害に関する多くのデータが各自治体等で作成されて来ているが、これらのデータのシステム的な整理、蓄積は遅れており、十分に効果的な利用が図られていない。データの各種の利用ニーズに応ずることを可能とするため、コンピュータシステムを用いたデータの蓄積、検索システムが考えられる。このようなデータバンク又はデータベースの作成の必要性が指摘されて来ている。しかしながら、環境公害のデータはその範囲が広く、かつ、ユーザの要求も多岐にわたる。ユーザの情報要求のカテゴリーとして、たとえば、①データそのものについての質内(数値情報)、②何をみればわかるかという質向(文献情報)、③どこに(誰に)聞けばわかるかという質向(情報源情報)などがあり、内容レベルも概略的な情報を要求するものから非常に専門的情報を要求するものまである。我々は、環境公害関係データについてのデータベースの作成の第一歩として、図-1のような構成のものを作成しているが¹⁾、このうち、大気環境データファイルと水質環境データファイルの作成とその利用について述べる。

データファイルの作成

(1) 考慮事項：大気環境データファイル、水質環境データファイルの作成にあたって、次のような事項の検討を行った。

① データ作成方法——どのようにデータを作成するか(全国統一の測定方法の問題、サンプリングタイムの問題) ② データ収集方法——全国的にデータを収集するにはどうすればよいか。帳票又は磁気テープで収集する場合の帳票フォーマット、磁気テープフォーマットの問題。③ 入力システム設計——WHO, WHAT, WHEN, WHERE, HOW の情報をどのように入力するか。④ ファイルの全体構成——将来、項目を追加する場合のことを考えて、全体のデータフォーマットを考える。⑤ コードの問題——全国の測定局コード、汚染物質・測定項目コード、測定単位コードについて合理的なものを考案し、全国統一を図る。⑥ データのチェック方法——マシンによるチェックと人間によるチェックの併用方法、データクリーニング方法の自動化を考える。⑦ データの蓄積方法——ディスクローディング、MT 保管等によるコンピュータシステムへのデータの蓄積の形態について考える。⑧ 代表的利用パターンの抽出——データがどのように利用されることが多いかを調べ、それに応じた利用プログラム、二次データファイルの作成、インバーティッドファイルの作成について考える。

(2) データファイルの内容：上記の事項を検討した結果、大気環境データファイル、水質環境データファイルとして、大きく① Raw Data, ② Statistical Summaries, ③ Site Description にわけて、かつその中に、表-1に示すような内容の情報を入力するのが適切であると考えた。米国 EPA の SAROAD (Storage and Retrieval of Aerometric Data)²⁾ などのデータベースのシステムも3つのファイルを作成しており、この点では、結果的にはよく似たものとなった。Raw Data ファイルとは別に Statistical Summaries ファイルを作成するのは、使用頻度の高い統計値をスピーディにかつ簡便にとり出すことを可能とするために、全国的な膨大なデータの収集と蓄積の実際を考慮すると必要性があるためである。

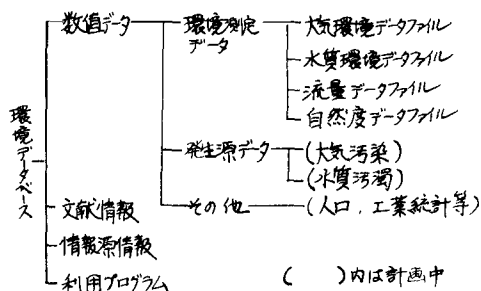


図-1 環境データベースの構成

(3) データの入手：水質環境

データの Raw Data, 大気環境データの Statistical Summaries & v. Site Description については、環境庁が法制度にもとづき、全国の自治体から収集した資料を用いて作成した。その他のものは、独自に調査したり、関係自治体に情報提供依頼をするなどして作成したものである。

表-1 データファイルの内容

		大気環境データ	水質環境データ
Raw Data	対象項目	SO ₂ NO ₂ NO _x CO O ₃ HC CH ₄ 非タンHC T-HC SP SPM 風向 風速	BOD COD DO NH ₄ -N NO ₂ -N NO ₃ -N 無機性N カルダレン T-N PO ₄ ³⁻ 無機性P T-P 水温 流量
	〃 年度	51年度	BOD COD...46~51年度, 水温...51年度
	〃 地域	主要自治体(10地域)	全国
Statistical Summaries	データ量	83.0 メガバイト	23.9 メガバイト
	対象項目	Raw Dataの項目(除く風向、風速)及びDF	Raw Dataと同じ
	〃 年度	45~51年度	Raw Dataと同じ
Site Description	〃 地域	全国	全国
	データ量	37.7 メガバイト	2.4 メガバイト
	名称位置等	測定局コード(国公研コード), 同(自治体コード), 測定局名称(かな字, ローマ字), 測定局所在地・設置場所, 地域メッシュコード等	測定局コード(国公研コード), 測定局コード(小規模メッシュコード), 水質(流域)・灘名称, 支流・湾の名称, 測定局名称, 緯度・経度(度分秒)
	その他	測定局付近の固定発生源の名称と距離, 測定局付近の道路・鉄道の名称, 交通量, 方位と距離等	環境基準類型, 達成期間, 基準指定年, 県庁所在地水域の区別, 測定機関等
	データ量	2.3 メガバイト	1.4 メガバイト

(4) データチェックの自動化：データのチェックはデータファイル作成業務の中で最も重要なものの一つである。これには非常に多くの人手と時間を要するが、この軽減を図るため、コンピュータによる異常値の自動検出法について考えた。大気環境データ(自動測定器により測定され、モニタリングシステムを通じてMTへ入力されたもの)を用いて、①上限、下限値による方法、②加重平均による方法、③項目間回帰分析による方法、④加重平均と上限、下限値による方法などを実験した結果、加重平均と上限、下限値によるチェックを行うと大部分の異常値を検出することができた。もちろん、コンピュータのチェックでは異常値として疑いものを検出するだけであり、これの確認と修正は人手で行わねばならないが、実業務適用への有効性が確認できた。³⁾

データ利用手法の簡便

(1) データの検索・出力：パラメータユーザが端末機を操作して、Raw Data あるいは Statistical Summaries の中から、望むデータを取り出し、望む形式の表にして出力するものである。大気環境データについては、①測定項目、②測定局あるいは都道府県・市町村、③期間(年度・年月・年月日・任意の期間)、④統計処理又は編集方法の4つを条件として指定することにより、一定の様式で編集された結果を検索・出力する。統計処理方法として20種類のものを作成した。水質環境データについても同様の処理ができるようにした。

(2) データのメッシュデータ化：環境測定データは“点のデータ”であるが、“面のデータ”に変換することができれば各種の解析に便利である。このため、各種手法のうち最も実用的と考えられるスプライン法と低次曲面近似法により、大気環境データを用い、データのメッシュデータ化を行うことを考えた。この方法は、ある少数の (X_i, Y_i, Z_i) の組から $Z = F(X, Y)$ となるめらかな曲面をもつ函数 $F(X, Y)$ を作成するものである。 $Z_i = F(X_i, Y_i)$ の前提のもとに、スプライン法は①式、低次曲面近似法は②式 $\Delta^2 Z - \sigma \Delta Z = 0 \dots\dots ①$ $(\Delta = \frac{\partial^2}{\partial X^2} + \frac{\partial^2}{\partial Y^2})$ を解くことにより求める。この結果を他のデータとのオーバーレイ解析などにより評価した。³⁾メッシュ化データは、光化学スモッグなどの広域の空間的、時間的変化の解析に役立つことの見通しが立った。 $\frac{\partial^2 Z}{\partial X^2} = 0, \frac{\partial^2 Z}{\partial Y^2} = 0 \dots\dots ②$

(3) 汚染シミュレーション等：多量ボックスモデル等のシミュレーションモデルのプログラム、濃度と平均化時間の関係の解析のための Larsen モデルのプログラム、その他を整備した。

おわりに 今後、データファイルの種類、内容の充実を図り、かつ、データの多角的な利用、簡便な操作による利用を可能にする各種の手法及びシステムの開発を行ってゆきたい。最後に、データの提供に協力いただいた自治体、システム開発やデータファイル作成に協力いただいた関係民間機関に感謝の意を表する。

<参考文献> 1) 国立公害研究所における環境データベースの概要, 1978. 2) US. EPA, Environmental Information Systems Directory, 1976 3) 国立公害研, 環境公害データ処理システムに関する調査研究報告書, 1978