

## 日本海洋データセンター (JODC) における 新しいデータ処理システムについて

三宅 武治, 中里 秀喜: 海洋情報課

### New data management system of JODC

Takeharu Miyake, Hideki Nakazato : Oceanographic Data and Information Division

#### 1. はじめに

JODC は、1965年以來、海洋観測データの収集・管理を主な業務として、我が国における海洋データセンターの役割を担ってきたが、この間、JODC を取り巻く環境は大幅に変化してきた。

近年、海洋観測データは単純に各海洋観測機関の活動成果としてではなく、一つの国際プロジェクトの総合成果として提出されるケースが増え、海洋化学、生物データ等の新しいデータ項目を管理する重要性が増した。

一方、計算速度の上昇、コンピュータネットワークの発達により、データの迅速な流通・管理が可能となったほか、観測機器の進歩によって高精度、高密度のデータが収集されるようになった。

また、ハードディスク(HD)、CD-ROM等の比較的安価な大容量記録媒体が出現したことにより、大量のデータを半永久的に保存することが可能となった。

このような背景を踏まえて、これまでのJODCにおけるデータ管理業務と、その問題点、そして、その対応策として現在構築中のデータ処理システムについて、以下に紹介する。

#### 2. JODC データ管理業務の現状と問題点

##### 2.1 JODC データ管理業務の概要

受領したデータは、一般的に

- ① 受領データの受付 (管理情報登録)
- ② フォーマット変換
- ③ 品質チェック (QC)

④ トランザクションファイル、マスターファイルへの追加

⑤ DB (J-DARS) への登録

⑥ DB (J-DOSS) への登録

の順で処理される。

①では、受領したデータの観測機関、データ項目等の管理情報を受領情報管理DBへ登録する。

②では、受領したデータファイルをJODCの各種データフォーマットに変換する。

③では、観測値が最小値、最大値の範囲に収まっているか等の簡単なチェックを行う。②と③は同一のプログラムで実行されることが多く、両者をまとめて標準化処理と呼んでいる。

④では、③まで処理されたデータを、各フォーマットごとにトランザクションファイルと呼ばれる一時ファイルに追加する。トランザクションファイルは1年に1度、マスターファイルに追加される。

⑤、⑥は、J-DARS、J-DOSSと呼ばれるデータベース(DB)システムを利用するための作業である。J-DARSは水路部内専用、J-DOSSはインターネットを通じてデータを提供するための部内外共用のDBシステムで、どちらもORACLE 7をベースとしたほぼ同様のDB形式で管理されているため、DBへの変換作業は1度でよいが、DB同士のデータエクスポート、インポート処理によるDBデータのコピーが必要になる。

これらのデータ処理の過程の問題点について、

- ・データ受領に関する問題点
- ・データの標準化に関する問題点
- ・データ提供に関する問題点



の3つのポイントに分けて述べてみたい。

## 2.2 データ受領に関する問題点

受領したデータを確実に保管することは JODC の重要な責務である。最近では、

- ① オリジナルデータの安全管理
- ② その基盤となる受領データ管理のための情報の管理

の2点が、特に問題となっている。

### ① オリジナルデータの安全管理

先に述べた記憶媒体の大容量化に伴い、オリジナルデータの重要性が注目されている。一般に、標準化されたデータは、オリジナルと比較したとき、データの一部が失われていることが多い。JODC の ADCP データを例にとると、オリジナルデータが2分又は5分間隔で取得されているにも関わらず、10分間隔で標準化されていること等である。また、化学系、生物系のデータでは特に、メタデータと呼ばれる観測数値以外のステータス情報の確保が大きな意味を持つ。計算機環境の変化により、高密度、大容量のデータを保管することが可能となった今、現状のシステムのまま運用するのは、大きな損失と言える。

### ② 基盤となる受領管理情報の管理

受領情報は、本来、受領したデータの所在を把握するためだけでなく、内容、提供機関名、各機関別のデータの内訳等についての情報を検索し、統計処理等に使用することも考慮されているが、現在のシステムではこの要求を満たすに至っていない。その理由は

- a データ入力項目の不備
- b 入力文字列の誤字や揺らぎの容認

の2点に絞られる。

a では、必須入力項目の設定がなく、重要な情報項目が未入力であっても登録を許してしまうため、登録機関名で検索したり、統計を取ったりすることができないようなケースが生じている。

b では、データ受領機関名等について「水路部海洋調査課」、「水路部 海洋調査課」、「海洋調査課」のような文字列の揺らぎや文字の欠落、誤字を容認してしまっている。

従って、ほぼ同一の入力内容でありながら、場合によっては、入力された内容にかなりの相違が生じ、実データと関連付けして DB を構築する場合や、データ統計を行う場合など、技術的困難に陥る可能性がある。

また、データの保存性、安全性という点で、システム設計そのものの対応が不十分である。本来 DB 化のメリットの一つはその堅牢性にあるが、現在のシステムでは、データテーブルやファイルが過失によって容易に破壊され得るなどの問題も抱えており、これらの支障が生じた場合は、場合によっては復旧に膨大な手間と時間を要する。

## 2.3 データの標準化に関する問題点

データの標準化に関する問題点は、

- ① データ処理進捗状況管理の問題
- ② フォーマットの問題
- ③ 品質管理の問題

の3点に絞られる。

### ① データ処理進捗管理の問題

JODC では、前述の受領情報管理 DB と併せて、データ処理に関する進捗状況等を把握するためのデータ処理情報管理 DB を運用しているが、両 DB はまったく独立した関係にあるため、受領からデータ処理に至る進捗状況を完全に把握することが不可能である。

データ処理情報管理 DB は、データ処理が終わってしまったデータについての処理方法や終了日時を登録したものであるから、処理途中又は未処理のデータについては、両 DB からの検索結果を照合する以外に進捗状況を知る方法がなく、DB 化のメリットがほとんど生かせていない。

### ② フォーマットの問題

フォーマットに関する問題として挙げられるのは、JODC で管理するフォーマットの種類の多さである。

これは、日本国内はもとより、海外から海洋データを受領する上で、フォーマット設定当初に予定していなかった種類のデータが増加したためである。

各データフォーマットごとに、専用のフォーマット変換、データ検査、マージ、抽出その他のアプリ



ケーションプログラムを必要とし、各データフォーマット間の関連はまったくない。このため各プログラムを個別に作成することによる時間的、金銭的損失だけでなく、プログラムの不備によるデータ処理の滞留、又はプログラム操作の不統一による作業効率の鈍化などの問題がある。

また、プログラム開発言語がまちまちであるため、プログラムにバグが見つかった場合や若干の変更を行う場合の障害となっている。

現在のフォーマットの抱える最大の問題は、新しい観測データ項目を追加することが困難なことである。近年重要となりつつある化学系、生物系データに対応できないため、早急に改善する必要がある。

### ③ 品質管理の問題

データの品質管理に関しては、QC フラグの基準が各フォーマットごとに異なること、フォーマット自体が QC フラグを有さないこと、QC プログラムの不備などが問題となっている。

このような QC の不徹底のため、一部のデータで観測点が陸上に存在する、観測データの最大・最小値を越えた異常データが多く存在する等のデータバグが指摘されていた。

また、バグではないが、1 ユーザに対し、異なるマスターファイルから同項目のデータ（例えば、水温データ、波浪データなど）を提供し、ユーザ側でそれらを複合して使用する場合、QC フラグが統一されていないための混乱が生じる等の問題もあった。

## 2.4 データ提供に関する問題点

現在のところ、データ提供の方法は

- ① マスターファイルからテキストファイルで抽出
- ② J-DARS から SQL \* FORMS で抽出
- ③ J-DOSS からブラウザで抽出

の3とおりが存在する。

①では、専用プログラム又は、その時々データの要求の条件に応じたスクリプト等を作成し、マスターファイルからテキストベースで抽出している。この方法の長所は、マスターファイルを使用するため最新のデータを利用できること、様々な条件に柔

軟に対応できることであるが、一方で、抽出者に専門の技術、知識を必要とすることが短所となる。

②については、操作性が良くないことから、J-DOSS に掲載されていない一部のデータのみについて利用されている。

③については、WEB ブラウザを利用したインタフェースで統一されているため、インターネットを利用できる環境で有れば、操作に習熟していなくても、また、どこからでもデータを得ることが可能である。

②、③の欠点は、あらかじめ決められた抽出条件でのデータ抽出に限定されること、及びマスターデータから DB への最終的な変換に日数を要するため最新のデータが取得しにくいことである。

## 3. 新受領管理・処理システム

以上の問題点を踏まえ、情報処理機器の進歩を積極的に利用すること、海洋データの変化に対応すること、省力化・自動化を推進すること等を目標として検討した結果、

- ① 受領データを安全に保管すること
- ② 受領データの管理情報及びデータ処理の進捗情報を安全確実に管理できること
- ③ 各種処理プログラム群の統一と汎用性を確保すること
- ④ データフォーマットを統合すること
- ⑤ データを自動的に受領すること

の5つの柱に主眼を置き、新システムを開発することとなった。

以下、新システムについて述べる。

### 3.1 受領データの一括保管

従来、受領・標準化処理後のオリジナルデータはあまり重要視されず、MT (磁気テープ) や FD のような受領時の媒体のまま保管するのが普通であった。しかしこの方法では、その媒体の所在管理が曖昧であったり、特に磁気媒体の場合、磁気劣化を起こすなどの問題点があった。

このため、新システムでは原則として、受領したデータは、データ受領管理用 PC の HD 上にデジタルコピーすることにしている。



同 PC は 12G バイトの HD 以外に、スキャナー装置、MO、PD、内蔵 CD-ROM の各ドライブ及び 500 枚 CD-ROM チェンジャーを有している。デジタル化された受領データは半年ごとにまとめて CD-ROM に焼き付け、CD-ROM チェンジャーで保管する。CD-ROM で受領したデータについてはそのまま CD-ROM チェンジャーに保管する。デジタル化になじまない書類情報や図画等はスキャナーで読み込み、画像ファイルとして保管する。

これらの受領データや画像データは、ネットワークを介して自由に参照できるため、マスターデータに準じた使用も可能となっている。

### 3.2 新受領管理システム

本来、データ入手からデータ処理にいたる作業は一連のものであるため、これまでの受領情報管理 DB と処理情報管理 DB を一本化し、データ受領以後の作業の進捗状況を一元的に管理できるよう受領管理システムを開発した。

同システムにおいては、西暦年 4 桁 + 年初からの連続番号 4 桁の計 8 桁からなる受領番号を一意的管理単位として扱い、この受領番号ごとに観測機関、件名、プロジェクト名、提供媒体等を定義する (図 1)。また、複数のファイルを同時に受領することが

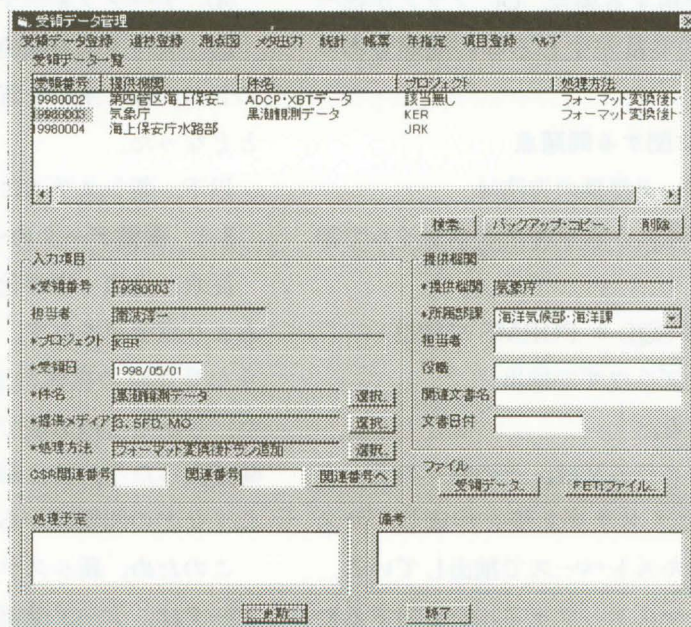
多い点を踏まえ、受領番号に更に 4 桁のファイル番号をもたせ、受領ファイルごとに観測項目、観測船舶、搭載機器、分析機器等の個々のファイル属性を定義している (図 2)。

入力した観測機関や船名、プロジェクト、観測項目等は、自動的にコードに変換され、データ処理情報管理用インデックスファイルとして、他のヘッダ情報とともに出力される。同ファイルは、後述の標準化処理プログラム群と密接に連動しており、フォーマット変換の際にはヘッダ情報として利用され、データの標準化処理の過程で、処理者の前処理作業を減少させるとともにミスを低減させることを狙っている。

また、データ処理プログラムは、フォーマット変換、QC、トランザクション追加等の進捗状況に関する情報、作業年月日、データ範囲等も併せて同インデックスファイルに登録するため、データ処理に関する情報の入力作業は不要となる。同システムの DB は一日一度、自動的にインデックスファイルを確認し、進捗状況を管理情報として DB に登録する。

データ入力インタフェースは、前述の入力時の問題点を解決するため、必須入力項目を設けて最低限の管理情報を確保するとともに、大部分をプルダウン

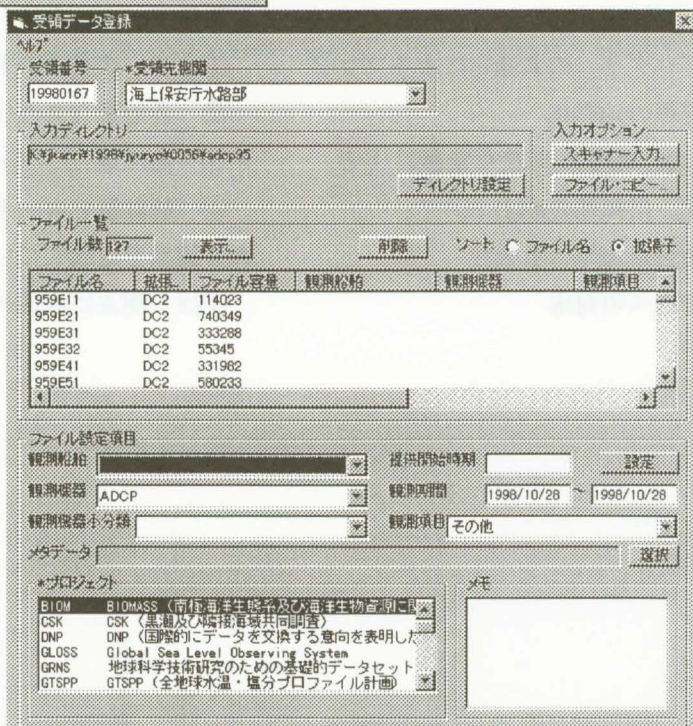
受領管理 メイン画面(受領単位登録)



第 1 図



受領管理 登録画面(ファイル単位登録)



第2図

ンメニューから選択する方法をとった。また、プルダウンメニューの表示内容も必須項目と連動させて、必要なメニューのみを表示・選択させることにより、データの相互関連性を高めた。

具体例を挙げると、ある観測機関を決定した場合、続いて入力すべき観測船舶のプルダウンメニューには、必ず同機関に所属する(又は関連のある)観測船舶のみが表示されるといった具合である。

また、検索結果からは、それらの管理情報のみでなく、受領したオリジナルフォーマットのデータを表示・ファイル化させることができるようにした。この結果、ある検索条件(プロジェクト等)で表示させたオリジナルのデータをまとめて保存するようなことも可能となっている。

更に、検索結果のリスト表示・印刷機能の付与に加えて、事前にプルダウンメニューの内容と1対1対応の英語のリストを登録することで、日英文両方によるリスト表示が可能となり、イントラネット・インターネット等でその検索結果を容易に利用することが可能となった。

また、同システムは、

- ① 更新系システム
  - ② 参照系システム
- の二つのシステムからなっている。

更新系システムは、ODBC を利用した DB への直接アクセスによるシステムで、専用開発したクライアントプログラムにより、DB ユーザ登録された職員のみがデータ入力及び参照を行うことができる。

これに対して参照系システムは、WEB を利用した参照専用のシステムである。本システムは、DB に登録されたデータから自動的に受領データ一覧(日本語及び英語)を HTML 形式で出力するため、DB 登録からのフィードバックが早い上、ユーザを問わず、受領データや各種統計による受領状況が簡単に参照でき、また受領ファイルの表示・ダウンロードも可能となっている。

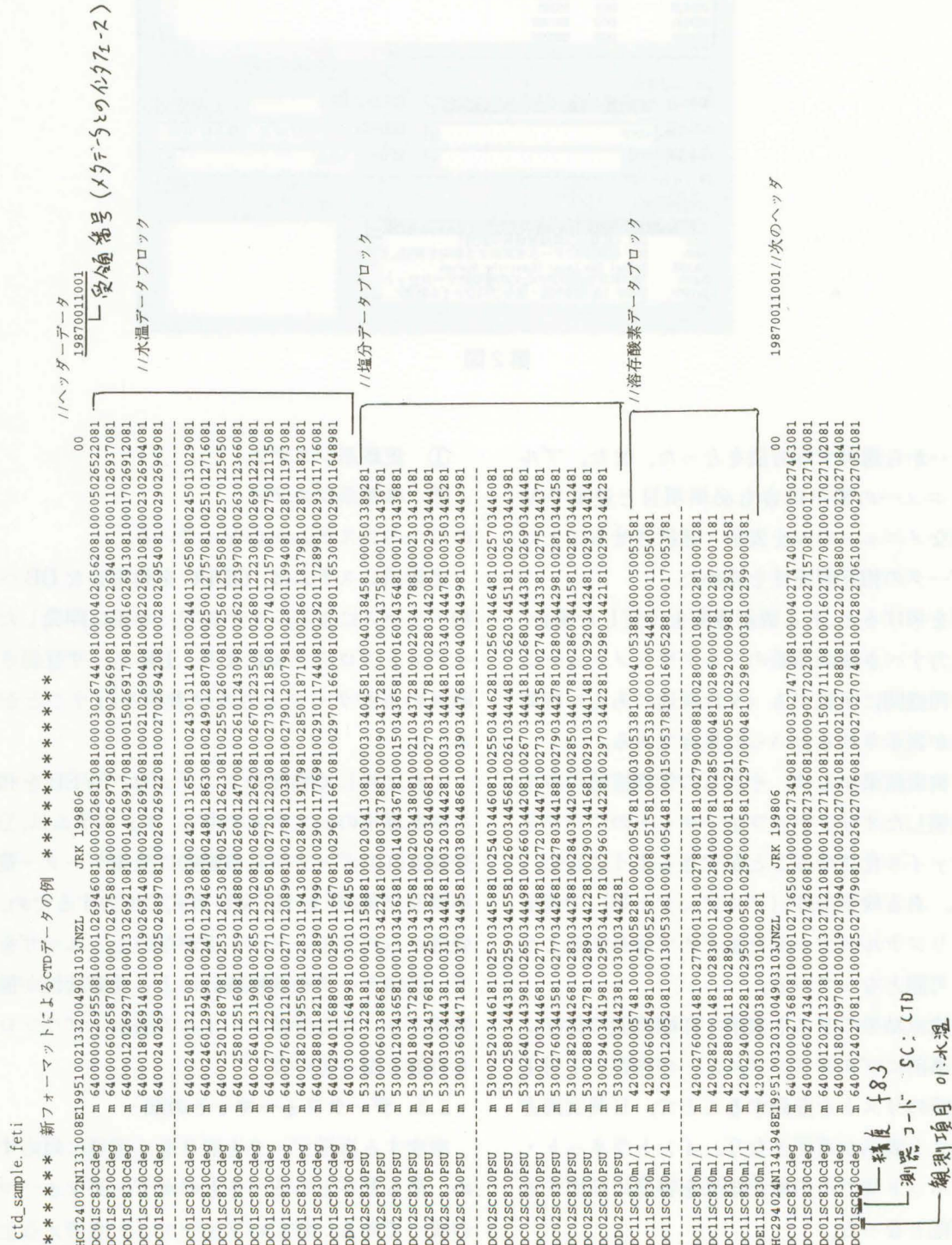
### 3.3 データフォーマットの統一

激変する海洋データを取り巻く環境に対応するため、最も基本となるべきものがデータフォーマットの統一である。新しいフォーマットを考える上で、具体的には、



- ① フォーマットの一元化によるデータ処理の簡素化
- ② データ処理プログラムの共通化によるプログラムの保守管理の効率化
- ③ QC フラグ、QC 処理の明確な定義による、データの精度、信頼性の向上
- ④ 新規データ項目追加等の柔軟性
- ⑤ メタデータ管理への対応

- に主眼を置いた。図3に新フォーマットの例を示す。
- 新フォーマットは、現在のところ、
- ① ヘッダレコード
  - ② 各層型（スカラ型）データレコード
  - ③ 海流型（ベクトル型）データレコード
  - ④ 統計型レコード
- の4種類を定義している。
- ①は1測点に共通のもので、観測時刻、位置等の



第3図

←精度↑83  
←観測項目  
←メタデータ SC: CTD  
←観測時刻



最大公約数的なデータを持つため、測点図の作成や各種抽出に利用する。また、単独で海上気象データとして使用することもできる。

②、③は主として観測データを扱う。

④は②、③から統計処理されたデータのためのフォーマットで、鳥瞰図、等値線図等用のメッシュデータとしても利用される。

簡単にフォーマットを説明すると、①、②は、1つの固定長ヘッダーレコード及び0個以上の可変長データレコードとコメントレコード（メタデータレコード）から構成されている。

これによって、従来、フォーマットの数だけ必要であったQC抽出等のプログラムが1種類ですむため、データ処理作業の省力化のみならずプログラム開発・管理が合理的なものになると期待している。

新フォーマットのデータ面における最大の特徴は、同一のデータ項目をブロック化し、新規データ項目の追加が容易になったことである。従来のフォーマットでは1レコード内に異なる項目のデータを定義していたため、データ項目の追加は困難であった。

また、新フォーマットでは、データレコードにデータ項目、測器コード、データ精度、水深等が記述されるため、将来観測機器が新たに追加されたり、分析精度が向上した場合でも対応が可能となっている。また、各観測値ごとに観測水深値を持つので、標準層以外のデータを持つことが容易になり、XBT, CTD, ADCP等のいわゆる連続データへの対応がなされている。

### 3.4 標準化処理とDBシステム

受領データは図4に示した手順で処理される。このうち、同図の①～④までの標準化を含む一連のデータ処理は、前述したデータ受領管理用PCの上で行われる。

②の作業では、通常受領作業と併せて、旧マスターデータのフォーマットから新フォーマットへの変換にも重きを置いている。各層、XBT, CTD, ADCP, アルゴスブイデータについては変換プログラムが完成している。定常的に受領するデータ以外のものについては、これまでどおり、その都度スク

リプト等を利用して変換作業を行うこととなる。上記プログラムは、同PC上のみで実行されることから、その利点を生かしGUIを使ったインタフェースを採用した。

新フォーマットに変換後のデータは、以後、共通のプログラムによる作業が可能となる。

②のQCプログラムは日時、位置、観測値等についてチェックを行い、エラーレベルに応じたQCフラグをデータに付与する。

フォーマットの変更は、テキストベースのファイルだけでなく、DBについても行う。従来のDBは水温データを扱っているテーブルが、各層、BT, 連続XBT等と多岐にわたっていたため、ユーザが水温データを要求するときに、それぞれのテーブルについて別々に検索しなければならなかった。

新フォーマットでは、各層型データからは水温、塩分等を含むスカラ型テーブルを、ADCP, GEK, アルゴスブイ（換算流）からなる海流型データからはベクトル型テーブルを生成する。

この結果、ユーザは1つのテーブルを検索するだけで同様の結果を得ることができるようになるため、従来のシステムに比べ、検索が容易になり、所要時間も低減する。また、JODCにとっては、マスターからDBへ変換するファイル数が減少することによってDBの管理が容易になるという利点を生み出す。

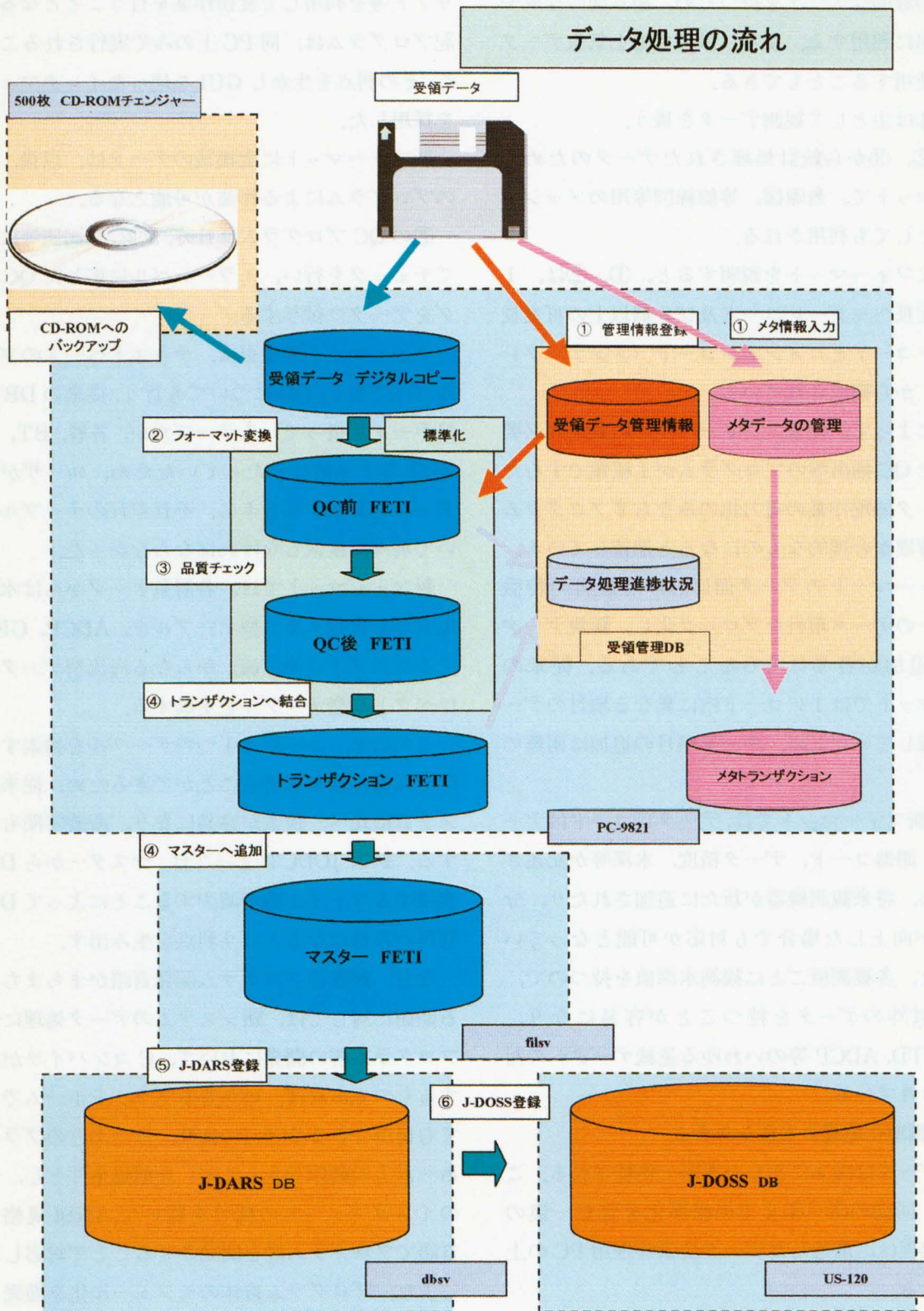
なお、前述のプログラム開発言語がまちまちである問題に対しては、新システムのデータ処理に係るプログラム群の開発において、Cコンパイラが動作するものであれば、いかなるプラットフォームであっても使用できること（つまり、ほとんどのプラットフォーム上で動作できること）を前提条件とし、一部のインタフェースの部分を除いてANSI規格のC言語でプログラム群を開発させることで対応した。

また、プログラム自体のモジュール化を徹底し、特殊な処理ルーチンについては完全に外部ルーチン化することによって、プログラムソースとしての価値を付加し、JODC職員によるプログラム修正・作成の助けとなるものとした。

### 3.5 使用する各種コード、単位等について

統一フォーマットの作成にあたり、各種のコード、





第4図



単位等の再規定も重視した。本来、フォーマット変換の前提となるべきコード定義が曖昧なことによって、データ処理において思わぬトラブルに見舞われたケースが少なからずあったからである。

別々のフォーマットのデータをマージする場合を例にとると、999と空白の両方を欠測値として認めていた場合のデータ検査時のエラーや統計値異常などである。

この例に関して言えば、新フォーマットでは、欠測値はXXXと定義したほか、原則として数字の桁は0で埋めることに規定しており、数字の左側（位の大きい方）は0で埋め、右の桁はデータ精度を考慮した空白に変えることとした。

その他、データ管理の上で特に重要な項目である船名については、従来使用してきた2桁の船舶コードに代え、一意かつ汎用性のある船舶コールサインを採用することとした。その詳細なリストはJODCのイントラに掲載しているので参照されたい。

全ての具体例を挙げるときりがないのでここでは言及を避けるが、原則として、ISOコード、WMOコード等の国内外の両方で主流と認められるコード、単位がある場合はそれを尊重した。

また、時刻、水深単位など、明らかに国内外で使用しているコード、単位が異なる場合は、国内研究者の使用を優先し、国内における一般的コード、単位を採用した。

### 3.6 自動受領システム

自動受領システムは、データ管理の省力化、合理化を考慮して開発したものである（図5）。

同システムには

- ① WWWを使った自動受領システム
- ② WSクローンによる自動受領システム

の2種類がある。

#### ① WWWを使った自動受領システム

同システムは、JODCのWWWサーバ上で管理するWebページの所定のフォームに必要事項（送付するファイルのパス、送付者名、データ種類、備考等）を入力することによって、データがWWWサーバ上の所定のディレクトリに送付されるとともに、JODCのデータ管理者に対しメール通知される

ものである。

これは、JODC職員の作業の低減というよりはむしろ、送付者のFTP等に要する作業低減に資するものであるが、入力項目を充実させることで、必要な管理情報・メタデータを確実に得ることができる。

#### ② WSクローンによる自動受領システム

同システムは、試験的に運用しているもので、LANで接続された管区のWS上にある所定のディレクトリに蓄積されたデータファイルの中から、JODCのWSで設定したクローン（自動実行プロセス）が新しいデータファイルのみを検出し、自動的に回収するものである。現在のところ、まだデータの受領のみにとどめているが、将来的にはフォーマット変換までの処理を自動で行うことを目指している。

### 4. おわりに

以上、データの管理及びフォーマットの変更等を中心に、新しいシステムについて大まかに述べた。

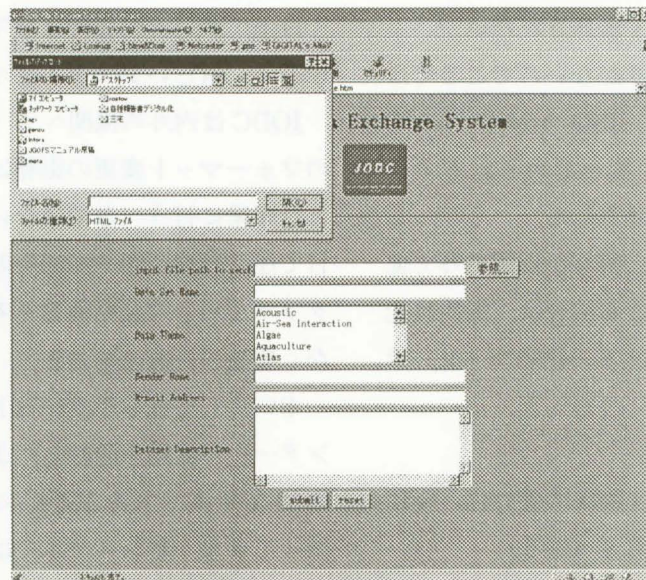
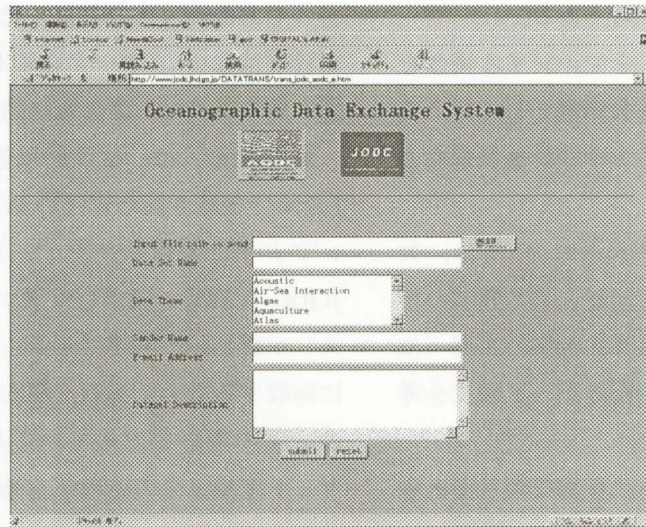
JODCは内外の機関へデータを提供しており、そのフォーマット変更の影響は大きいため、軽々しく、また頻繁に行うべきではない。しかしながら、2項目で述べた諸問題や2000年問題、水路部全体のデータファイルとの互換性等を考慮する必要から、やむなく実施したものである。

幸いにも平成9年度からMIRC（海洋情報研究センター）が活動を開始し、より高度の品質管理や技術的なサポートをJODCに対して行える環境が整ってきた。新システムではこれらを含めた最新の技術を取り込みながら、より使いやすいものを目標に改良を加えていきたい。

なお、ここに述べた受領システム及び新フォーマットの詳細については、JODCのイントラページ（<http://192.168.0.45/jodc>）以下に掲載しているので、そちらも参照されたい。



### 自動受領システム



第 5 図