

## 漁業センサデータベースの更新について

楫取和明<sup>†</sup>, 瓜倉 茂, 青木邦匡

### Updating the Fishery Census Database

Kazuaki Kajitori<sup>†</sup>, Shigeru Urikura and Kunimasa Aoki

**Abstract** : It is not an easy task to update a census database by adding the data of the most recent survey since many changes have been made to the new survey according to the various changes of the social circumstances. For the electric data of the Fishery Census, the fishery areas and the survey items have been changed over the surveys. We therefore had to maintain the method to compare different surveys of the Census by keeping track of those changes of the fishery areas and survey items. We had designed a comparison scheme for our database of the 6-10th Census and by using it we added the 11th Census data to our database. In this paper, we describe the update of our database and argue new points which arose because of the new features of the 11th data and the developments of the computing environments.

**ASFA keywords** : Data, Data collections, Census, Surveys, Data processing

#### はじめに

我々は漁業センサをはじめとする各種社会調査データをデータベースに格納し検索分析手段を提供するシステムを構築してきた<sup>2)</sup>。社会調査は、毎月、毎年、5年毎と定期的に行われるものが多いが、同じ調査でも各次の調査間には様々な違いが存在する。

理由としては、調査項目が時代に合うものに見直される、市区町村合併等のため集計対象地区が変更されるなど、主に時代に即した変更がなされる事情があるからである。

そのため、各次の調査データの比較分析のためにはどの部分が同じなのかをはっきりしておく必要がある。6~10次漁業センサ(1978~1998)の電子データについてはこれらの同定をして比較分析ができるスキームを提案し<sup>3)</sup>、システムに実装した<sup>1)</sup>(以下、我々のシステムという)。

このたび11次漁業センサ(2003)の電子データ<sup>1)</sup>が入手できたので、早速上記スキームに従って我々のシステムに追加してみた。今回の11次センサデータとして入手

したものは、6~10次のものとはかなり違っている。また、コンピュータ環境などとりまく社会状況も変わってきている。このため、新たに多年度データベースの更新の問題点が浮かび上がってきたのでそれらについて考察する。

#### 11次漁業センサ電子データについて

11次漁業センサの電子データは、Microsoft Windowsのアプリケーションになっていてトップ画面で「目次検索」「表頭・表側検索」「地域範囲検索」の3種類の検索方法が選べる。これらの中から選択しさらに見たい項目や地域を選んで検索結果をExcelで表示するようになっている。「目次検索」ではセンサ報告書各巻と同じ統計表を見ることができる。「表頭・表側検索」では「結果表」「一覧表」の2種類の表が見ることができ、「地域範囲検索」では同じことを日本地図の中から地域を選んで見ることができる。電子データとして我々のシステムに取り入れることができるのは、「一覧表」と「結果表」のデータであるので、以下これについて述べる。

2007年11月30日受付. Received November 30, 2007.

水産大学校水産情報経営学科 (Department of Fisheries Information and Management, National Fisheries University)

<sup>†</sup> 別刷り請求先 (Corresponding author) : kajitori@fish-u.ac.jp

漁業センサスの基礎データは全国の各漁業経営体毎の調査データである。これを社会統計一般の慣習に従って個票データ (microdata) と呼ぼう。個票データを各調査項目について漁業地区毎に集計 (合計や平均を取る) したものが「一覧表」(6~10次電子データでは「漁業地区整理表」と呼ばれていたもの) である。

以下は「一覧表」の例として、漁業経営体の基本構成 (総括) なる表から、北海道のいくつかの漁業地区レコードを拾ったものである。

Table 1.

地域名	市区町村	漁業地区	経営体数	無動力船隻	船外機付船隻	動力船隻	平均漁獲金額
北海道	別海町	野付	294	4	334	81	1232
北海道	別海町	別海	105	2	220	6	715
北海道	根室市	湾中	126	7	208	53	1125
北海道	根室市	根室	106	7	153	83	8506
北海道	根室市	齒舞	467	35	874	116	1579
北海道	根室市	花咲	61	2	43	44	1658
北海道	根室市	落石	207	-	180	109	2235

An example of an area records table.

「結果表」とは個票データから作られたクロス集計表である。「結果表」におけるクロスを取り方は、表頭・表側の組み合わせも非常に多く、集計単位も全国から各漁業地区まで幅広いもので、様々なクロス集計表が供給されている。

以下は「結果表」の例として、経営組織別 (表側) 一経営体基本構成の一部 (表頭) のクロス集計表を北海道別海町野付漁業地区についてとったものである。

Table 2.

経営組織	経営体数	無動力船隻	船外機付船隻	動力船隻	平均漁獲金額
計	294	4	334	81	1232
個人	212	4	292	1	336
会社	-	-	-	-	-
漁業協同組合	-	-	-	-	-
漁業生産組合	-	-	-	-	-
共同経営	82	-	42	80	3551
官公庁・学校・試験場	-	-	-	-	-

An example of a cross tabulation table.

6~10次センサスの電子データも閲覧は同様のアプリケーションを通して行うようになってきている。しかし、6~10次の電子データは11次の「一覧表」にあたる「漁業地区整理表」であって、11次の「結果表」にあたるものがない。従って我々はとりあえず11次の「一覧表」データを我々のシステムに追加した。その詳細を次章に述べる。

11次の「結果表」は「一覧表」では消えてしまう個票データの情報をいろいろな形で見せてくれる。「結果表」の活用については最終章で触れる。

## システムへの「一覧表」の追加

我々のシステムはリレーショナルデータベースを基礎としている。各次のデータは、調査項目をフィールドとし各漁業地区のデータをレコードとするリレーションを成している。調査項目は階層構造を持っており、検索時に上位項目を指定すればその下位項目すべてを指定することになるようにSQLに拡張がなされている。また、多年次のデータを比較検索できるようにも拡張がなされている。

上述の多年次の比較分析スキームでは、新しい (このたびは11次の) データを新しいリレーションとして追加した上で、各次の漁業地区と調査項目の対応をマスター表を作るとることにしている。以下の多年次スキームの要点は拙稿<sup>3)</sup>に従っているが、コード体系をリレーショナル的により厳密なものにしている。

- ・年次によって調査 (集計) 対象が変わる場合、全年次にわたるマスター対象表を作る。

漁業センサスの「一覧表」データの場合、集計対象の漁業地区についてマスター地区表のほか、マスター市区町村表も作る。

- ・複数年次表の比較を行う場合は各年次表とマスター対象表の結合によって行う。

- ・年次によって項目が異なる場合、または大項目がある場合は、全年次にわたる全ての項目 (大項目, 小項目) を挙げたマスター項目表を作る。各項目レコードはマスター項目コードと各年次表項目コードを持つ。

- ・新旧調査項目, 新旧集計対象地区など, 新旧の同定は年次間共通のユニークなコードを割り当てれば分かりやすいし, 処理的にもコードをキーにして結合を取れば比較は簡単である。コード系はそのようにしてある。

架空のセンサスデータを例にとり具体的な処理方法を示そう。簡単にするため年次は10次と11次のみとする。まず10次, 11次の単年次データ表が以下のようなものとする。cc10, cc11は漁業地区コードである。

Table 3.

cc10	x1	x2	x3	x4
1	250	11	20	51
2	300	14	101	92
3	450	10	30	82

10th Census Table : gl0

Table 4 .

cc11	x1	x2	x3	x4	x5	x6
1.11	400	0	3	12	11	25
2.11	100	5	3	10	9	20
2	270	22	9	44	51	61
3	500	5	0	20	38	40

11th Census Table : g11

マスター項目表miはつぎのようなものとなる。

Table 5 .

b	bname	a	aname	code10	code11
NULL	NULL	cc	地区コード	cc10	cc11
b0	地域名	a0	市区町村名	sn10	sn11
b0	地域名	a1	漁業地区名	cn10	cn11
NULL	NULL	a2	平均漁獲金額	x1	x1
NULL	NULL	a1.11	活魚販売経営体数	NULL	x2
b1	男子就業者数	a3	15~30 歳	x2	x3
b1	男子就業者数	a4	30~45 歳	x3	x4
b1	男子就業者数	a5	45 歳以上	x4	NULL
b1	男子就業者数	a5.1	45~60 歳	NULL	x5
b1	男子就業者数	a5.2	60 歳以上	NULL	x6

mi

10次就業者数の項目「45歳以上」は11次では2つの項目「45~60歳」「60歳以上」に分離している。

マスター対象表は漁業センサの場合、漁業地区表chikuと各次の漁業地区の対照表mcからなる。chikuの列は左からcc（漁業地区コード）、cn（漁業地区名）である。mcの列は左から、cc（漁業地区コード）、cc10（10次漁業地区コード）、sc10（10次市区町村コード）、cc11（11次漁業地区コード）、sc11（11次市区町村コード）である。（都道府県コード、大海区コードは簡単にするため省略している。）

豊津町豊津は豊津町村津、川玉に分離している。

Table 6 .

cc	cn
1.11	村津
2.11	川玉
1	豊津
2	安浦
3	西関

chiku

Table 7 .

cc	cc10	sc10	cc11	sc11
1.11	1	20	1.11	20
2.11	1	20	2.11	20
1	1	20	NULL	NULL
2	2	30	2	10.11
3	3	40	3	10.11

mc

これに市区町村のマスター表が加わる。豊瀬町は西関市に吸収合併されている。

Table 8 .

sc	sn
20	豊津町
30	豊瀬町
40	西関市
10.11	西関市

Table 9 .

sc10	sc11
20	20
30	10.11
40	10.11
NULL	10.11

shikuchoson

ms

複数年次表の比較はつぎのように行われる。

ユーザはマスター項目表を見て例えばつぎのように問合せを指定する。

集計地区：10次

表示：10.b0, 10.a5, 11.a5\_1, 11.a5\_2

条件：10.a5 < 11.a5\_1 + 11.a5\_2

これは、11次の45歳以上就業者数が10次より多い10次地区の、地域名（10.a0, 10.a1）と就業者内訳（a5, a5\_1, a5\_2）を表示するための指定である。ユーザ指定用の汎用コード10.a0, 10.a1, a5などは、翻訳ルーチンによりマスター項目表にしたがって、実コードsn10, cn10, g10.x4などに翻訳される。

集計地区が10次と指定されているので、システムはmcとshikuchosonとの左結合をsc10=scで、chikuとの左結合をcc10=ccの条件で取り、g10, g11を左自然結合した後、cc10でgroup byする。

関連する列だけ抜き出す形でSQLを示せば、

```
SELECT sn10, cn10, g10.x4, sum(g11.x5), sum(g11.x6)
FROM mc LEFT JOIN shikuchoson ON sc10=sc
      LEFT JOIN chiku ON cc10=cc
LEFT JOIN g10 USING (cc10)
LEFT JOIN g11 USING (cc11)
GROUP BY cc10
```

こうすると、地区名は10次の地区名で表示され、10次で分離された地区に関しては11次のデータは加算（sum）され分離前の地区で比較がなされる。

就業者数なのでsumをしたが、平均漁獲金額などの平均値であれば経営体数の重みをつけた重みつき平均を取るようにシステムは判断しなければならない。

Table10.

sn10	cn10	g10.x4	g11.x5	g11.x6
豊津町	豊津	51	20	45
豊瀬町	安浦	92	51	61
西関市	西関	82	38	40

The joined table for comparison of 10th and 11th.

この表からの選択結果として次を返す。

Table11.

sn10	cn10	g10.x4	g11.x5	g11.x6
豊津町	豊津	51	20	45
豊瀬町	安浦	92	51	61

The result of the selection.

以下, 11次の「一覧表」を上での多年次スキームに従ってシステムに追加した作業を手順に従って述べる。

1. 11次の「一覧表」データは数多くのMS ACCESSの表に格納されている。これをダンプしたものをシステムのデータベース上でリストアし結合して, 11次の単年次テーブルができた。元のACCESSのデータは, 全国を9地域に分割した9個のデータベースに分かれ, さらにその各々が調査項目をいくつにも分けた多くの表に分割されている。したがって文字通りの「一覧表」ではない。本来1つの一覧表であるものを多くの表に分けるメリットは我々のシステムにとっては存在しないのでこれら多数の表のJOINやUNIONをとって本来の一覧表を作らねばならなかった。
2. 調査項目は11次で見直しがされ, 10次とは変化している。廃止された項目, 新規に追加された項目の他, 従来の項目が細分化されたケース, 従来の関連項目が統合されたケースなどがある。従来と変化ないと思われる項目が大多数ではあるが, しかし同じ項目名であるからといって同じ調査項目である保証はない。その同定は調査主体が行う以外にないのであるが, 新旧項目の対応は今回与えられていない。とりあえず, 11次の調査項目で10次以前に同じ名前のものがあるものについて, マスター項目表に11次データとして追加した。
3. 11次の集計対象漁業地区も10次から変わっている。まず11次の調査年2003年は市町村合併が盛んになり始めた時期で, 市町村名に変化が見られる。その他全体で2000地区余りあるうち, 100以上の漁業地区

名が削除され100以上の漁業地区名が新しく追加された。しかし, ここでも新旧漁業地区の対応が与えられていない。これは項目の対応より深刻である。同名の地域が同じ地域であるか, 名前の異なる地域が違う地域であるか, 合併がなされているか, これらも調査主体による関係の明示以外に判断材料はない。とりあえず, 11次の地区で明らかに10次以前と同地区と思われる地区をマスター対象表に追加した。

4. システムはウェブアプリケーションとして構築されているが, ウェブスクリプトの部分の変更は, 11次のデータテーブルを追加したことに対応するわずかな変更であった。

## 考 察

### 多年次スキームについて

我々のシステムの多年次スキームでは, 更新はシステムへの新規データの追加では済み, プログラムの変更はわずかで済むことが実際に確認できた。上記のスキームのうち特定次の地区について集計比較する機能はSQLモニタ上で動かしただけであるが, その他の機能は我々のシステム上で更新後も問題なく俊敏に動いている。

年次比較の検索指定の方法については独自の拡張SQLに基づいているのでその文法が表に出る形になっている。関係データベースとしてシステムが発給してその本質を表に出すことでシステムを理解しやすくなるメリットを取っているのであるが, 年次比較で上記例題のような10次の地区に基づく指定をするとすると拡張SQLに対する慣れが必要であろう。この辺は拡張SQLを表に出さずにより簡単に使えるように自動化する方向もあるので, 機能的にさらに複雑化したら考慮しようと考えている。

一方, 調査項目と地区の新旧対応については厳密な対応をシステムに取り込むことはできなかった。従って, 累年統計としては不完全なものになっている。

これは, 前節で述べたように供給データにおける(確実な)情報が不足していることによる。6~10次データについては各次の調査項目の対応表がありそれをうまくシステムに取り入れることができた。漁業地区についても6~10次で同じコードが用いられていた所以他们のコードについて結合をとればよかった。11次ではデータ供給側の多年次比較のためのデータ整理が十分なされていなかった。少なくともデータ受入れ側で整理・加工ができるような情報

提供が望まれる。多年次スキームはデータ受入れ側のシステムだけでは完結しない。以下に多年次スキームにとってどのような情報が必須かを述べる。

一つの旧項目（対象）が複数の新しい項目（対象）に細分化されているケース。これは旧と比べれば推測はつくが明示されるべきである。

複数の旧項目（対象）が新しい一つの項目（対象）に統合されているケース。これは旧と比べれば推測はつくが明示されるべきである。このケースの具体例として11次漁業センサ時に市区町村のレベルで合併が多く起こっている。11次の調査時（平成15年）はいわゆる平成の大合併の途中であり、その後成された多くの合併は調査データには反映されていない。よって非常に過渡的な市町村データであり、やはりはっきりした新旧の対応づけが必要である。

新旧の対応は明確に述べられねばならない。例えば、三宅島のように今回災害のため調査ができなかったケースはそのことが明示されるべきである。

地域名データでは、巖原か巖原か、鰯ヶ沢か鰯ヶ沢か、諫早か諫早か、といった字の違いによる混乱が調査次間で見られた。正しい表記がある場合はもちろん、統一された表記がなされるべきである。

分離・合併など比較可能な新旧の対応は明記されるとともに、新旧比較研究に大いに利用されるべきである。一方、定義の変更などのように項目（対象）名は同じでも実質異なる場合も誤解のないような明記が必要である。

これらの情報の記述は、「利用上の注意」などとして別の文書（ファイル）になされることが多いが、何がどこに書かれているかが分かりやすいこと、データユーザの用途が何であれできるだけ自動処理が可能であることが望まれることから、XMLなどの汎用の文書記述フォーマットを採用することが考えられよう。社会統計データに関するXMLベースの規格に関して後に述べる。

### 「結果表」について

従来、漁業センサ電子データは漁業地区単位で集計されたものしかなかったが、11次では他に「結果表」として多数のクロス集計表が供給されていることはすでに述べた。

次節に述べるように個票データの活用が望まれる中で、個票データの供給の代わりとして豊富な「結果表」の供給は一つの答えではあろう。多くのクロス集計表が各漁業地域ごとに供給されており、それらは「一覧表」の部分的な

詳細バージョンと見ることができる。

例えば、漁業経営体の基本構成（平均漁獲金額などの項目を含む）の統計は、漁業地区毎に経営組織別の表になっている（Table 2を参照）。ここに経営組織とは、個人、会社、漁業協同組合、漁業生産組合、共同経営、官公庁・学校・試験場である。「漁家」に着目してその漁獲金額を漁業地区毎に比較する場合、個人の平均と会社の平均では大きく異なることが多いので個人経営体平均を用いることが望まれる。「一覧表」の項目との対応が分かりやすく示されていると、漁業地区単位の研究には大変貴重であろう。しかしすべての統計が経営組織別に与えられているわけではない。また実際の供給形態はExcelでの個々の表の閲覧のみを想定しているの、一覧表との連携は難しい。システムに結果表のデータを加えるのは部分的な詳細化という形で可能であるが、ユーザに詳細化の結果を分かりやすく提示するのは難しそうである。「結果表」のシステムへの活用は検討課題としたい。「結果表」群は一つ一つ独立したものとして扱うなら、印刷された報告書を閲覧するのと同じくして変わらない。

## 個票データの活用と汎用社会統計

### フォーマットについて 一まとめに代えて

漁業センサの元データは個票データであるから個票データを直接分析できることが望ましい。漁業センサにおける漁業地区は年次毎に変化していく集計単位にすぎない。個票データが開示されるかはともかく、多年次スキームを組むに当たっても個票ベースで組む方が自然である。実際、11次センサの「利用上の注意」には継続経営体、新規着業経営体、旧廃業経営体の別が記されているので経営体を多年次にわたって同定していくことは可能と思われる。11次と10次の比較を地区毎にすると、11次にあって10次にない（あるいはその逆の）地区が（それぞれ）100以上あるためにこれらの地区に属する経営体のデータは比較に反映できない。もし例えば11次個票データの各経営体レコードに10次地区情報があれば10次地区について集計することで10次データとの10次地区ごとの比較が情報損失極小で可能になる。また、集計表の集計項目（特に上位項目）は調査後に集計者の集計意図に基づいて変わるのに対し、個票の調査項目は次期センサ設計時に綿密な議論を経て決まるのであるから、やはり項目の累年比較も個票レベルで行うのが自然であろう。

個票データはプライベート情報を含むことから漁業センサにおいては個票データの開示はされていない。これに

関しては「個票開示問題」として個票データ開示の基準や方法が研究されている<sup>8)</sup>。また実際に個票データが利用されている例としては、学術研究用途に個票データが開示されている例、あるいは個票データそのものの提供はされないがインターネット上で個票に関する分析プログラムを受け付けてその分析結果を提供するサービスの例などがある<sup>9)</sup>。

最新データの追加によるデータベースの更新には、具体的には上記のような、また上記以外にもいろいろな問題がありうる。これらは一つ一つ対処するしかないが、そのために必要なことはデータおよび年次データ間の関係を表現できるフォーマットとそこから効率よく（できれば自動的に）ユーザーの必要とするデータの形を取り出せることである。ここでいうフォーマットは標準化されていることが望ましい。広く認知されることでユーザの採用や関連ツールのサポートが促されるからである。社会統計記述に関するこのようなフォーマットとしてはDDI (Data Documentation Initiative)<sup>6)</sup>が提案されている。

DDIは近年普及し出したXMLマークアップ言語を利用して社会統計データを記述するための枠組みであり、社会統計データ記述の国際標準を目指している。DDIはバージョン3.0 (DDI3<sup>7)</sup>) になって、groupingの概念を導入し調査間で比較可能な変数を扱えるようになってきている。しかしバージョン3.0はそのCandidate Draftが出された段階で公式リリースは2008年初期とされている。また、比較可能な変数の記述方法に関してはまだまだ模索の段階のように見られる。したがって漁業センサスなどに適用するのが適当かはまだ判断しかねるが、もしこのような汎用の社会統計記述の枠組みが広く採用されることになるとすれば、以下のようなことと考えられる。

- ・データの利用に必要なことはすべてこのフォーマットの中で記述することができる。
- ・よって、フォーマットに対応したツール群を使えば、利用者の必要な情報が必要な形で効率よく得られる。
- ・しかし、汎用性のあるフォーマットは抽象的で理解が難しく (DDI3.0はすでにそうである)、データをそのフォーマットで記述するにはエキスパートか専用ツールが必要であろう。
- ・しかも、多用途への汎用性を確保するために記述は特定の用途にとって不便なものにならざるを得ないであろう。
- ・したがって、別フォーマットへの変換ツールとか、特定用途に対するツールの整備が必要である。これらツールの作成にもエキスパートが必要である。

汎用社会統計フォーマットを採用するメリットは著しいが半面エキスパートをどう育てるかという問題が起こるのである。エキスパートは社会統計にもフォーマットの技術的知識にも通じていなければならない。これはこれまでの情報技術の利用には情報技術者が必要という状況とは違って、よりニッチな分野ごとにエキスパートが求められる点で各分野に問われる問題としてより深刻である。

これまで漁業センサス電子データとして提供されたアプリケーションは、データを取り出す検索方法は項目と地域の選択だけである。項目と地域だけ指定して得た細切れの結果をExcelで出力するインターフェイスでは汎用的なデータ利用に対応できていない。本システムのように拡張SQLで検索する用途にデータを利用しようとする時、用意されたインターフェイスは利用できず、元データファイルを探してその構成を推測しながら利用するしかない。十分なデータの意味記述のなされた汎用フォーマットでの供給を切に望みたい。

## 文 献

- 1) 農林水産省統計部：2003年（第11次）漁業センサス海面漁業に関する電子統計書、(2007)
- 2) 楫取和明, 井元康裕, 三輪千年：漁業センサスをはじめとする水産関連電子データの利用方法に関する一考察. 農林統計調査, 50, 19-32 (2000)
- 3) 楫取和明：水産関連電子データの利用手法について. 釜慶大学校学術交流懇談会講演論文集 (IV), 83-92 (2004)
- 4) 井元康裕, 楫取和明, 三輪千年：漁業情報システム化推進に向けて. 漁業情報利用, 第1号, 6-15 (2003)
- 5) 農林水産省：2008年漁業センサス研究会（第3回）議事概要, <http://www.maff.go.jp/www/council/council%20cont/toukei/2008gyocen/03/summary.pdf> (2007)
- 6) DDIホームページ：<http://www.ddialliance.org/>
- 7) DDI3.0ホームページ：<http://www.ddialliance.org/ddi3/index.html>
- 8) 竹村彰通：個票開示問題の研究の現状と課題. 統計数理, 51, 241-260 (2003)
- 9) 福重元嗣：わが国における官庁統計の個票利用と経済分析. 統計数理, 51, 373-388 (2003)