

# みんなくりポジトリ

国立民族学博物館学術情報リポジトリ National Museum of Ethnology

著作目録の編集にあたって：  
梅棹忠夫著作目録データベースをつくる

メタデータ	言語: jpn 出版者: 公開日: 2010-03-23 キーワード (Ja): キーワード (En): 作成者: 及川, 昭文 メールアドレス: 所属:
URL	<a href="https://doi.org/10.15021/00001114">https://doi.org/10.15021/00001114</a>

# 梅棹忠夫著作目録データベースをつくる

及川 昭文

## 1. はじめに

最初の『著作目録』は1970年早春に梅棹忠夫氏自身が、50歳の自祝事業として作成をおもいつき、作業が開始された。そのうち、中央公論社から著作集をだすことが決定し、その編集者の協力のもとに、1940年から1978年12月までの著作物（約2,200件）を収録した著作目録（以下、目録還暦版）が1979年6月に完成した。そして、1989年10月に中央公論社から「梅棹忠夫著作集」の刊行が始まり、1994年6月に別巻『年譜・総索引』が刊行され全23巻の著作集が完結した。

この目録還暦版に収録されていなかった古いものや、1979年以降の著作すべてを網羅した目録作成にかかわることになった。かかわるきっかけは、2007年たまたま民博を訪れる機会があり、その際に梅棹資料室に寄り道したことである。「2008年に梅棹先生が米寿を迎えられるので、その記念に、全著作の目録を作ることにしている」ということを三原喜久子女史から聞き、「じゃあ、何か手伝いましょうか?」ということになり、この大事業に参画することになった。

当時、客員教授として併任していた人間文化研究機構では、所属する各共同利用機関が保有する研究データベースを統合的に検索するために、研究資源共有化システムの開発事業が進められており、筆者もデータベース・システムの開発にたずさわっていた。このシステムを利用すれば、目録データベースも比較的簡単に作成できると考え、気楽にデータベース作りを引き受けた。ところが、この著作目録には著作物に対する梅棹哲学ともいべきものがあり、一筋縄ではいかないことがすぐに判明した。本稿では、2007年に梅棹資料室を訪れたときから、今日までの悪戦苦闘について報告する。

## 2. 著作物とは——梅棹忠夫の場合——

今回のデータベース作りにおいて、もっとも苦勞したことの一つは、「著作物」の単位をどのようにするかということであった。すなわち、著作物に対して梅棹氏は独特の考え方をもっており、それをいかにしてデータベースに反映するかということが大きな課題であった。前述の目録還暦版に、「著作目録をつくる」（本書にも転載）と題して、梅棹氏が著作物に対する考え方を述べている。

公表された刊行物で、それが自分の著作物であるといえるための基本的条件は、ふたつある。それは、権利と責任とである。その著作の内容を、無断で転載されたり、盗用されたりしたとき、法律にうたえなくても著作権を主張できるか、ということ。もうひとつは、その著作物の内容について、ほかからなんらかの発言があった場合、それに応答する用意があるか、ということである。このふたつの条件を満足させる形式的要件として、わたし自身は、その著作物における署名ということをもっとも重要だとかんがえたのである。その著作物の著作者として、その著作物に自分の名が明記されているかどうか、それがきめ手だ、というのである。

わたしは、原稿または校正刷の段階で自分で内容に手をいれることができたものは、インタビュー記事や談話記事でも、著作にかぞえることにしている。内容についての権利はもちろん、責任ももてるからである。新聞社などの電話インタビューは、わたしは、おことわりすることにしている。記事の正確さについて、とうてい責任がもてないからである。著作者のかかわりあいかたとして、編、共編はよいとして、編集委員、監修というのはどうであろうか。じつは、わたしは最近まで、編集委員や監修者として名をつらねている刊行物を、わが著作とみなしていなかった。著作棚にもなく、カードもなかった。しかし、江阪氏と議論しているうちに、これらのものもやっぱりわが著作物とみなすべきだという結論になった。

このような考え方から図1にあるような他人の本を推薦する言葉を「オビ」に書いていた場合、この「オビ」もまた論文と同じように著作物として扱われることになる。この場合、一般的な目録規則にいう「タイトル」や「ページ数」は存在しないため、目録規則に準拠したデータベースのそれぞれの項目をどのようにして完成させるかという問題が生じることになる。

編集委員や監修者になっているものを著作物として扱うことには、データベース作

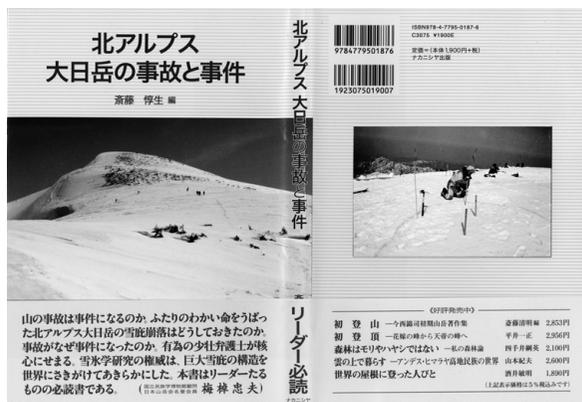


図1 「オビ」の例

成上それほど大きな問題は生じない。すなわち、編集委員や監修者を「著者」として扱えばよいからである。しかし、次のようなケースが生じることがある。

「朝日講座 探検と冒険」(全8巻)の編集委員は「加納一郎, 泉靖一, 梅棹忠夫, 樋口敬二, 本多勝一」となっているが, ある巻を「梅棹忠夫」が編集したとすると, その巻に対する関わり方は「編集委員」と「編集担当」という二つの関わり方が生じることになる。すなわち, 図2にあるように「編集委員」としての著作と「編集担当」としての著作の二つのレコードが作成される。このことについては, 梅棹氏は前述の「著作目録をつくる」の中で次のように述べている。

#### 著作物の異同と照合

いずれも, 一見複雑なようだが, 論理的には首尾一貫して, 検索上は便利である。要するに, 前節にのべたような書誌的記載事項, すなわち, 標題, かかわりあいかた, 所載刊行物名, 刊行年月日, 発行所, 所載ページの各事項がすべて一致する場合は同一著作物であるが, そのうちの1項あるいはそれ以上の事項が一致しない場合は, 別項目をたてる。たとえば, 重刷や定価改定などは, 上記の事項について差がないから, 別項をたてない。……

筆者注) ここでいう「別項目」は別レコードのことである。

新聞記事もデータベースの一つの著作物として作成することになるが, 作成上問題となる場合がある。これについては, 次のように述べている。

もうひとつこまったのは, 通信社の場合である。記事は棒ゲラの形で全国の各新聞社に配信されるが, それが, 何日にどの新聞に掲載されたかは, 通信社においても, とうてい確認できないのだという。もちろん, 掲載紙の現物は, 著者の手にわたらないのが通例である。やむをえず, 配信原稿を入手して, その配信の日づけをもって刊行の日づけとしたが, 刊行物の現物の検索という原則からは, この場合だけは, はずれることになった。

これは現物(印刷されたもの)が存在しない著作物ということで, どこかの新聞社で記事として掲載されて初めてその存在が確認できるということになる。その場合には, これらの両者の関係を明確にしておく必要がある。すなわち, 「配信元」, 「配信先」といった相互参照ができるような仕組みをデータベース上で作り上げておかなければならない。

以上, いくつかの課題について述べてきた。手作業で編集をおこなう場合は, おおまかな編集ルールを定めておき, 問題があれば臨機応変に対応することも可能である。しかし, データベースの場合は, 事前に十分な検討を行い, あとで変更する必要のないような編集ルールを定め, それに対応したデータベースの構成を考えておくことが重要になってくる。

項目名称	入力内容
レコードID	11495
文献番号	1972012502
種別コード	7
種別	
種別表示	編集委員
言語	
タイトル	「朝日講座 探検と冒険」1
サブタイトル	
著者名	加納一郎; 泉靖一; 梅棹忠夫; 樋口敬二; 本多勝一
司会・聞き手	〈編集担当〉梅棹忠夫

項目名称	入力内容
レコードID	11496
文献番号	1972012503
種別コード	7
種別	
種別表示	編集担当
言語	
タイトル	「朝日講座 探検と冒険」1
サブタイトル	
著者名	梅棹忠夫
司会・聞き手	〈編集委員〉加納一郎; 泉靖一; 梅棹忠夫; 樋口敬二; 本多勝一

図2 編集委員としての著作（上）と編集担当（下）としての著作

### 3. データベース作成ツール——nihuONE について——

今回のデータベース作成のプラットフォームとしては、nihuONE というデータベース・システムを利用することにした。その最大の理由は、このシステムは筆者が設計したもので、その機能や使い方に熟知していたということである。また、今回のようにデータの修正や項目の変更が頻繁に行われると予想されるデータベースに対応した便利な機能が豊富にあるということも理由の一つである。

このnihuONEは、人間文化研究機構が2005年度から3年計画で実施した「研究資源共有化」事業の中で、筆者がその責任者として開発を担当したシステムである。nihuONEは2007年の4月に試行運用を開始し、2008年の12月から本格的な運用が始まった。以下、このnihuONEについて、その開発理念、特徴、機能などについて説明する。

#### 3.1 nihuONE で目指したもの

nihuONE 開発においてもっとも重視した目標は2つある。ひとつは「人文系の研究

者でも SE やプログラマなどの情報技術者の支援なしで、データベースの作成から Web での公開までできる」ことである。

多くの研究者は研究に必要な資料やデータをカードにしたり、図表にしたり、あるいはファイルフォルダーにまとめたりして整理している。データベース化とはこれらの資料やデータをコンピュータ上の仮想空間に移し替える作業に他ならない。言い換えれば、それはコンピュータにとって理解しやすい、処理しやすい、管理しやすい形態への変換作業である。このことは、その形態がどのようなものかを熟知していないと、データベース化に失敗する恐れが大きいことを意味しており、そのことがデータベース化を難しいものになっている。したがって、その変換作業が研究者が日常的に行っている資料やデータの整理作業の延長として行えれば、データベース化は困難なものではなくなってくる [及川・山元 2001]。nihuONE では、エクセルやワープロなどで作成したデータを簡単な手順でデータベース化できるようになっており、それらのツールは文系の研究者でも容易に使いこなせるインターフェースとなっている。

もう一つの目標は「単なる検索ツールとしてではなく、データベースを活用できる研究支援ツールとして機能する」ことである。本来データベースは蓄積されたデータを分析したり、人事システム、給与システムあるいは大学における学務システムなどの業務を効率よく運用するために開発され、発展してきたものである。ところが情報化時代の到来とともに大量の情報の中から必要なものを探し出す、いわゆる検索機能が重要視されるようになった。その結果検索やそれに関連した機能や性能は、ハードウェアや OS の発展とともに格段に拡充してきているが、データベースに格納されているデータを分析したり、活用するための機能は不十分なままの状況にある。nihuONE は、検索機能のみでなくデータベースの分析やそれらを活用するための機能を充実させることを大きな目標とした [及川・藤沢・洪・山元 2007]。

### 3.2 システム運用とデータベース運用の分離

一般的に DBMS (DataBase Management System) の運用においては、システム管理とデータベース管理は同じ管理者が兼ねることが多い。このため DBMS に登録されるデータベースが多くなればなるほど、管理者の運用に関わる負荷は大きくなる。結果として、それぞれのデータベースのアップデートや保守に支障をきたすことが少なくない。nihuONE ではこのような状況を回避するために、システム運用とデータベース運用を完全に分離している。

nihuONE の利用者は表 1 のように 4 種類に区分される。nihuONE での基本的なデータベース作成プロセスは、①データベースの登録 → ②データベースの定義 → ③項目の定義 → ④表示 (一覧, 詳細) 形式の定義 → ⑤データのアップロードとなる。このプロセスにおいて、admin が関与するのは、①のデータベースの登録のみである。

この登録処理では、「データベース識別記号」と「データベース管理者」の設定を行い、それ以後の作業はすべてデータベース管理者の仕事となり、システム管理者はデータベースの管理・運用の作業から解放されることになる。このことは多数のデータベースを公開、運用している部署にとっては、システム運用にのみ専念すればよく、非常に大きなメリットとなる。

一方データベース管理者は、自分が作成し、その内容を熟知しているデータベースのみを管理するわけであるから、たとえ複数のデータベースを管理するとしても、運用の作業量は大きな負担とはならない。むしろ、システム管理者に依頼しなくても、自分の判断でそれぞれのデータベースを管理・運用できることは大きな利点といえることができる。

表1 nihunoONEの利用者区分

区 分	権限・役割など
admin	システムの管理者で、一人だけ登録できる。その役割は利用者の管理、データベースの登録・削除、利用統計の閲覧、システムのバックアップ・リカバリなどである。
DB 管理者	データベースの管理・運用を行う者で、各データベースに必ず一人だけ設定する。
登録利用者	認証を必要とするデータベースの閲覧やデータのアップロード・ダウンロードの権限を付与された利用者である。
一般利用者	一般公開されているデータベースのみを閲覧できる利用者である。

### 3.3 マイ・データベースの実現

一般的に人文系のデータベースは標準化が困難で、研究者それぞれの研究内容や成果と深く関連している。したがって、対象となる資料が同じであっても、作成されるデータベースは研究者ごとに異なったものとなる性質を本質的に持っている。研究資源の共有化はその分野の研究の発展には重要な要素であるが、個々の研究者の知的生産を支援するためには、まずそれぞれの研究者のニーズに応えたデータベース（以下、マイ・データベースと呼び、MyDB と称する）が作られることが肝要となってくる [及川・山元 2007]。

多くの場合、データベースは周到な準備を経て、組織的、計画的に作成されるものであるが、研究者個人が分析の対象として、あるいは研究の成果をまとめるためにMyDBを作成する場合には、まずは作ってみて、不具合や不都合があれば項目の定義を変更したり、再編集したり、試行錯誤を繰り返しながら作られていくことになる。これらの作業を人文系の研究者が容易に行うことができるためには、以下のような要件が満たされている必要がある。

### 1) 作るのが簡単

MyDB の作成過程としては、一般的に以下の3つのケースが想定される。

- ① 1次資料を集めることから始める。あるいは1次資料は収集済みであるが、電子化はまだで、まったくゼロから MyDB を作成する。
- ② 1次資料の電子化はできており、Excelなどのソフトを利用してそれなりに活用している。このデータをもとに MyDB を作成する。
- ③ 既存のデータベースを加工、再編集したり、あるいは新しい資料を追加したりして MyDB を作成する。

いずれのケースでも、単純かつ容易な作業で MyDB を作成できることが、システムとしての機能要件になる。たとえば①の場合、Excelなどは使えなくても、ワープロなどで決められた書式（できるだけ簡単な）でデータをテキストとして入力すれば、それがそのままデータベースにアップロードできること。②の場合では Excel ファイルを CSV 形式のファイルに変更するだけでデータベースにアップロードできること。③の場合であれば、既存のデータベースから必要な項目を CSV、タブ区切りなど多彩な形式で簡単にダウンロードできることなどである。

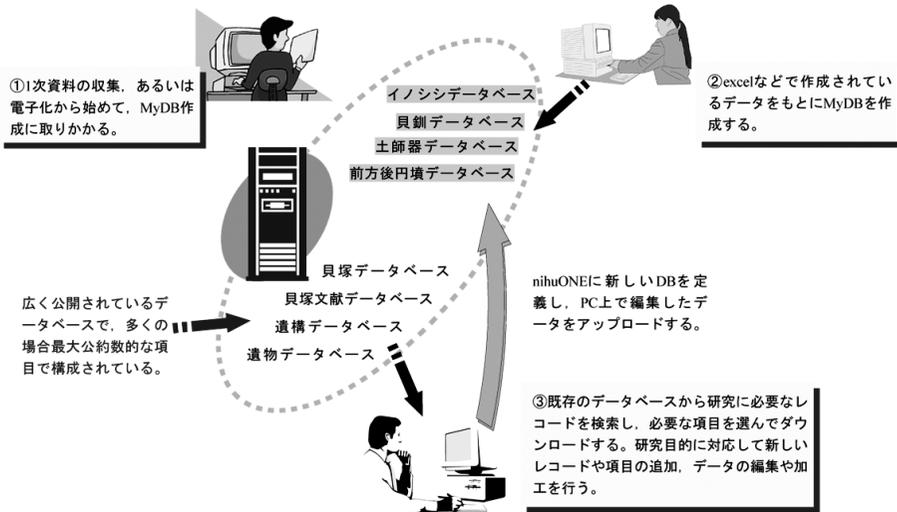


図3 マイ・データベースのイメージ

表2 項目定義 — 属性一覧 —

属性	内容
タグ番号	項目を識別する3桁の数字
名称	項目名称
表示名称	一覧表示などの際に表示する項目名称
ID	ID項目かどうかの指定 (Yes/No) <sup>1)</sup>
データ型※	データの種別を表すコード 1 = テキスト (全角のみ)                      12 = 数値 (カンマを含む) 2 = テキスト (半角のみ)                      13 = 数値 (小数点を含む) 3 = テキスト (全角, 半角)                    14 = 数値 (カンマ, 小数点を含む) 11 = 数値 (カンマ, 小数点を含まない)      20 = ファイル
最小値※	データ型が数値の場合の最小値を指定 (不明の場合は空白)
最大値※	データ型が数値の場合の最大値を指定 (不明の場合は空白)
テキスト長※	データ型がテキストの場合の最大バイト長を指定 (不明の場合は空白)
必須項目※	必須項目の指定 (Yes/No)
検索対象	検索対象の指定 (Yes/No)
KWIC 検索	KWIC 検索対象の指定 (Yes/No)
区切り文字※	デリミッタ項目 <sup>2)</sup> の場合, 使用する区切り文字
ソート	デリミッタ項目の場合, 区切られた文字列単位でのソーティングの指定 <sup>3)</sup>
html	html テキストかどうかの指定 (Yes/No)
頻度統計	検索結果の集合に対して頻度統計の対象とするかどうかの指定 (Yes / No)
基本統計	検索結果の集合に対して基本統計の対象とするかどうかの指定 (Yes / No)
Table 変換※	Table を利用したテキスト変換の指定
リンク対象	他のデータベースへのリンク指定
ペアチェック※	著者名と著者名読みのように必ず同時に存在する項目の指定
辞書チェック※	辞書を参照してその辞書の中に存在するかどうかをチェックする
公開・非公開	公開・非公開の指定
ダウンロード	ダウンロードの可否の指定

<sup>1)</sup>: 複数の項目を ID 項目として指定可能。その場合すべての項目値を連結したものが ID として認識される。  
<sup>2)</sup>: 「デリミッタ項目」とは、一つの項目に複数のキーワードなどがあらかじめ指定された区切り文字 (デリミッタ) で区切って入力されている項目のことである。  
<sup>3)</sup>: デリミッタ項目の場合, 区切り文字で区切られた文字列単位でソーティングして格納しなおすかどうかを指定する。例えば, 「アサリ, シジミ, マシジミ, ハマグリ」と入力されていた場合, これを「アサリ, シジミ, ハマグリ, マシジミ」とソートする。  
 ※印: QC (Quality Control) 項目

## 2) 容易な定義変更

MyDB は研究と密着したデータベースであり, 研究の進展に伴って, 項目追加や項目の属性や内容の変更が必要となることが少なくない。また, 研究の結果をデータ

ベースに反映し、そのデータベースを再び分析に活用するということもある。つまり、研究の進展にもなってデータベースの構成や項目定義の変更を容易に行えることが不可欠になる。

nihuONE ではこれらの処理を画面上で簡単に行えるインターフェースを備えている。項目定義では表2で示す属性を各項目ごとに設定することになるが、事前にExcelなどで作成しておき一括してアップロードすることもできる。また定義が終了した時点でCSVファイルとしてダウンロードする機能もある。

項目定義は、すでに入力されているデータの内容との整合性が保証されれば、データベース作成後に変更することもできる。たとえば、検索項目となっていない項目をあらたに検索対象としたり、区切り文字を追加したり、最小値や最大値の値の変更などである。また、項目そのものを新しく追加したり、削除することも可能である。

### 3) 高度な品質管理

データベースの利用価値を左右するものの一つはデータの品質である。多くのデータベース・システムにおいては、品質管理（Quality Control, 以下 QC）は原則として利用者の責任とされており、十分な QC 機能は備えていない。nihuONE においては、データの品質を高めるための QC 機能を充実させ、システムでさまざまなエラーを発見できるようにしている。QC は項目定義でパラメータを入力しておくことで、データのアップロード時に実行される。表2は項目定義で設定しなければならない属性の一覧であるが、そのうち属性に※印がついているものは QC 項目である。

**データ型チェック：**その項目が表2の「データ型」のどれにあたるかは必ず指定しなければならない。それぞれの項目データがこの指定に適合しているかどうかを調べる。

**ペアチェック：**文献目録データベースなどで「著者名」「著者名よみ」のように必ず同時に存在しなければならない項目があった場合、いずれかが欠落していないかを調べる。

**Table 変換：**入力されたデータが、DB 管理者によってあらかじめ登録されているテーブルに含まれているかどうかを調べる。テーブルには変換する値（数値でもテキストでも可）も設定されており、エラーがなければその値を指定された項目のデータとして展開する。

**デリミッタ項目の2重チェック：**たとえば「シジミ、アサリ、ハマグリ、カキ、アサリ」と入力されたデータを、「アサリ、~~アサリ~~、カキ、シジミ、ハマグリ」とソーティングし、同じ文字列がある場合は自動的に削除する。

**辞書チェック、辞書変換：**あらかじめ nihuONE にデータベースとして登録されている辞書と照合し、入力されたデータがその辞書に含まれているかどうかを調べる。エ

ラーがなければ、その辞書の任意の項目値を新しい項目データとして展開する機能もある。

### 3.4 その他の便利な機能

nihuONE は電子化されたさまざまなデータを容易に活用するための機能も充実している。これらの機能は、特別な知識や技術がなくても利用できるユーザ・インターフェースを備えている。

#### 1) リンク機能

XML 型のシステムをベースにしていることから、階層構造を持ったレコードを構築することも可能であったが、nihuONE では“simple is best”の精神で、レコード構造は可能な限り単純な構造とした。その代わりに、階層的な関係を実現するために「リンク機能」を実装した。この機能は次のような手順で実現する。

- (1) 項目定義の際に、あらかじめリンクするデータベースの識別記号を設定する。この項目を「リンク項目」と呼ぶ。設定できるデータベースは nihuONE 上のデータベースであれば何れでも、すなわち同じデータベースでもよい。
- (2) リンク項目にリンク対象となるデータベースのレコード ID を入力しておく。

データベース名	UMEO3 梅棹目録_090106	データベース名	UMEO3 梅棹目録_090106
項目名称	入力内容	項目名称	入力内容
ID	14393	ID	10013
文献番号	1990042001	文献番号	1940111701
種別コード	1	種別コード	1
種別		種別	
種別表示	著	種別表示	著
言語		言語	
タイトル	「梅棹忠夫著作集」第1巻	タイトル	白頭山をこえて満州へ
サブタイトル		サブタイトル	
著者名	梅棹忠夫	著者名	梅棹忠夫
司会・聞き手		司会・聞き手	
掲載誌	『探検の時代』	掲載誌	『時報』
掲載誌属性	〈編集担当〉栗田晴之 〈編集委員〉石毛直道; 栗田晴之; 佐々木	掲載誌属性	第15号
シリーズ名など		シリーズ名など	
連載回		連載回	
出版年	1990	出版年	1940
出版年月日	1990.4.20	出版年月日	1940.11.17
出版者	中央公論社	出版者	関西学生山岳聯盟
判型	A5判		
頁	589p.		
掲載箇所			
著作集収録巻			
リンク元	10013; 10014; 10017; 10018; 10019; 10042; 10092; 10123; 10134; 10137; 10371; 10394; 10439; 10583; 10616; 10637; 10638; 10641; 10646; 10648; 10880; 11058; 11497; 11796; 11828; 11946; 12058; 12501; 12775; 128		

リンク元には「梅棹忠夫著作集」第1巻に収録されている著作物の「レコード ID」が入力されており、このレコード ID をクリックすると、右図のように当該著作物が表示される。

図4 リンク機能の実例

(3) 詳細表示画面で、リンク項目に表示されている ID をクリックすれば、当該レコードが別画面で表示される。

## 2) html テキストの活用

項目定義で項目の属性を“html テキスト”と指定しておく、その項目に入力された文字列は html テキストと認識される。この機能を利用することによって、nihuONE 以外のサーバにアクセスし、そのサーバにある画像を表示することのようなことも実現できる。図 5 はその例で、ここで表示されている画像は国立民族学博物館（以下、民博）のサーバにある画像である。

また、このデータベースは民博のコスチューム・データベース（MCD: Minpaku Costume Database）を nihuONE 評価のために移植したものであるが、nihuONE の諸機能を活用することによって、短期間のうちに移植することができた [中川・高橋・及川 2007]。



図 5 国立民族学博物館にある画像の表示例

## 3) SDF 形式

nihuONE ではシステムへアップロードするファイルは、原則として DB 管理者の PC 上にあることを想定している。利用者が管理できるファイルの形式としては一般的な「CSV」「タブ区切り」と呼ばれるファイル形式と SDF (Standard Data Format) と名付けた nihuONE 独自のファイル形式の 3 種類がある。

SDF とは図 6 のように「タグ番号」「|」（区切り文字として使用）」とデータから構成されたテキストファイルである。したがって、ワープロなどで作成したり、編集することが可能となる。この SDF 形式を設定した主な理由は、① CSV, タブ区切りの形式では挿入できない「改行」文字を、テキストに含めることができる、②ワープロなどで編集することができるの二つである。

010		10713
020		1965040101
030		1
050		著
070		知的生産の技術について
080		第 1 回 発見の手帳
090		梅棹忠夫
110		図書
140		4 月号 第 188 号
150		1965
160		1965.4.1
170		岩波書店
180		pp.2-6
230		10795 ; 10848 ; 11107 ; 11529
240		抜粋 ; 抜粋 ; 収録 ; 載録
910		701
930		1966020501 ; 1966112001
932		抜粋 ; 載録
940		377 ; 404 ; 631 ; 1091

図 6 SDF 形式の例

#### 4) ファイル型項目の活用

項目属性のデータ型に「ファイル」がある（表2, 図7参照）。この項目に入力されるのはファイル名で、ファイル本体は事前に nihuONE のサーバ上に転送しておく。詳細表示画面でこのファイル名をクリックすると、ブラウザがファイルの拡張子に応じた処理を実行する。図7は幕末明治地勢地図の詳細表示で項目「索引地図」に入力されている pdf ファイル名をクリックすると下図のような地図が表示される。利用者はこの地図をダウンロードして自由に利用することができる [鎌田・及川 2007]。

データベース名	EMPO1 幕末明治地図
項目名称	入力内容
国ID	K01
国名	山城
国名よみ	やましる
郡数	8
郡ID	G01001; G01002; G01003; G01004; G01005; G01006
郡名	葛野; 愛宕; 乙訓; 紀伊; 宇治; 久世; 綴喜; 村
郡名よみ	かどの; あたご; おとくに; きい; うぢ; くせ; つづ
隣接国数	6
隣接国ID	K02; K03; K05; K06; K21; K41
隣接国名	大和; 河内; 摂津; 伊賀; 近江; 丹波
府県ID	M26
府県名	京都府
師管ID	S04
師管名	第4師管
索引地図	K01-1
pdf_files	K01-1; K01-2; G01-1; G01-2

↑

項目「pdf\_files」にある地図番号をクリックすると、右図のような pdf ファイルになった地図が表示される。

図7 ファイル型項目の事例

## 4. データベースを作る・分析する

近年 FileMaker, Access, Oracle などのパーソナルなデータベース・ソフトが開発され、以前と比べれば格段にデータベース作成は容易になったといえる。しかしながら、いずれのデータベース・ソフトを利用するにしろ、高品質なデータベースを作るには、周到な準備と専門家の支援が必要不可欠である。今回はデータベース・ソフトとして、開発されたばかりの nihuONE を利用したが、データベース作成とともに nihuONE の性能・機能評価も行った。

### 4.1 データベースを作る

一般的なデータベースの作成手順は図8のようになる。今回のデータベース作成を

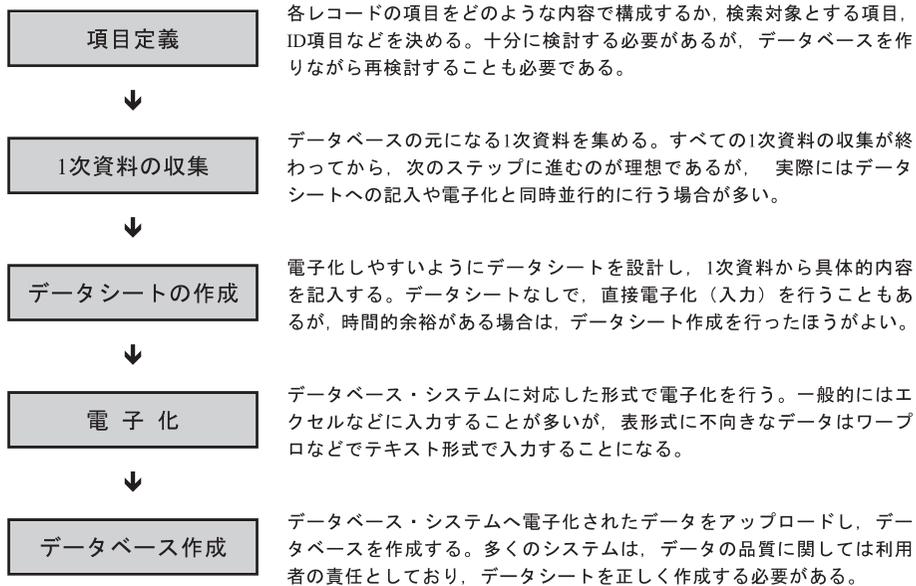


図8 一般的なデータベース作成手順

この手順に照らして検討したが、いくつかの課題が明らかになった。

- ・ 項目定義は基本的には標準的な目録規則に準拠すべきであると考えるが、著作物に対する梅棹哲学を反映するとともに、1979年に作成された「梅棹忠夫著作目録」と大きく異ならないようにしなければならない。
- ・ 最終的には冊子体として印刷することから、印刷のための編集がやりやすいように設計されている必要がある。
- ・ 1次資料の収集については、すでに梅棹資料室に保管されているから、ほとんど問題はないが、すでに電子化されているデータがあり、ワークシート作成や電子化は、それらのデータを可能なかぎり利用できるようにする必要がある。

これらの課題を検討した結果、梅棹忠夫著作目録データベースは、表3のような手順で作業を進め、まずどのような項目で構成するかを決めた（表4参照）。

表4に示した項目定義にもとづいて、具体的な入力マニュアル（図9）を作成した。このマニュアルに即して、新規レコードの入力、およびステップ-3の編集処理をおこなった。ステップ-3でのプログラムによる編集はかなり複雑な処理となった。主な処理としては以下ようになる。

表3 梅棹忠夫著作目録データベース作成手順

ステップ-1	既存の電子化データを元データとして、nihuONE上に仮データベースを作成。
ステップ-2	作成された仮データベースを分析し、項目定義を行った。
ステップ-3	ステップ-2での項目定義と整合性をとれるように仮データベースのすべてのレコードを編集。この編集はプログラムで行ったため、編集できなかった項目もあり、それらは手作業での修正となった。
ステップ-4	電子化されていないデータを項目定義にしたがいエクセル・データとして入力。
ステップ-5	ステップ-3で編集した仮データベースと、ステップ-4で作成したデータを元にしてデータベースを作成。
ステップ-6	ステップ-5で作成したデータベースの全レコードをSDF形式でプリントアウト。ページ数は約4,300であった。
ステップ-7	SDF形式のプリントアウトの校正、データ修正、入力漏れの著作物のデータ入力などを手作業で実施。
ステップ-8	ステップ-7で作成したSDF形式のデータをnihuONEにアップロードし、最終的なデータベースの作成。
ステップ-9	データベースの項目を編集して印刷用データを作成。

表4 梅棹忠夫著作目録データベース項目一覧

タグ	項目名	備考
010	レコードID	5桁の数値
020	文献番号	yyyymmdd + 2桁で10桁の数値
030	レコード種別コード	著作物とのかかわりあい方の種類コード
040	レコード種別	レコード種別コードによるテーブル変換(自動生成)
050	レコード種別表示	表示用著作物とのかかわりあい方
060	言語	日本語以外の言語を記述
070	タイトル	著作物のタイトル
080	サブタイトル	著作物のサブタイトル
090	著者名	著者名(座談会、編集委員などの場合は、全員を記述)
100	司会・聞き手	対談の聞き手、パネルディスカッションや討論会などの司会者
110	掲載誌	
120	掲載誌属性	
130	シリーズ名など	
140	連載回	
150	出版年	yyyyの4桁
160	出版年月日	yyyymmddで格納
170	出版者(出版社)	
175	判型	
180	ページ	
190	掲載箇所	
200	著作集収録巻	
210	リンク元	元になった文献のレコードID
220	リンク元との関係	参照関係(掲載、転載、再録、抜粋など)
230	リンク先	リンク先文献のレコードID
240	リンク先文献との関係	参照関係(転載、収録、再録、抜粋など)
250	配信リンク元	
260	配信リンク先	
900	備考	

項目データ入力マニュアル	
010	レコード ID
	半角数字
	現在アクセスに入力されているレコードに対しては、「10001～」の ID を設定する。今後 nihUONE で新たに追加されるレコードに対しては、「20001～」の ID を設定する。
020	文献番号
	半角数字
	出版年月日を基に以下のようにつける。
	①年月日が既知の場合
	2001年3月21日 → 20010321**
	②日が不明の場合
	2002年11月 → 20021199**
	③月日が不明の場合
	2005年 → 20059999**
	④春号, 秋号などで月日が不明の場合
	2006年春号 → 20062199**
	2006年夏号 → 20062299**
	2006年秋号 → 20062399**
	2006年冬号 → 20062499**
	⑤配信レコードの場合は, 原稿依頼日あるいは原稿提出日を基にする。
	1996年5月19日に原稿を提出したとすれば, → 19960519** とする。
	⑥年月日がまったく不明の場合は, → 9999999999 とする。
030	レコード種別コード
	半角数字
	以下の種別にしたがって, コード (2桁の数字) を入力する。
	1 = 著作
	2 = 共著

図9 入力マニュアルの一部  
注) 実際の作業過程ではかなりの変更があった。

## 1) 不要項目の削除

すでに電子化されているデータは、パーソナル・コンピュータのない時代に、ワープロ専用機やホスト・コンピュータを利用して入力作業を行ったものが含まれていた。当時のコンピュータの機能や性能の制約から、タイトルの読みをカタカナで入力した項目、1項目に入力できる文字数の制限から、本来一つの項目とすべきところを複数の項目に分けて入力された項目など、負の遺産ともいえるべき現在では不要と思われる項目が多数あったため、これらの項目を削除した（図10参照）。

## 2) 項目の分離, 新項目の作成

データベースとしては別々の項目として格納しておいたほうがよいと思われる項目

	A	B	C	D	E	F
1	整理番号	著者	著者2	著者3	著者4	著者(ヨミ)
2	1952122001	[編]梅棹 忠夫				メノウメサオ タダオ
3	1953070501	[著]梅棹 忠夫				メノウメサオ タダオ
4	1953071501	[著]梅棹 忠夫				メノウメサオ タダオ
5	1953080101	[著]梅棹 忠夫				メノウメサオ タダオ

図 10 現在では必要ない仮名振りの項目例

F	G
著者(ヨミ)	掲載誌
メノウメサオ タダオ	都府立第一中学校山岳部編『山城三十山記 上篇』「部報」第3号
メノウメサオ タダオ	京都府立京都第一中学校 山岳部 編『山城三十山記 上篇』「部報」第3号
メノウメサオ タダオ	京都府立京都第一中学校 山岳部 編『山城三十山記 下篇』「部報」第4号
メノウメサオ タダオ	京都府立
メノウメサオ タダオ	掲載誌
メノウメサオ タダオ	京都府立京都第一中学校山岳部編『山城三十山記 上篇』
メノウメサオ タダオ	掲載誌属性
	シリーズ名など
	部報
	連載回
	第3号
	出版年
	1934

掲載誌の項目に

「京都府立第一中学校山岳部編『山城三十山記 上篇』「部報」第3号  
と入力されているのを

掲載誌=京都府立第一中学校山岳部編『山城三十山記 上篇』

シリーズ名など=部報

連載回=第3号 と複数の項目に分離。

図 11 分離した項目例

が、印刷用に編集した形で格納されていた。たとえば、「著作物とのかかわりあい方」と「著者名」が、「〔著〕梅棹忠夫」、「〔編集委員〕梅棹忠夫」のように1項目になっているのを、「著作物とのかかわりあい方=〔著〕」、「著者名=梅棹忠夫」と2つの項目に分離し、新しい項目を作成した。また、逆に2つ以上の項目に分けて入力されているものを、一つの項目にまとめるという処理も行った(図 11 参照)。

### 3) 参照関係データのチェックと再構成

目録還暦版には図 12 のように「転載→641001」、「収録→670106」、「全集に収録→740502」という記載がある、これはこの著作「文明の生態史観序説」が、「641001 中央公論 10月号(1964.10.1)」に転載、「670106 文明の生態史観」に収録、「740502

戦後日本思想体系 15」に収録されていることを意味している。表 4 の項目定義でリンク項目となっているところであるが、このデータが不確かなレコードが数多くあった。そこで、プログラムでこの参照関係のデータを相互に適切かをチェックした。不適切である場合は、エラーリストを出力して手作業での修正を行った(図 12 参照)。

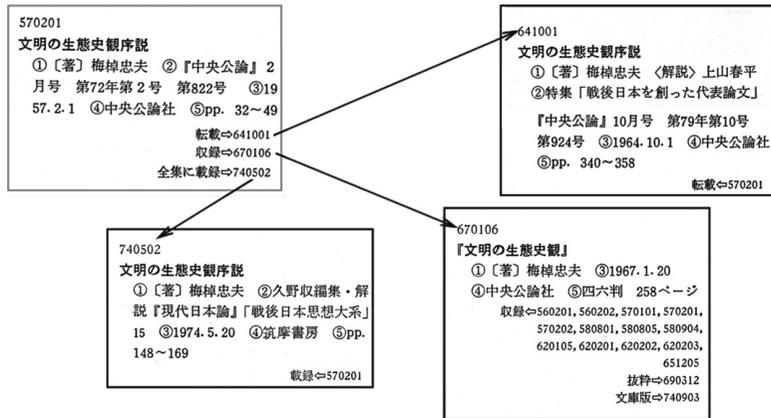


図 12 参照関係のあるレコードの例

#### 4) 全角, 半角文字の統一, 余分な空白の削除

アルファベットや数字が全角と半角で入力され、不統一であったため、アルファベットと数字は基本的に半角文字とすることとし、プログラムで自動的に変換した。また、余分な空白を削除した。この空白は「分かち書き」的意味を持っていたのではと推定される。

## 4.2 データベースを活用する

2008年12月にほぼすべてのレコードのチェックが終了し、その修正作業を終えることができた。まだ細かなエラーが残っているが、nihuONEの持っている機能を活用していくつかの統計資料を作成したので、ここで紹介する。ここに表れている数字は、最終的に変わる可能性があるが、傾向としては大きな差はないものと思われる。

### 1) KWIC リストの作成

KWICとは、Keyword In Contextのことである。KWICリストは、ある言葉がどのようにテキストの中で使用されているかを、その言葉の前後のテキストを表示してリストを作成するものである。nihuONEにはこのKWICリストを作成する機能を標準的に備えており、図13のように「詳細検索画面」で、「KWICリスト」を選択し、キーとなる言葉を入力すれば、瞬時に図のようなKWICリストが作成される。このリストをExcelなどにダウンロードして同じようなリストを簡単に作成することもできる。

### 2) 頻度集計

nihuONEの便利な機能に「頻度集計」機能がある。これは、検索結果の集合に対し

検索語  キーとなる言葉「知的」を入力

データベース ■ UME03 梅棹忠夫著作目録\_200

表示件数  10件  20件  30件  50件

文字列項目検索

検索方法  AND  OR

検索項目

<input type="checkbox"/> 種別	<input type="checkbox"/> 種別表
<input type="checkbox"/> 著者名	<input type="checkbox"/> 司会・問
<input type="checkbox"/> 出版者	<input type="checkbox"/> 判型
<input type="checkbox"/> リンク元と	<input type="checkbox"/> リンク先
<input type="checkbox"/> 備考	<input type="checkbox"/> 備考ト
<input type="checkbox"/> 備考リン	<input type="checkbox"/> 備考ト

数値項目検索

比較方式  等しい (=)  等しくない (!=)

検索項目  レコードID  文献番

KWIC検索

左右表示文字列長  20  30  40  50  60

ソート  前方  後方

検索項目

<input checked="" type="radio"/> タイトル	<input type="radio"/> サブタイ
<input type="radio"/> シリーズ名	<input type="radio"/> 出版者
<input type="radio"/> リンク先	<input type="radio"/> リンク先

知的 生産の技術の三つの原理(その1)
知的 生産の技術の三つの原理(その2)
知的 生産の方法
知的 生産の空間設計
知的 発見の楽しみ
知的 遊戯の場
知的 生産都市・京都
どうする 知的 生産都市・京都
ワードプロセッサは“ 知的 生産の新しい道具”になりうるか
ビジネスマン “ 知的 武装” の技術
「 知的 生産の技術」のひとつ “手紙の形式”について
情報産業社会における「 知的 生産」
情報産業社会における「 知的 生産」
『 知的 市民と博物誌』
『 知的 生産の技術』その後
『 知的 生産の技術』その入門書や
新聞『日刊総合雑誌』『 知的 発見』活字ならでは
創造的な知的生産のノウハウ『 知的 生産の技術』
役立つ 知的 な行動人
創造的な 知的 生産のノウハウ『知的生産の技術』
みえないなりの 知的 生産
現代の 知的 状況の裏側のいい見取り
私の 知的 生産の技術
私の 知的 生産の技術
『私の 知的 生産技術』
中東研究の 知的 フォーラム
中東研究の 知的 フォーラム
博物誌の 知的 サービス
最も 知的 会話を楽めたいのは司馬さん
信頼できる 知的 案内書
肥沃なる 知的 領域への新規参入を
新・ 知的 生産の技術の明確

知的 生産の空間設計  
 知的 発見の楽しみ  
 知的 遊戯の場  
 どうする 知的 生産都市・京都  
 ワードプロセッサは“ 知的 生産の新しい道具”になりうるか  
 ビジネスマン “ 知的 武装” の技術  
 「 知的 生産の技術」のひとつ “手紙の形式”につ  
 情報産業社会における「 知的 生産」  
 情報産業社会における「 知的 生産」

上：ダウンロードしたファイルを編集して作成した KWIC リスト

図 13 KWIC 検索画面と KWIC リストの例

て、ある項目の値についてその頻度を集計するものである。具体的には、

- ①検索結果一覧が表示されている画面で、「頻度統計」をクリック。… 図 14-①
- ②集計を行う項目を選択。… 図 14-②  
注) 複数の項目でクロス集計を行うことも可能である。
- ③集計表の表示。… 図 14-③

となる。この例では、データベースの全レコードを対象に「著作物とのかかわりあい方」について、頻度の集計を行っている。

この頻度集計結果は、CSV 形式でダウンロードすることができるようになっており、それをエクセルなどで編集して表を作成することも簡単に行える。

表 5 は、「著作物とのかかわりあい方=対談」となっているレコードを検索し、その相手を頻度集計したものである。この表では、「司馬遼太郎」が 1 位となっているが、

①「頻度統計」ボタンをクリック

検索

絞込検索 頻度統計 基本統計 ダウンロード

梅棹忠夫著作目録\_200809 <検索結果 6170 件>

Next >

No.	レコー	文献番号	タイトル	著者名
1	10001	19341232	鷲峰山	梅棹忠夫
			三ヶ岳	梅棹忠夫
			雲岩山	梅棹忠夫
			芦見谷	梅棹忠夫

頻度統計

データベース ■ 梅棹忠夫著作目録\_200809

頻度統計項目1

頻度統計項目2

実行

種別コード  
種別  
種別表示  
言語  
著者名  
司会・聞き手  
掲載誌  
シリーズ名など  
出版年  
出版年月日  
出版者  
掲載箇所  
著作集収録巻  
リンク元  
リンク元との関係  
リンク先  
リンク先との関係  
配信リンク元  
配信リンク先

②集計する項目を選択

レコード種別表示	出現頻度	比率(%)	累積
	6113	100.0	
1 著	2447	40.03	40.03
2 対談	773	12.65	52.67
3 編集顧問	338	5.53	58.20
4 共同討議	332	5.43	63.63
5 監修	303	4.96	68.59
6 編集委員	284	4.65	73.24
7 座談会	281	4.60	77.83
8 インタビュー	234	3.83	81.66
9 講演	151	2.47	84.13
10 討論	134	2.19	86.32
11 共著	122	2.00	88.32
12 談	95	1.55	89.87
13 編	79	1.29	91.17
14 すいせん文	68	1.11	92.28
15 鼎談	44	0.72	93.00

③頻度集計結果

図 14 頻度集計の例

これでは不十分である。というのは、この「著作物とのかかわりあいかた」について「著作目録をつくる」には

記載事項については、こちらで枠をもうけたり、分類したりしないで、原典すなわち著作物そのものに記載されているところを尊重した。たとえば、「かかわりあいかた」の記載も、こちらで分類したのではない。原典に「鼎談」とあれば、あえて座談会に分類せずに、「鼎談」のままにしてある。共同討議とか討論会とかも、こちらの分類ではない。もとの著作物にそう記載されているのをそのまま採用したのである。

と記されており、さまざまに表記されている。したがって、「著作物とのかかわりあい方」が「対談、座談会、討論、鼎談、討論会、懇談会」のいずれかになっている

レコードを検索した結果の集計を行うと、トップ5は、

小松左京	61
川添 登	52
木村重信	39
加藤秀俊	39
石毛直道	38
司馬遼太郎	36

となる。

表6は同じように全レコードの出版社の頻度集計を行ったものである。この表からわかるように朝日新聞社が最も多く、講談社、中央公論社と続く。トップ5でほぼ全体の30%、トップ10で40%を超えている。

表5 対談相手と対談回数

対談相手	回数
司馬遼太郎	23
川添 登	18
小松左京	17
木村重信	13
矢野 暢	10
瀬戸内寂聴	8
樋口敬二	8
加藤周一	8
井上 靖	8
湯川秀樹	7
加藤秀俊	7
下川辺淳	7
上田正昭	7
黒川紀章	6
川勝平太	6
小山修三	6

表6 出版社の頻度集計

	出版社	点数	%	累積%
1	朝日新聞社	478	9.01	09.01
2	講談社	429	8.08	17.09
3	中央公論社	348	6.56	23.65
4	国立民族学博物館	245	4.62	28.26
5	財団法人千里文化財団	198	3.73	32.00
6	日本放送出版協会	133	2.51	34.50
7	平凡社	113	2.13	36.63
8	共同通信社配信	98	1.85	38.48
9	財団法人民族学振興会	85	1.60	40.08
10	「知的生産の技術」研究会	84	1.58	41.66
11	岩波書店	77	1.45	43.11
12	日本アイ・ビー・エム株式会社比叡会議事務局	76	1.43	44.54
13	筑摩書房	74	1.39	45.94
14	角川書店	72	1.36	47.30
15	財団法人21世紀ひょうご創造協会	54	1.02	48.31
16	文芸春秋	53	1.00	49.31
17	国際交流基金	51	0.96	50.27
18	小学館	48	0.90	51.18
19	河出書房	46	0.87	52.04
20	朝日放送	44	0.87	52.91

「出版年」で頻度集計をおこない、10年単位に区切ってグラフにしたものが図15である。60代が一番多くなっているのは、ちょうどこの年代に「梅棹忠夫著作集」が刊行され、多くの著作物がこの著作集に再録されたことが影響していると想定される。

70代に入っても年平均100の著作物が刊行されているのは、それだけ社会が梅棹忠

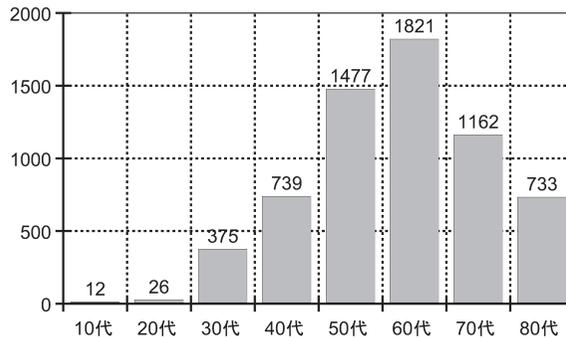


図 15 年代別出版著作物点数

表 7 掲載紙の頻度集計 (上位 10 誌)

掲載紙	回数
朝日新聞	343
国立民族学博物館編集『月刊みんぱく』	196
毎日新聞	97
神戸新聞	83
日本経済新聞	80
ちけんだいがく	63
産経新聞	56
千里眼	55
京都新聞	55
読売新聞	51

夫の著作物を必要としていることであり、80年代になっても年に100近い著作物があるのは驚きである。

ひとつの著作物が他の著作物に転載されたり、一部が抜粋されたりした場合、リンク情報として記録されている。このリンク情報について頻度集計を行ったところ、もっとも転載などが多いのは1969年に著された『知的生産の技術』である。47の著作物に転載されたり、抜粋されている。もっともあたらしいのは2003年の『国語表現 I』（教科書）に抜粋されている。次に多いのは、『中公新書 解説目録』の中の「良質かつ広領域の知識源」で31である。

表7は掲載誌の頻度集計である。もっとも多いのは『朝日新聞』で最初の掲載は、1951年7月10日の夕刊に載った「『大阪的』が意味するもの」というタイトルの座談会で、相手は桑原武夫、鶴見俊輔である。もっとも最近のものは、2007年10月20日に掲載されたインタビュー記事「『みんぱく』30年 役割は？——ナショナリズム超え世界に目開く場に——」である。

## 5. おわりに

ようやくデータベース作りも山を越えつつあるが、作業を開始して足かけ3年の長丁場となった。よいデータベースをつくるということは、ひとことで言えば、いかにして一つひとつのレコードの内容を正確なものにするかということにつきる。コンピュータを駆使したとしても、最終的には人の目による精査、修正作業が必要になってくる。したがって、時間がかかるのは当然であるが、もっとも大きな要因は「著作物」が多すぎることである。

私のもともとの専門は考古学だが、考古学関連のデータベースだけでなく、多種多様なデータベースをつくってきた。たとえば、いまは「ブックカバー（本のカバーで、図書館ではすべて廃棄される）」や「鶏に関する民芸品」などの画像データベースを手がけているが、データベースをつくっていて、つくづく思うことが二つある。

ひとつは「データベースは使うために作る、だから利用者が使って役立つように作らなければならない」ということである。当然といえば当然であるが、時として作成者以外の利用者にとって、とても使いづらい、あまり役立たないデータベースも見受けられる。

もうひとつは「データベース作りはひとりではなく、チームでやるべきだ」ということである。1次資料の収集、データシートへの記入、コンピュータへの入力といった作業は、ほとんど手作業の世界で、ひとりでやっていると、めげてしまいがちである。研究者仲間であれ、仕事仲間であれ、チームでお互いに励まし合い、時として叱咤激励しながら、作業を進めていくことがいいデータベースを作る要である。今回のデータベース作りも、梅棹資料室の三原喜久子女史、民博・情報管理施設の中川隆君などとチームとしてやってきたからこそ、やっところまでこれたのである。

## 文 献

及川昭文・藤沢桜子・洪政国・山元啓史

2007 「研究支援機能を強化したデータベース・システムの開発」『じんもんこん 2007 論文集』：229-236

及川昭文・山元啓史

2001 「Web 公開のためのデータベース・エンジニアリング」『情報処理学会研究報告』CH-49：49-56

2007 「研究者のためのマイ・データベース・システムの開発」『第13回公開シンポジウム「人文科学とデータベース」論文集』：pp. 25-34

鎌田聖子・及川昭文

2007 「幕末明治地勢地図境界データの作成」『じんもんこん 2007 論文集』：205-212

中川隆・高橋晴子・及川昭文

2007 「民博コスチュームデータベース（MCD）の過去・現在・未来」『じんもんこん 2007 論文集』：229-236