Osaka University Knowledge Archive

Title	原子核物理における学術情報の収集と検索
Author(s)	冨樫, 雅文
Citation	大阪大学大型計算機センターニュース. 1976, 22, p. 79-90
Version Type	VoR
URL	https://hdl.handle.net/11094/65318
rights	
Note	

Osaka University Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

Osaka University

原子核物理における
学術情報の収集と検索

北海道大学理学部　冨　樫　雅　文

0. はじめに

荷電粒子核反応実験のデータをファイル化し，研究者の利用に供するという計画が，昭和49年度からはじまった特定研究「広域大量情報の高次処理」の一環として立案・実行され，データの蓄積・検索システムNRDF－1が開発された*。ここでは主に核反応データの様相とそれに対応した情報システムの形成という面からの解説及びシステム開発の過程で折々に出会った問題や展望についてふれていくことにする。

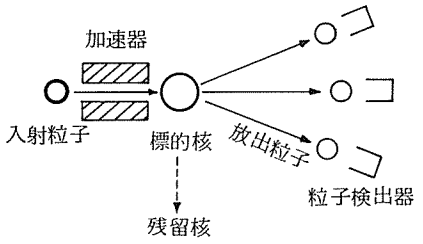
1. 計　　画

1.1　荷電粒子核反応データ

荷電粒子核反応データとはおおよそ次のようなものである。

図1に示すように，対象とする原子核に，加速器で加速された荷電粒子を衝突させて反応を起こし，その結果とび出してくる粒子の様子を観測する。こうして得られる種々のデータを，荷電粒子核反応データと呼ぶ。反応の全体は，化学反応式のようにa＋A→B＋bと表わしたり，あるいはA（a，b）Bのように表わしたりする。（ここでaは入射粒子，Aは標的核，Bは残留核，bは放出粒子を示す）。

図1　核反応実験



核反応を情報変換機構としてとらえるならば，入力データにあたるものは，入射粒子の種類，エネルギー，標的核の種類等であり，出力データのそれは，放出粒子の種類，角度分布，残留核の種類，エネルギー準位等である。また，反応自体も何種類かのタイプに分かれる。

ここで，可能な出力データの量を，a＋A→B＋bの核反応の場合についておおざっぱに

＊この研究は，北大・理　田中　一教授を研究代表者とする次の7人の研究者による共同研究である；田中　一（北大・理），池上栄胤（阪大・核物理研究センター），大沼　甫（東工大・理），村岡光男（阪大・理），河合光路（九大・理），阿部恭久（京大・基研），富樫雅文（北大・理）。

見積ってみよう。入射エネルギーが10点，標的核が300種，反応が40種類，放出粒子の角度分布を70点で測定，残留核のエネルギー準位を100点とする。測定量が反応の微分断面積$\sigma(\theta)$であるとして，その誤差もいっしょに記録するとすれば，データの量は；$10\times300\times40\times70\times100\times2=1.68\times10^9$個。1個のデータに4B(バイト)をあてるとすると，これは6720MB(メガバイト)に相当する情報量である。

1.2　NRDFの構想

これまでの学術情報，殊に研究論文の流通体系は学術雑誌という印刷媒体を介したものが主要なルートになっている。しかし，共通の印刷媒体を通した学術情報の流通は，ひとつにはそのchannelの太さの点で，今ひとつは研究者間のcommunicationの速さという点で，さらには情報のコストの面で多くの問題を含むと思われる。核反応実験データに即していえば，雑誌に発表される論文には，数値データではなくグラフの形で，しかもすべての実験データではなく，主要な部分だけが，のせられるというケースが多くなってきている。original dataの参照を希望する研究者に対して，雑誌はある意味でフィルターの役割を果たしているといえる。道路や河川の場合と同じく，学術情報の流通体系にも迂回路を切り開く必要を感じないわけにはいかない。核反応データファイルおよびそのシステムはまさにその迂回路としての役割りを担うべく構想されたものであった。

1.3　利　用　法

データファイルを利用する立場からその可能性をいくつか探ってみると，まず，データの評価，解釈を将来に委ねるということができる。またすでにファイルへの登録の段階で評価や解析がなされているものでも，再評価を受けるチャンスを保留できる。逆にいえば，過去のデータを詳細にわたって検討することを可能にする。さらに，人手によるデータ管理の弱さからくるデータの自然損失が防げる。これは次のようなケースからのアナロジーが有効であると思われる；東南アジアの農業生産は近年その収穫率を高めているが，一方で，収穫後の輸送・保存の面での遅れから全収穫量の30％にもおよぶ部分が最終的に失われてしまうという。このアナロジーはしかしデータの品質ないしは信頼性・有用性という面での検討に空白を残してはいる。次に積極的なdata compilationを可能にするという点があげられる。蓄積されたデータを種々のやりかたで編集することによって，いろいろな方向から全体を見通すことができる。これらの利用法を可能にするための機能をひとまとめにいうならば，それは，data freezing＋communication channel＋data compilationということになるだろう。

1.4　核反応データの様相

荷電粒子核反応データの特徴のひとつは，多様性と変化にあるといえる。核子の多体系としての原子核はそれに荷電粒子を衝突させる実験に対して種々の反応をひきおこす。この反応様式の多様性に対応して実験方法が多くのバリエーションをもち，さらに，放出粒子の観

測に際して種々の測定系が構成される。また測定量も多岐にわたる。一方，原子核理論は絶えず発展を続けていき，その結果，原子核を理解する上での概念や物理量が新しく現われ，また古いものがその意味するところを変えていくということが起こりうる。したがって獲得すべき実験データもまた時間とともに変化していく。こうしてみると核反応データは，対象の多様性と，対象の理解へ向けての我々自身の発展的変化という二重のひろがりの中にあるといってよい。

一方情報システムの形成という観点からこれをみるならば，情報の多様性と変化は必然的に情報要求の多様性と変化をもたらす。この結果，データファイルは一定程度の汎用性をもつことが要求される。しかしここで注意しなければならないことは，この「汎用性」はあくまで相対的であり，化学反応データでも，社会学データでもなく核反応のデータであるという大きな枠組の中でのひろがりをおおうためのものであるという点である。したがってその枠組に対応した個別最適化によるメリットを追求することはやはり必要である。

1.5　原　情　報

実験データを，計算機を利用して蓄積・検索することの利点は，その正確さや，情報処理の量と速度の大きさにある。しかしながら，よく指摘されるように，計算機による情報管理能力は，人間が従来計算機なしで行なってきた様々な情報管理のすべてをカバーするものではない。したがって例えば情報管理の正確さという面においても，それは計算機に与えられた論理の枠内におけるものであり，蓄積された情報が現実の世界によく対応するものであるかという意味での正確さは，人間と計算機の双方にまたがる情報系全体の問題である。

核反応データは現実の世界で発生した情報のひとつであるから，これを正確に記録し，計算機に入力する必要があるが，ここで情報の正確さというものに次の尺度をあてることができる；それは精度と確度である。情報の精度とは，ひとつの情報を識別するための分類の詳細度をあらわし，数値における有効桁数にあたる。したがってこれは通常の意味の情報量に直結している。一方情報の確度というのは original image の再現性の良さを示すもので，与えられた情報から現実世界における，発生地点での情報をどの位再現できるかを示している。いいかえれがノイズの少なさといえる。いかに詳細に伝えられた情報でも　それが単に入づてに聞く噂話であったのなら，現実の役には立たないのと同様に，精度のよい数値データファイルでも，そのデータの確度について考慮がなされていなければ　利用することができない。

データファイルの精度は，核反応データの場合，実験装置系全体の性能によって上限がおさえられてしまうが，その上限に近づくことは容易である。しかしながら確度については問題がある。

原情報として生データ（ぎりぎりの発生地点でのデータ）を登録することは確度の点でのぞましいが，そのためには実験の状況のごく詳細な記述を与えておかなければならない。と

いうのは，生データはその生成環境との密接な関連の下でのみ理解されるもので，その意味で情報としては環境に強く依存しているといえるからである。ところが，実験条件のうちには実験をした本人でなければ理解できない情報や，計算機に受理可能な format を与えることが困難な情報がでてくることが多い。そこで，このふたつ，つまり生データと生成環境を切り離す作業がなされる。それは，生データに，実験者が生成環境を考慮した上で前処理を施し，複雑な環境条件を参照しなくとも他の研究者がデータを利用できるようにするのである。このとき，生データに施された処理過程は，計算機を介した他人の利用という観点から計算機に受理可能な生成環境記述とともにデータファイルに登録される必要がある。

このようにしてはじめて計算機の内部論理に情報管理の正確さを期待することが意味をもつのである。

2. NRDF－1

2.1 概　　要

NRDF－1（Nuclear Reaction Data File version 1）は1974年の4月から10月までの間に仕様を検討し，その後コーディングを開始して1975年6月に一応の完成をみたものである。このシステムの特徴は次の点に要約される；

- 固定 format，固定 field へのcodeの記入方式ではなく，syntaxを設定し，可変な<文>を用いて，書誌事項や実験条件の記述ができる。この<文>の集合を記述部という。
- 数値データは文字の型式で蓄積される。また data field は空白を区切りとして可変長である。
- 検索時には，key 項目に対する inverted table を用いて，file全体のsequential search をさけている。さらに，検索 logic は bit map に対するブール演算によっている。

2.2 データ収集

データの収集は次のようにして行なわれる；まず研究者自身が，各自の実験データを整理して，定められた文法に従って書誌事項や実験条件を書き下し，対応する数値データを添える。これらのデータシートを集めて，パンチャーがこれをカードにパンチしてシステムに投入する。システムは，この入力データについて種々のチェックを行い，それにパスしたものをファイルに登録し，同時に記述部の中のkey項目から inverted table への情報を抽出する。key項目は変更可能である。

入力データの例を図2に示す。

図2　入力データの例

```
IDN=1;
RFC=74NOJ1; ATH=(T.NOJIRI, H.TOYAMA); REF=JUPSA; VLP=37(74)100;
RTY=IAR; RCT=50TI(P, P)50TI;
PRJ=P; EIN=(3.588, 3.606);DEIN=0.003; PLP=0; ACC=VDG; INT=0.5;
BAC=C; CHM=T1.02;
TGT=50TI; PLT=0; THK=0.01; DTHK=(+0.003, -0.003); ENR=67.7; IMP=
(D, C);
DTF=NAI+SIL;DET=(NAI(DBS=G, PSI=90), SIL(OBS=P, ERS=50, THTC=45
(15)165, DTHTC=1.0));
PHQ=(SIG/SR, EXF, FLP, CSP, CPT, PWD);
ANL=RES;
INE=3.588; EXE=1.55;
MEQ=(THTC, SIG/SR, DSIG/SR);
NDTA;
     45.0     8.0     0.2     60.0     9.6     0.3
     75.0    11.0     0.4     90.0    11.6     0.4
    105.0    10.6     0.4    120.0     9.0     0.3
    135.0     6.3     0.2    150.0     4.3     0.2
    160.0     2.9     0.1
DATA END;
IDN=2 1;
INE=3.588; EXE=1.55;
MEQ=(THTC, FLP, DFLP);
NDTA;
     45.0    28.5     4.0     60.0    27.3     3.0
     75.0    29.5     4.0     90.0    29.5     3.5
DATA END;
END;
```

2.3　記述部とデータ部

記述部の<文>の syntaxは，実験データのひとまとまりに対する属性－属性値の組の複合として表現される。すなわち；

X＝A……<項目>＝<値>；

を基本型とし，これに

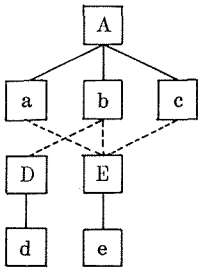
(X，Y，Z，…)＝A；

X＝(A，B，C，…)；

X＝A(Y＝B)；

図3.a

A=(a, b(D=d), c)
(E=e);



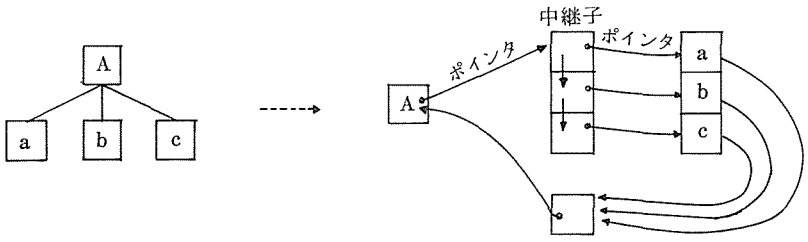
X＝A（Y＝B，Z＝C，…）；

X：＝A；

という変型規則をくりかえし適用して得られるものを<文>としている。例えば「測定データは，角度と微分断面積で，微分断面積の単位は mb/sr である」という内容は，MEQ＝（THT, SIG/SR（UNIT＝MB / SR））；と表わされる。このような<文>は項目と値の network 的な結合を持ち得るが（図3.a），NRDF－1ではこの関係を pointerを使って明示的に表現している（図3.b）。

図3．b

A＝（a，b，c）；



数値データの集合（データ部）においては，いくつかの異なる種類のデータの組がくりかえしあらわれるという以外の制限はなく，数値の型も，利用者とシステムの間に合意が存在すれば何でもよい。これはデータ部の内容が文字の形でそのまま蓄積されていることによる。

2.4　データ検索

データの検索は基本的には，記述部の中の key 項目が満たすべき条件を示すという形で指示される。また，ふたつ以上の key 項目に対する条件指定は 論理演算子で結合されて「条件Aかつ条件B」というような複合条件が提示される。こうして示された条件を満たすデータセットは，各 key 項目についての inverted table を調べて得られた適合データセットへのアドレスの集合または各集合間の集合演算の結果をもとに引き出される。いったん検索されたデータ集合は，それへのアドレスの集合の形で自動的に保持され，指定された名前をつけておく。

検索操作のためには，検索コマンド，消去コマンド，出力コマンド，終了コマンドの4つのコマンドを自由な順序でシステムに投入して，段階的に目的データへ接近する。

検索コマンドは，検索結果につけるべき名前のあとに各 key 項目に対する複合条件（論理式）をつづけた質問式の形をしている。このコマンドを受けて，システムは検索を実行し，適合するデータセットの数を表示する（図4.①，②）。

消去コマンドは，自動的に保持されている過去の検索結果のうち不要になったものをシステムのリストから削除することを，当該質問式名を与えて指示する（図4.③）。

図 4

```
①→Q1 :=( INE>=10 )*( INE<=100 ); ……10≦INE≦100の意
②→HITCNT= 46
   Q2 :=( EXE>=2.5 )*( EXE<=5.5 );……2.5≦EXE≦5.5 の意
   HITCNT= 10
   Q3=Q1 *Q2 ;
   HITCNT= 8
③→ERASE Q1 ;
   Q3=Q1 *Q2 ;                          Q1は引用できない
                                          ↓
   *** SYNTAX ERROR(USER)***            エラーメッセージ
   Q3 :=Q3 *( PRJ=P )*( TGT=13C );
   HITCNT= 1
                        部分指示 ; ALL又はINTRO
④→OUTPUT Q3, ALL, ATH, RCT, DTF, PHQ, ANL;
          出力の対象default値は直前の結果   出力する項目default値は全ての項目
   *** DATA-SET NO. 1 ***
      ATH=(H. OHNUMA, K.KOYAMA, S.KUBONO, K. I. KUBO, H.AMAKAWA);
      RCT=13C(P, P)13C;
      DTF=MAG+PPC*PLS;
      PHQ=( SIG/STR, EXE, TRL, EMT );
結果  ANL=(OPM, DWB, TWSTP, CCB, COL, SHL);
    NDTA;
                 THT        SIG/STR      DSIG/STR
                  6.5         .27           .06
                 19.2         .135          .02
                 40.7         .295          .02
    DATA END;
⑤→END;
```

出力コマンドは，すでに検索され，アドレスの集合の形で保持されている結果を，print outすることを指示する。この時，どの結果のどの部分を実際に出力するのかをパラメータで示すことができる。またこれらのパラメータにはdefault 値が設定されている（図4.④）。終了コマンドは検索操作の終了を指示する（図4.⑤）。

3. 情 報 構 造

3.1 単 位 情 報

核反応データを収集する際，実験者はひとつの実験で得られたデータを format 化して提供する。ここで＜ひとつの実験＞とは，実験者が得ようとする，あるいは光をあてようとする，対象の一側面についての知見を引き出すために構成された一連の実験を指すのがふつうである。これら一連の実験は，少しずつ実験条件を変えて行なわれ，個々の条件の下で得られた実験データをまとめた形で書き下されたものが，データファイルへの登録の一単位とする。このように，収集時のデータの基本単位は，実験者の意図や主観を反映した可変の大きさを持つ。ところが，データファイル内の情報を引き出そうとする利用者は，データ登録時の実験者の意図や主観とはほとんど独立に　各自の検索操作を行なうのがふつうである。つまり利用者は，各自のデータ群への視点および分解能というものに規定されながら一連の検索コマンドによって要求するデータの像を表現しているのである。したがって，データ提供者が登録時につけた各データ間の差異を，利用者が無視している場合もあれば，データ提供者の考えなかった区別のしかたを利用者が前提としている場合もありうる。これらデータ提供者とデータ利用者の間の，単位情報に関する規定の差異は，前者が自然界を対象とし，後者がデータファイルを対象としてはいるが，ともに情報獲得様式の，各研究者間における差異に強く関係していると思われる。データファイル内における単位情報の規定は，したがって両者の差異を吸収し，かつ相互の橋わたしを可能とするだけの客観性をもたなければならない。

情報の収容体制として，それではどのような枠組を準備すれば，異なる主観にもとづく情報の授受を媒介しうるのであろうか。

そこで，ひとつのてがかりとして「原子情報」という概念を用いてデータファイルの構成を考えてみることにする。原子情報というのは，それ以上には分解できない情報の単位であって，その意味するものについて，人間を含めたひとつの情報系全体で一意性をもつものとする。一個の測定値は，この意味で原子情報といえる。しかし一個の測定値の数値として以上の意味づけは，パラメータ，実験条件，書誌事項等を通してみてはじめて考えられる。すなわち，個々の測定値すべてに　何らかの方法でこれらの，いわば環境情報を結びつけておくことがデータファイルでは必要となる。また，いくつかの測定値は環境情報の一部を共有するという形で連合を形成する。したがって一個の測定値は，ファイル内におけるその「位置」を確定するために，環境情報や，環境情報を通した同類へのつながりをもつことになる。

このような考え方をすることによって，情報の授受における主観の差異は，原子情報同士を結びつける，あるいは関連づける方法の差異と　とらえることができる。

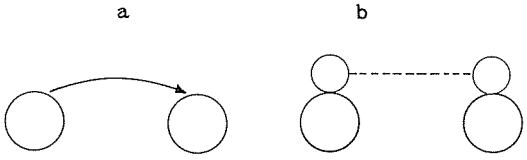
3.2　連　　合

原子情報とは限らず，任意の情報要素同士を関連づけるには　種々の方法が考えられるが，それはすべて次の２種類の方法の変形と考えることができる；　ひとつはアドレスによる連合の形式であり，いまひとつは値（内容）による連合の形式である。アドレスによる連合の形式とは，ひとつの情報要素に，関連づけるべき他の情報要素のアドレスを付加する方法で，いわば，矢印をつけてふたつの情報要素をつなぐ操作である（図5.a ）。これは explicit で，かつ方向性をもった関連を表わしている。

一方，値（内容）による連合の形成は，ある情報要素の全体またはその一部分と，他の情報要素の全体または一部分が同じであるという事実をもって関連づける方法である。これはふたつの情報要素に同じ旗を立てて関連していることを示す方法といえる（図5.b ）。さらにこれは　implicit で方向性をもたない関連づけの手法である。

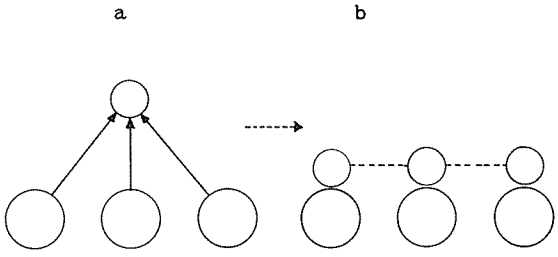
いくつかの情報要素が　ともにあるひとつの情報を共有するという事情は，アドレスによる連合の形成によれば図6.a のようになるが，値による方法では　図6.b のように個々の情報要素に同じ情報を付加させなければならない。一般に，アドレスによる方法は，関連づけられた情報というものをコンパクトに表現できる。したがって，値による方法での表現に変換するためには余分な情報の付加がさけられない。これは強く結合しあった多体系を分解するための解離エネルギーのようなものと考えることができる。

図５



アドレスによる連合　　値による連合

図６



3.3　検　　索

情報検索というのは，利用者がファイルの中を「うろついて」めぼしい情報をひろいあげる操作であると説明することができる。このとき，ファイルの中の「道標」にしたがって，特定の小径沿いの情報をひろい集めるというやり方が　構造検索とよばれるものであり，てあたりしだいひろいあげた情報の内容をみて，取捨選択をしながら集めるのが内容検索である。すなわち，構造にに沿った検索は，アドレスによって形成された連合に依拠して，アドレスをたどりながらそのひとかたまりをとり出すのに対して，内容による検索は，値によって，implicit に形成されている連合のうちのひとつを，当の値を示すことによってとり出すとい

う操作である。

値によってimplicit に形成されている連合というものの枠をいま少し拡げてみると，ひとつの情報要素の内容をA，他の情報要素の内容をBとして，A＝Bというのが連合を形成するための条件であったが，ここで，f，gをある関数としてf（A）＝g（B）である時にふたつの要素が関連しているとみなすこととすれば，f，gのとりかたによって関連したり，しなかったりという状況が現われてくる。

ここで，ファイルの中での情報要素の関連のしかたをあらためて構造とよぶことにすればアドレスによる連合は，ファイルの客観的な構造を与え，誰にとっても「存在する」構造としてとらえることができるのに対し，一方，値によって形成された連合は，さきほどの広い意味にとるならば，各要素をどのように関連づけるのかという，ファイルの外側からの主観を反映した，いわば「見える」構造として前者とはちがう可能性をもつことになる。

そして情報検索は，これらの構造にもとづいて情報をとり出すという意味で，狭い意味での構造検索と内容検索の両者を統一的にとらえることができる。

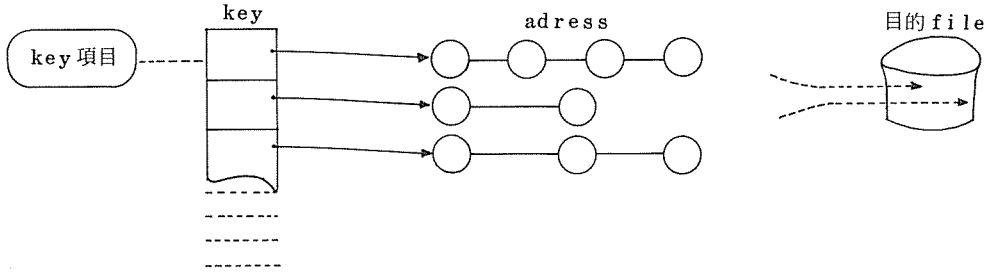
3.4　inverted file

inverted fileとは，目的fileの中の各情報要素がいくつかのkey項目についての情報を含んでいる場合，key項目ごとにその内容（key）と，そのkeyを含んでいる情報要素へのアドレスあるいはアドレス群を対にして保持しているfileで，検索時にkeyを指定することによって情報のとり出しを行なう場合に利用される（図7）。inverted fileを利用することによって，目的fileの中を直接探しまわるかわりに，inverted fileの中で，単純な構造に沿って指定されたkeyとそれに続くアドレス群を見出すことで，実質的な検索が行なわれることになり，検索の速度を上げることが期待できる。

inverted file の中で，ひとつのkeyにつづくアドレス群は「〜のkey項目が〜の値をもっている」という単純条件指定に適合する情報要素を各々指しており，これらの情報要素は，この条件を通しての同類とみなしてもよい。したがってこれは，情報要素に対するimplicitな連合形成のひとつの方法を示していると思われる。

このような見方をとるならば，explicitな連合形成にもとづく客観的構造をもったfileの

図　7



上に，ある種のinverted fileを重ねるという操作によって，各利用者が独自の主観を反映させたいくつものfile構造定義を独立に与える可能性が開けてくる。
この利用者が個々に用意するinverted fileは例えば次のようなものが考えられる；（すべての項目について展開したfull inverted fileがあればそれを利用して）異なるkey項目にまたがる複合条件指定に適合するデータのアドレスと，指定した条件に利用者がつけた識別名を対応づけて，種々の条件指定について対応表を作成する。ひとつの識別名から引きだされるデータ群は，利用者が複合条件指定という言葉で表現した，あるひとまとまりの情報を指しており，この意味でfile全体の情報に対して，利用者の主観を反映した分類法に基くものである。

3.5 code system

code systemはfile内における語の体系である。codeは何らかの意味を担う媒体として機能するのであるが，その担い方を2通りに大別することができる。ひとつはsimple codeとでも呼ぶべきもので，単一の概念，実体に対応づけられていて，ひとつのcodeをそれ以上分解してみても意味のないものである。もうひとつはcomplex codeで，これは，いくつかのsimple codeをsub codeとして用い，それをsub code間の関係構造に対応した方法で，ある文字列の中にたたみ込むことによってcodeを形成するものである。これによってcodeを複合概念や，構造の見える実体に対応づけることが可能になる。
fileへのデータ入力時の情報表現は，これらのcode systemとcodeを布置する枠組としてのsyntaxの両方に依存してなされるのであるが，表現すべき情報を主にcodeに担わせるのか，あるいは主にsyntaxに担わせるのかによって，code systemの大きさや各codeの担う情報の量がかわってくる； 高度のsyntaxの下では，基本的な概念や実体に対応するcodeを用意するだけで複雑な状況を記述できるのに対し，simpleなsyntaxの下では，豊富なcodeの用意がなければ，これと同等の表現力をもつに至らない。
また，安定に存在しうる概念は，たとえそれが複合概念であっても，syntaxに依存するよりは，全面的なcode化による表現がとられやすいのは，機能のblack box化と事情が似ている。
一方，codeとそれが担うべき意味との対応関係はpublic fileを前提とするかぎり客観性をもたなければならないが，利用者の主観に基くpublic fileの，外側からの再編成を利用者個人のprivate file内で行なう場合は，独自のcode systemを定めることが考えられてよいが，この時はpublic code systemとprivate code systemとの間の相互のmappingが自動的に行なわれる必要がある。

このprivate code systemの考え方は，各利用者が独自の言葉によってfile systemとcommunicateすることの可能性を開くことになる； 検索時のmacro commandの利用や質問ライブラリー，入力時のmacro statementの導入などはその例といえる。
また，このようなことを可能にするためには，programがcode systemに対する独立

性を保つことが必要になるが，これは， code system とその意味するものとの間，および code system 間の変換のための辞書を装備することを意味する。さらにこの辞書は，code system の時間的な変化を追えることが望ましい。

4. 情報検索……まとめ

▽ 情報検索とは，部分的情報から全体像を引き出す操作である。いいかえれば，これは関係にもとづく連想である。

▽ fileを構成するためには

1 原子情報

2 関 係

3 code

の三者を明らかにすることが必要である。

▽ 情報表現における，情報密度と操作性の間の相反性と相補性は情報系に対する制限であると同時に，臨機応変のための主要な自由度でもある。

これらの事項を原則として，NRDF－1は今，NRDF－2へと移り変わりつつある。