

p45

LECTURE 2.1, READING 1

Organizing Knowledge in Libraries

An Introduction to Information
Retrieval

C. D. Needham

Fellow of the Library Association

Principal Lecturer, School of Librarianship,
Polytechnic of North London

Second Revised Edition

Reprinted by permission of the publisher from: Needham, Christopher D. *Organizing knowledge in libraries: An introduction to information retrieval*, Seminar Press, Ltd., 1971, 2nd ed.

Grafton Library Science Series

 SEMINAR PRESS

FOSKETT, A. C. *The subject approach to information*. London, Bingley, 1969. A valuable new textbook with an integrated, systematic approach.

MILLS, J. *A modern outline of library classification*. London, Chapman and Hall, 1960.

There are also valuable chapters in:

ASHWORTH, W. ed. *Handbook of special librarianship and information work*. London, Aslib, 1967.

Other readings will be indicated in the chapters that follow.

Chapter 7

Classification

The *Shorter Oxford English Dictionary* defines a class as 'a number of individuals (persons or things) possessing common attributes, and grouped together under a general or "class" name; a kind, sort, division'. Classification, 'the action of classifying', is a basic process in indexing – whether the final product of the process be a systematic sequence of classes known as a classification scheme, a sequence of documents systematically ordered into subject classes on the shelves (or a similar sequence of entries in a catalogue), or an alphabetical sequence of classes (as in an alphabetical subject catalogue or a list of subject headings). In creating any of these tools we are naming classes of things – documents and the subjects contained within them – according to the presence or absence of attributes. Further, by the arrangement of classes and through systematic references etc., we are demonstrating the relationships that exist between them. Classification is also a fundamental operation in reference work, and Coates has written an interesting chapter on this use of classification in his book *Subject catalogues*.

Although in this and the next chapter we shall be dealing with *schemes* of classification, the basic theoretical points concerning the analysis and categorization of subjects underlie all forms of subject indexing.

The Making of a Classification Scheme for a Particular Subject

A classification scheme is simply defined as 'an orderly arrangement of terms or classes'. Given a notation, a classification scheme can be used to arrange documents or entries in the catalogue and is, as we have seen, the only way of achieving a preferred order of subjects.

Examine the following titles: (i) *The literature of the Eighteenth*

century, (ii) *A history of English literature*, (iii) *French drama*, (iv) *Russian influence in the English short story*. A good scheme for Literature should be able to accommodate each of the subjects denoted by these titles with a distinct place set aside for that subject only – otherwise specific entry will not be possible* – and setting each in proximity to related subjects.

Of the making of such subjects there is no end. Hundreds pour off the presses each week. How then can a scheme be made to accommodate them whenever they occur? Clearly *enumeration* of all subjects is impossible.

STAGE ONE: ANALYSIS

As all substances are either simple chemical elements or compounds† of such elements, so subjects in documents are either simple concepts or compounds of such concepts. The subjects expressed in the titles listed above are compounds of: (i) Literature, Eighteenth-century, (ii) History, English, Literature, (iii) French, Drama, (iv) Russian, English, Short story. Although the number of such simple concepts increases with the increase in knowledge, the rate of increase is slight compared with the rate at which new compounds are created. It follows that if a classification scheme is composed of simple concepts only, and if some means of combining them in a preferred order is devised, then the scheme will be capable of accommodating the vast majority of new subjects as they arise. For example, if the notational symbol L represents Literature, b Novel and e English, then the combination 'Leb' will stand for the subject 'The English novel'. *Synthesis* is a term used to refer to the creation of compound subjects by the combining of simple elements.

* A classification scheme should allow specific entry even if the classifier using the scheme prefers broad classification.

† In this chapter and elsewhere in the book I have not differentiated between 'compound' and 'complex': both terms are used to connote subjects comprising more than one simple element. However, other writings on classification make the following distinction.

Compound: a subject comprising a relationship between two or more foci within a class, e.g. the use of visual aids in teaching mathematics.

Complex: a subject comprising a relationship between two or more foci from different classes, e.g. mathematics for engineers, psychology of politics. *Complex relationships* are sometimes referred to as *phase relations*.

Number-building refers specifically to the notational aspects of this process.

The first job of a maker of a subject classification scheme will be to examine a representative sample of documents in that subject, covering all topics within the subject and all levels – such as could be found in a bibliography of the subject – listing concepts as they arise. In this way all simple concepts will be discovered and they will be concepts actually arising in documents – they will have *literary warrant*. Concepts in this unorganized state are often called *isolates*.

Vickery has advocated the use of organized textbooks in the subject to discern the general structure of the subject. He also shows how glossaries of the subject can be of assistance, particularly in connection with Stage Two – the creation of facets.¹

STAGE TWO: CREATION OF FACETS

The isolates now need grouping so that those which are related are proximate. Examine the following isolates in Education: (i) Primary, (ii) Secondary, (iii) Further, (iv) Visual aids, (v) Tutorials, (vi) Geography, (vii) Mathematics. (i)–(iii) are obviously related, so are (iv)–(v), and (vi)–(vii). There are three basic groups here: (a) Education according to age, (i)–(iii); (b) Teaching methods, (iv)–(v); (c) Subjects taught, (vi)–(vii). The relationships arise because in each group the concepts have one characteristic in common (Age or Method or Subject).

A *characteristic* is an attribute by which concepts are grouped or subjects divided. Traditionally we talk of *characteristics of division*, because, following logical analogy, the process of division predominated. The isolates listed above could have been arrived at had we divided the class Education by the characteristics Age, Method, and Subject. The grouping process has this advantage: the isolates have literary warrant. However, division might help to yield isolates which by chance have remained undiscovered in the survey of the literature.

Other examples:

<i>Subject</i>	<i>Characteristic</i>	<i>Examples of isolates</i>
Literature	Period	Eighteenth century, Nineteenth century

<i>Subject</i>	<i>Characteristic Form</i>	<i>Examples of isolates</i>
	<i>Author</i>	Ode, Ballad, Drama, Novel
	<i>Language</i>	Shelley, Blake, Hugo English, German
Occupational safety and health (CRS)*	<i>Industries</i>	Agriculture, Food, Textiles
	<i>Source of hazard</i>	Fires, Tools, Electricity
	<i>Industrial disease</i>	Asthma, Eczema

The sum total of isolates formed by the division of a subject by one characteristic of division is called a *facet*. In the above examples we may refer to the Form facet, the Period facet, the Source of hazard facet, and so on. The isolates within a facet are called *foci* to distinguish them from the unorganized concepts called isolates. This process of analysis is called *facet analysis* and a scheme produced after such analysis a *faceted classification*.†

Characteristics of division should be *mutually exclusive*. Thus the Period facet in Literature should not include foci belonging to the Form facet, nor the Hazard facet in Occupational safety and health include foci belonging to the Industry facet. This is very important. Where such impurities occur there is liable to be *cross classification*: the situation where documents on the same subject are to be found classified at different places in the scheme.

The enumeration of foci should be *exhaustive* according to the literary warrant. That is, all foci should be listed and there should be room for expansion as new foci arise, e.g. Esperanto in the Language facet for Literature. The latter is a notational problem that will be discussed below.

It has been discovered that there are recurrent *categories of facets* in many subjects. For example, a Materials facet appears in Engineering, Building, Architecture, Painting, and so on;

* A scheme for this subject devised by D. J. Foskett and others, was published by International Occupational Safety and Health Information Centre, Geneva, 1960, and is referred to in this book as CRS.

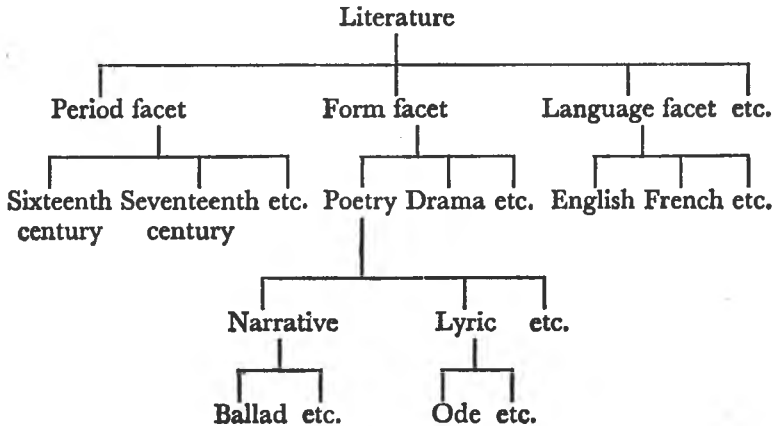
† Many of the terms and concepts in this chapter – and indeed throughout this book – are based on the fundamental research of S. R. Ranganathan.

an Operations facet appears in these subjects also; an Age facet appears in many subjects in the Social sciences, e.g. Education, Social Services, Labour, Economics. The foci within these categories may vary according to subject, of course, as the following table illustrates:

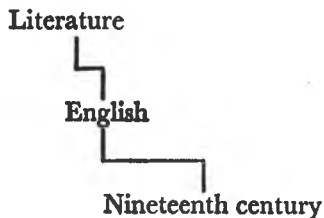
<i>Category of facet</i>	<i>Subject</i>	<i>Examples of foci</i>
Operations	Librarianship	Cataloguing, Reference work
	Building	Site preparation, Bricklaying
	Medicine	Surgery, Nursing
Materials	Engineering	Steel, Plastics, Copper
	Textiles	Wool, Cotton, Nylon
	Painting	Water paints, Oils, Crayon

This fact – that categories of facet recur – is of great assistance to the maker of a scheme as an aid to the initial definition of relevant characteristics and, consequently, facets. By asking what is the Materials facet, the Operations facet, and so on, the grouping of isolates is facilitated. Ranganathan has suggested that there are, basically, five categories which he calls Personality, Matter, Energy, Space, and Time (PMEST). Other fundamental characteristics have been suggested. Vickery,² with particular reference to scientific literature, has postulated the following: Thing (product) – Part – Constituent – Property – Measure – Patient – Process/Action/Operation – Agent (Space – Time). Differences between sets of fundamental categories arise from differences in level of generality and differences in the purposes for which they have been created. The important thing to grasp here is the principle of facet analysis and the recurrence of categories of facet.

Once the facets are established, the foci within them may need further grouping into subordinate facets or *sub-facets* according to subordinate characteristics applicable only to the facet under consideration. Thus in the CIS scheme, the foci in the Pathology facet (M – N) are grouped according to the location of the disease (diseases of the eye, respiratory system, nervous system, and so on); in the Safety and health engineering facet (S) there are sub-facets for fire protection, explosive prevention, radiation and electrical protection, and so on, and within each of these a cluster of related foci. Again, in the class Literature, foci in the Form facet may be grouped as follows:

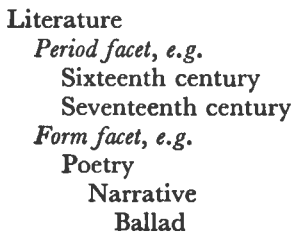


Care must be taken to see that elements from other facets are not introduced, for, as we have seen, characteristics of division should be mutually exclusive. For example, it would be wrong to sub-divide the Language facet of Literature by divisions from the Period facet, e.g.:



because this results in a compound: Nineteenth-century English literature. It is better to form such compounds by synthesis rather than by enumeration.

The groups of foci could be listed in a classification schedule as follows:



Lyric
 Ode
 Drama
Language facet, e.g.
 English
 French

Such a series of ranked classes and subclasses reflecting various steps of division through the application of a variety of characteristics, as shown here, is referred to as a *hierarchy*. Within the hierarchy foci of equal status, i.e. those formed by dividing a class by *one* characteristic, are called *coordinate foci* or *coordinate classes* or *arrays*. Thus (a) Poetry and Drama are coordinate, as are (b) Narrative and Lyric. This distinguishes their relationship from the *subordinate* relationship existing between, respectively, Poetry, Narrative, and Ballad, and Poetry, Lyric, and Ode, for here, in each case, the second and third foci are produced after further steps of division. Hierarchy can only be properly demonstrated by a two-dimensional 'family tree' layout; classification schedules with a *linear* or one-dimensional sequence demonstrate it with only partial success by indentation and typography, and on the shelves and in catalogues the order is even less apparent.

Each stage in the creation of sub-facets is called a *step of division*.^{*} It is important that relevant steps of division (i.e. those for which there is a literary warrant, such as Narrative and Lyric in the above example) are not missed. As Mills has pointed out in his *Modern outline of library classification*, the omission of a division for Prose in DC Class 800 has led to the lack of accommodation for documents on Prose.

STAGE THREE: ARRANGEMENT OF FOCI

So far we have a series of facets and sub-facets with coordinate foci enumerated in each. The next task is to arrange the foci in some order. Ranganathan³ has listed the following principles for the arrangement of foci: evolutionary, chronological, geographical, increasing complexity, alphabetical, canonical

^{*} This term is also used, as we shall see, to refer to steps in the formation by synthesis of a compound subject, e.g. Literature - English - Nineteenth century, see p. 183.

and consistent. What order is chosen for a particular group of foci will depend on the subject: thus, arrangement of Place names according to geographical proximity is superior to the arbitrary alphabetical arrangement; some foci – Biological species, for example – are best arranged according to the evolutionary principle; chronological arrangement is clearly called for whenever the development or sequence of a subject is to be stressed – in Printing processes or Agricultural operations for example. Arrangement of coordinate foci has been called *order in array*.

STAGE FOUR: DECIDING ON THE COMBINATION ORDER OF FACETS

Remember that the scheme will consist of simple concepts which will have to be combined or synthesized to accommodate the complex subjects found in documents. A combination order (or citation order as it is sometimes called) must therefore be laid down, i.e. *a combination order of facets*. Without this documents on the same subject may be classified at different places. For example a person classifying a document called *Sixteenth-century English literature* may, unless he is given an indication of combination order, place it in the Period facet or the Language facet. Again, a person classifying a document called *English language teaching in secondary schools* must know whether to place the document in the Subject facet or the Grade facet; if the notational symbol for English language is ϵb and the symbol for Secondary schools is gd , shall the final symbol be $\epsilon b g d$ or $g d \epsilon b$? In the first place the combination order is Subject – Grade, in the second Grade – Subject. *This is a very important point, for it determines the extent of collocation*. If the order is Subject – Grade, then the teaching of a particular subject will be collocated whatever the type of school involved, but a person wanting all material on, say, Secondary schools will have to search not only under the symbol for Secondary schools, where he will find documents on Secondary schools in general only, but also under the various *subjects* taught in Secondary schools.

Classification inevitably separates as it collocates because a document can go in only one place on the shelves, no matter how complex its subject; and though in theory any number of entries

can be made in the catalogue, there are economic limits, and few cataloguers would make entries for books on the teaching of subjects in certain types of schools under both subject *and* school. And the more complex the subject the greater the separation; if the combination order for Literature is Language – Form – Period, as in DC, then all the Literature in a particular Language is collocated, but Forms of literature (Drama, Poetry, etc.) are scattered through those Languages, and Periods of literature are further scattered through the Forms – and thus a person who wants all documents on the Literature of a certain period, regardless of Language or Form, will have a long search. Related subjects scattered by the classification in this way are called *distributed relatives*, a term which we shall use frequently when discussing the subject index to the classified catalogue (Chapter 9). The index to the scheme itself – particularly if enumerative – should gather together any distributed relatives.

The best combination order for a subject will depend to some extent on *the needs of users*, and their approaches to the subject should be carefully analysed. This is impossible if the library caters for a wide variety of readers, but even here there is, to use Bliss's phrase, an *educational and scientific consensus of opinion*. Thus in the field of Literature, students normally confine their studies to the Literature of a particular Language, rather than Period or Form, hence Language should take precedence in the combination order for compounds; within a particular Language they normally study Period rather than Form (see the style of GCE examinations) and therefore the combination order should be Language – Period – Form.

There have been several attempts to express a standard citation order in terms of fundamental categories. The categories of Ranganathan and Vickery have already been noted (p. 113), and the order in which they are given (PMEST, Thing – Part, etc.) is, in each case, the one regarded by its author as being most appropriate for general application. However, Vickery makes the point that 'such a set can only serve as a guide, a reminder of a pattern that has been helpful in other situations'.

STAGE FIVE: ARRANGEMENT OF FACETS IN THE SCHEME:
SCHEDULE ORDER

The final decision before the notation is added is the relative order of the facets *in the scheme itself* (i.e. the facets as wholes – not the foci within them as already discussed above). For example, shall the layout be:

(a) Education <i>Grade facet</i> Primary Secondary Higher etc. <i>Subject taught facet</i> Science Literature History etc.	OR	(b) Education <i>Subject taught facet</i> Science Literature History etc. <i>Grade facet</i> Primary Secondary Higher etc.
--	----	--

This is relatively unimportant compared with the question of combination order. However it is traditional for 'general' to precede 'special', e.g. Wheat and Harvesting in general would precede the compound Harvesting of wheat; Secondary schools and The teaching of mathematics would file before The teaching of mathematics in secondary schools. If we wish to achieve this order of *increasing speciality* (and some recent schemes have abandoned it), then *the order of facets in the scheme must be the reverse of the combination order*. For example, if the combination order in Education is Grade – Subject, then the schedule order must be Subject – Grade; in this way the most significant facet (the one which collects most compounds – and is therefore more specialized) will file after the less significant (the one where there are fewer compounds and which is therefore more general). If however an order of increasing speciality is not considered important the order of facets in the schedule should be simply that of the combination order (as in CIS).⁴

STAGE SIX: NOTATION

The last stage apart from the index, is to add a notation to mechanize the arrangement. This subject will be treated below.

STAGE SEVEN: MAKING THE INDEX

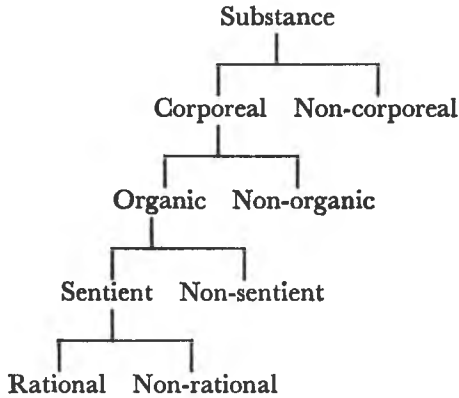
An index is necessary: (a) to indicate the location in the schedules of a sought term, and (b) to collocate distributed relatives (see page 117). Indexes which meet the second objective have been called *relative indexes*. We shall investigate the purpose and construction of a relative index in Chapter 9. Suffice to say here that *the index should not repeat the order of the schedules* as this adds greatly to the length of the index, is uneconomic, and is likely to be unsystematic. For example, the index to the partial schedule given above (list (a)) should read:

	<i>Class number</i>
Education	...
Higher education	...
History: Teaching: Education	...
Literature: Teaching: Education	...
Primary education	...
Science: Teaching: Education	...
Secondary education	...
NOT Education	...
Education: Higher	...
Education: Primary	...
Education: Secondary	...
Education: Teaching: History	...
etc.	

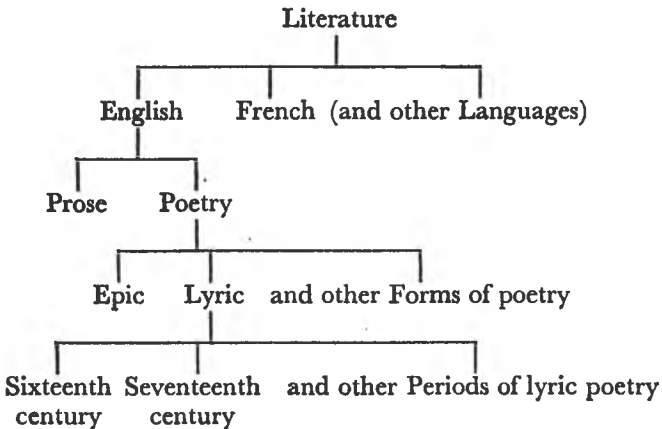
Traditional Classification Theory

The method outlined here is today recognized as being superior to the traditional methods of library classification. It has been used in several special classification schemes, e.g. CIS and the BNB *Music classification* used in the *British catalogue of music*. The traditional methods were based on a close analogy with logical classifications where the classification structure is a hierarchy formed by the *successive* application of characteristics of division thus enumerating compounds. This can be diagrammatically expressed on the next page.

Its main purpose is the *definition* of terms and categories, and it is concerned with relations between *Genus* (e.g. Corporeal substance) and *Species* (e.g. Organic substance) – the *thing* and its *kinds*. Strictly *specificity* should be used only to refer to this



relationship between a thing and its kinds (but see page 97). In library classification we are not concerned solely with the Genus-species relationship (Harvesting is not a *kind* of Agriculture, Electricity is not a *kind* of Physics), nor is our aim definition. The limitations of the traditional hierarchical structure for our purposes are easily demonstrated by the following diagram where the subject Literature is divided successively by various characteristics by analogy with the above pattern.



Similar divisions would be found under each of the Languages. Because all Forms can only be expressed as sub-divisions of Languages, documents *simply* on Poetry, Prose, Epic, etc.,

cannot be accommodated; documents on Periods of literature cannot be accommodated either – except where the document also concerns a particular Form and Language, e.g. Seventeenth-century English lyrics. Moreover, *just as simple subjects cannot be accommodated, neither can compounds other than those enumerated in the hierarchy* – thus the scheme above could not accommodate Seventeenth-century English literature. This is what invariably happens in enumerative classification schemes.⁵

Clearly the method outlined earlier in this chapter (contrast the hierarchy on page 114), where, instead of creating hierarchies based on the successive application of characteristics, simple subjects only are listed with the means of combining them whenever necessary, is much more satisfactory. Ranganathan has likened the latter method to a Meccano set – the pieces and fixings are provided, and innumerable compounds can be created.

General Classification Schemes: Special Problems and Traditional Features

The general principles and methods outlined above, though relating specifically to the creation of classification schemes for particular subjects, could no doubt be used in the creation of a general scheme of classification. A major problem here is that of 'main classes'.

(a) 'MAIN CLASSES'

With the exception of SC all existing schemes choose the major disciplines as their 'main classes': Economics, Medicine, Law, Engineering, and so on. Apart from the fact that the barriers between these main disciplines are increasingly difficult to distinguish, there is a further disadvantage in having such disciplines as the primary divisions of a scheme: documents which concern subjects (often 'concretes') treated from several aspects cannot be classified satisfactorily. For example, a book on Railways dealing with the Economic, Administrative and Engineering aspects, has to be classified according to *one* of these aspects – under Economics for example – with the result that it is lost to the students of Transport organization and

Engineering. Nor is it an easy matter to rectify this in the catalogue, as we shall see. The problem arises frequently, e.g. documents on *substances* (Oil, Coal, etc.), on *processes* (Measuring, Testing, etc.), *places* (GB, USA), *periods* (Eighteenth century), and so on. Investigation of this problem is in progress at the present time;⁶ though there are difficulties, there seems to be no fundamental objection to the faceted method outlined in this chapter being applied to knowledge generally in order to produce a general scheme of classification. The basis of such a scheme will not be the traditional 'main classes' but a series of facets capable of being compounded according to a preferred order with the possibility of alternative orders to suit particular needs.

The Classification Research Group has been largely responsible for exploratory work on a new general scheme, their investigations having been assisted for a time with a research grant from Nato. The Group has, through its members, close connections with BNB, and Derek Austin of the BNB has written several papers on the progress of the research.⁷

It would be optimistic to suggest that a new general classification is imminent, but certain fundamental features appear to have been established. The 'vocabulary' of the scheme is to consist of two schedules of terms designated *Entities* and *Attributes*. Examples of the former include terms arising from energy (force, gravity, heat), matter (molecular states, elements, compounds), earth, plants, animals, man. Attributes include such 'positional terms' as time and space; properties such as shapes, sounds, states, structures; and activities, e.g. equilibrium, kinetic conditions, motions and transfers. The order of these vocabularies is based on the philosophy of 'integrative levels', whereby terms are categorized according to level of complexity and arranged in an ascending order - as can be glimpsed from the outline of Entities given above. Each concept is represented once only in the vocabularies and it is hoped that basic generic relationships will be accounted for by the hierarchies of terms. Compounds will be built up according to an established citation order and the relationships between elements in the compounds will be made explicit by the use of a series of *Relational Operators* (e.g. property of a system, sub-system, interaction within the system, effect produced on the

system, attribute defining a type or class). The notational symbols for these operators will link the symbols representing Entities or Attributes to demonstrate the precise relationships between them.

Austin has stated that such a highly specific, synthetic scheme will inevitably produce lengthy class numbers: '... any idea of a short notation will now have to be abandoned. This is where enumerative schemes must score every time, but it should be remembered that an enumerated scheme fed into a computer will not allow retrieval of any particular element, whereas this proposed system of organized concepts will, and the computer, thank goodness, has no aesthetic scruples about long numbers.'⁸

To illustrate the sort of class number that could arise Austin gives the following: c35(2)q24(29)x75(299)v6(22)B(223)t44 - a hypothetical number standing for the document 'Energy balance in the turbulent mixing layer of a gas'. The basic citation order is Gases/Turbulent mixing layers/Energy balance. The numbers introduced by capital letters (c35 and B) are taken from the Entities schedule and stand for gases and energy respectively. Those introduced by lower case letters (q24, x75, v6 and t44) are taken from the Attributes schedule and stand in turn for layers, to mix, turbulence, and to balance. The numbers in brackets are Relational Operators all stemming from the basic operator (2), standing for a subsystem: thus layers is seen as a part of gases and so we get the sequence c35(2)q24.

Whether or not a new general scheme emerges in the next few years the work of analysing concepts and relationships is fundamental. Its influence can be seen already in the development of PRECIS indexing at BNB (see Chapter 9). There is no doubt that a new general scheme is badly needed to provide an alternative to the essentially nineteenth-century conceptions in general use today. A scheme such as this with its systematic analysis of concepts and relationships would also provide an extremely valuable basis for the creation of special schemes, effecting a much-needed measure of standardization.

Where *traditional* 'main classes' are used, some criterion for determining schedule order must be devised. (In a faceted

scheme this would be determined largely by the combination order of facets.) Bliss, who has systematically investigated this problem⁹ has said that 'main classes' should be carefully collocated so that related subjects are proximate. For example, Literature and Language should be proximate, as should Commerce and Economics and Business, Psychology and Medicine, and so on. Of course, there are limits to such collocation in schemes which are essentially one-dimensional linear sequences. In some cases *alternative locations* should be allowed to cater for special needs (some examples of these will be found in the section on BC below). The prefaces to BC are extremely valuable commentaries on the content of 'main classes' and the relationships between subjects.

(b) FACET ANALYSIS IN TRADITIONAL SCHEMES

Although enumeration of compounds is the traditional basis using the hierarchical model shown on page 120, all schemes have some degree of facet analysis – and in some it is considerable. The outstanding examples found in all schemes (save LC) are the *common subdivisions*, the main ones being for Place, Period, and Form of presentation.* These divisions can be applied at most points in the schedules – thus, in UDC (41) stands for Great Britain and 62 (41) would therefore represent the compound Engineering in Britain (where 62 means Engineering); similarly 338 (41) would stand for Economic situation in Great Britain. Again '19' stands for Twentieth century and thus 8 '19' would stand for Literature in the twentieth century. (03) represents the form Encyclopaedias and therefore 62 (03) would represent Encyclopaedias of Engineering. Further examples can be found in the notes on the schemes (Chapter 8).

By definition these are common *subdivisions* and cannot be used except in conjunction with a subject in the main schedules. The *Generalia class* results from this limitation; basically it houses *general forms* (general encyclopaedias, general collected essays, general periodicals, etc.). Such a class would be unnecessary if a *common facet* were created which could accommo-

* See Chapter 13.

date such documents and also be applied to any subject in the schedules in the manner of common subdivisions. Note: in practice the generalia class, as we shall see, is also used as a rag-bag for miscellaneous items that will not fit anywhere else.

Other examples of facet analysis in general schemes will be noted in Chapter 8. In general it is sporadic – for example, in DC there is the occasional instruction to ‘*divide like . . .*’ as in the Literature class where all Languages can be further subdivided by the divisions enumerated under English; thus 820 is English literature, 821 English poetry, and the 1 can be added to any other literature to stand for poetry, e.g. 831 German poetry, 841 French poetry, etc. However, in recent years there has been a notable trend in DC towards a more systematic analysis of categories and the use of synthetic devices. The most extensive use of analysis and synthesis in general schemes is to be found in UDC and, of course, in Ranganathan’s *Colon classification*.

Notation

Ranganathan has defined notation as ‘the system of ordinal numbers used to represent the classes in a scheme of classification’. To be efficient each class number should be unique, standing for only one distinct subject and used constantly for that subject.

PURPOSE

As already noted (page 93), the main purpose of notation is the mechanization of a preferred order; the order must be decided first. *Yet without notation the classification scheme cannot be put into operation at all*, for the only order without a notation is the alphabetical order of terms. A notational symbol is a shorthand sign and as such necessary for the arrangement of documents on the shelves; it also provides briefer and more satisfactory headings than words for the filing of entries for complex documents in the catalogue. The notational symbol is an easily memorized link between catalogue and shelves.



KINDS OF NOTATION

Notation can be *pure* or *mixed*: a pure notation uses only one kind of symbol (e.g. arabic numbers or roman letters), a mixed notation uses more than one kind of symbol (e.g. arabic numbers and roman letters). Numbers can be used as decimal *fractions* (e.g. 0 . . . 1 . . . 12 . . . 13 . . . 131 . . . 2 . . . 21 . . . etc.) or as *integers* (e.g. 1 . . . 2 . . . 12 . . . 13 . . . 21 . . . 131 . . . etc.).

Notation can be *expressive* or *non-expressive*. Expressive notation not only mechanizes the order but reveals the structure of the classification scheme. For example, in the Literature class in DC the hierarchy is reflected in the notation, a further digit being added at each stage of division:

8	Literature
82	English literature
821	English poetry
821.3	Sixteenth-century English poetry

Again, subjects of equal status in the hierarchy (coordinate subjects) have class numbers that *look* equal (82 English literature, 83 German literature, 84 French literature, etc.).

Expressiveness is not always found and is generally considered less important than it once used to be. In UDC, though this scheme is largely expressive, we find:

	531	Mechanics
	532	Fluid mechanics
	533	Gas mechanics
and	450	Italian language
	459.0	Roumanian language
	459.0	Ladin-Romansch. Rhaeto-Romanic
	460	Spanish language
	469.0	Portuguese language
	469.9	Galician language

A distinction should be made between the expression of generic relations (i.e. those existing between simple foci within a facet) and compound relations (i.e. those between two or more foci from different facets). As we shall see there is no inherent difficulty in expressing compound relations through the use of facet indicators, retroactive notation, etc. Generic relations are

another matter: there are obvious limits to the number of co-ordinate divisions that can be placed within any facet or sub-facet so long as we are aiming at expressiveness – the number being determined by the notation employed. Thus a numerical base will allow only ten such divisions, whilst an alphabetical base will allow twenty-four. Ranganathan has suggested the use of the various *sectorizing devices* for extending the capacity of a numerical base. One of these, the *octave device*, reserves the final digit 9 for a further set of coordinate divisions. With such a device 9 is never used alone but always introduces a further series of coordinate divisions: 91, 92 . . . 991, 992 . . ., etc. Though this certainly provides places for any number of co-ordinate divisions the resulting numbers do not *look* coordinate and are evidently so only to the initiated. Moreover this device alone is still incapable of allowing insertion of a new focus at a predetermined point in the schedule: for example, it will not allow the placing of a division between 2 and 3 which is still expressive.

REQUIREMENTS OF NOTATION

(i) *Hospitality*

The basic purpose of notation is to mechanize a preferred order, and if knowledge were static it would be easy to add a notation to the final scheme and there the matter would end. But new subjects are constantly being created and therefore notation must be able to accommodate these subjects as they arise – and in their proper place in the scheme: in a word, it must be hospitable. If the notation is inhospitable, it will *determine* order not merely mechanize it; the cart will be before the horse. For example, in an arithmetic notation (i.e. one using integers) if 62 represents the Novel and 63 Drama, there is no means of placing the Historical novel in its correct place – which is clearly *between* 62 and 63. Instead the nearest vacant number will be used (numbers are always left free for this purpose in an arithmetic notation). An arithmetic notation can never be fully hospitable because no amount of foresight can ensure that gaps will always be left as required. A decimal notation, on the other hand, is infinitely hospitable; had the

numbers in the above example been decimal fractions, the Historical novel could have been accommodated at, say, 625 (or 62.5); again, if 010 represents Bibliography and 020 Librarianship, and Documentation (considered as a subject including both Bibliography and Librarianship) had to be placed ahead of 010, then 009 might be used – or, if that were already in use, 0099, and so on. This is equally true of an alphabetical notation. For example, if Gb represents Bibliography and Gc Librarianship, then Gab could be used for Documentation and if Gab were already in use Gaab could be used, and so on.

No notation can ever be fully hospitable and maintain its expressiveness; in the examples above, the symbols for Librarianship and Bibliography do not *look* subordinate to the symbol chosen for Documentation. Nor is there any symbol that would look superordinate to these. Expressiveness is useful in that it expresses the structure of the scheme to some extent, but it must be sacrificed in the interests of hospitality – and many modern schemes have abandoned the concept, at least, so far as generic relations are concerned.

So far we have considered hospitality to *new foci* and simple *steps of division*; equally important is hospitality to *new compounds*. Such compounds are, as we have seen, the result of synthesis – the combining of foci from different facets in a scheme.*

One of the problems in notation is to create a device to indicate that the resulting symbol *is* a compound so that it cannot be confused with a straight division within the facet; for example, in UDC 531 stands for Mechanics and 62 for Engineering, but the compound Mechanics for engineers cannot be expressed as 531.62 because this is a division of 531, standing for Law of conservation of energy – in other words, the point cannot be used as a facet indicator in a scheme using a decimal notation and a point separator. *Combination cannot be indicated by an indicator that is already used for class divisions.*

Bearing this general consideration in mind, *combination can be indicated by the following methods*:

(a) *By such devices as colons, brackets, inverted commas, and so on* – as in UDC, e.g. 531:62 Mechanics for Engineers; 373.5:371.27

* The process of combining notational symbols to cater for compounds has been called *number building*.

Examinations in Secondary schools (where 373.5 stands for Secondary schools and 371.27 for Examinations); 5(09) History of science (where (09) stands for History).

(b) *By capital letters being used to indicate the different facets* – as in CIS where gtz means Dust (in the Substance facet G) and qfz means Sampling (in the Method of investigation facet Q) and so gtzqfz means Dust sampling.

(c) *By the reservation of certain sections of the base for certain facets* – thus, for example, in UDC class 656 (Transport organization) the series .01/.09 is reserved for the Operations facet and .1/.9 is reserved for the Type of transport facet. 656.032 stands for Passenger rates and 656.132 for Buses and motor coaches. As .032 is used here in the Operations facet only, the compound, Bus fares, can be expressed as 656.132.032.

A scheme which sets out its facets in inverted citation order, reserving sections of the base for each facet as illustrated here, is said to employ a *retroactive notation*.

When symbols with no ordinal value – such as brackets, colons, etc. – are used, they have to be given one and this makes filing and searching more difficult (thus in UDC, colons file before inverted commas, which in turn file before brackets, and so on). On the other hand it gives a certain measure of flexibility to the classifier who, by the simple expedient of altering the filing value of the connecting symbols, can achieve a filing order to suit his requirements.

(ii) *Notation should be easily comprehensible and have ordinal value*

This means that only roman letters and/or arabic numbers should be used; other symbols (triangles, for example) being unusual and having no ordinal value. It also means that as far as possible, facet indicators should also be symbols having ordinal value as seen already.

(iii) *Notation should be easily memorized, written and spoken*

This is partly dependent on its *comprehensibility*. It is also and primarily dependent on the *simplicity* of symbol. In general, numbers are more easily remembered than letters, but they have some disadvantages (see *Length of symbol*, below). It has been shown (Coates)¹⁰ that too much mixing of letters and figures is detrimental to memory and that much depends on the

placing and style of *separators* to break a long sequence (e.g. the point in DC and UDC). The best separators are those which have that function in everyday usage (points, brackets, and so on). 33847 (62148) and 33847K62148 are better than 3384762148 (no separators) or 3384U2K21P8 (three separators), and so on. If the function of separator is combined with that of facet indicator (as in CIS), and letters or figures are used to indicate order, this may lead to compounds with rather too many separators for easy memorizing. The use of such mixtures as BB211 is to be avoided. *Pronounceable* or *syllabic notation*, pioneered by Cordonnier and others, may well be increasingly used in new schemes – see, for example, D. J. Foskett's *London education classification*.

Another factor here is *length of symbol*. Inevitably class numbers will become longer as more complex subjects are specified. *Maximum brevity can be achieved by:*

(a) *A long notational base*: in a decimal system a numerical base (0/9) will result in longer class numbers than an alphabetical base (A/Z) – because whilst the former can accommodate only ten classes using one digit and a hundred using two digits (00/99), etc., the latter can accommodate twenty-six classes using only one letter, and 676 classes using two letters (AA/ZZ), etc. Bliss uses a base of thirty-six digits – 0/9 followed by A/Z.

(b) *The proper allocation of the notation*: subjects on which there is much literature and which are rapidly expanding (as Science and Technology) should be given more space than the relatively static subjects. DC gives Philosophy and Religion (100 and 200) the same space as the whole of Science and Technology (500 and 600) and therefore, as may be expected, class numbers in the former classes are much shorter than those in the latter where important classes start from a four-figure symbol or more (e.g. Electrical engineering 621.3, Aeronautical engineering 629.13).

(c) *Rejection of expressiveness*. Expressiveness will also lengthen class numbers – as every new subdivision will require a further digit in the notation.

(d) *Synthesis also tends to lengthen numbers*. The compound subjects on page 129 could be more shortly expressed if, instead of forming them by synthesis, they had been enumerated, e.g. Examinations in secondary schools could have been 373.51 (or

373.6). But enumeration is, as we have seen, to be avoided. Good allocation of notation and carefully designed synthesis can do much to reduce length.

Memory may also be assisted by *mnemonics*, i.e. the constant expression of certain concepts by particular symbols so that they become familiar in time to staff and even perhaps to users – e.g. periodicals (05), Great Britain (41). In this way enquiries can be rapidly translated into class numbers and class numbers on documents and in catalogues translated back into natural language. This is clearly an aid in reference work – not to mention its value in the process of classification and cataloguing. Sometimes ‘literal’ mnemonics are used (e.g. in LC the letter T is used for the class Technology, in BC Chemistry is C, and U stands for Useful arts). Literal mnemonics should be incidental – there is danger that in striving for them order and economy may be affected.

Throughout these brief notes on notation a manual system has been assumed. Where machine systems are in operation the characteristics and the priorities given to them will change. For example, length of notation is no longer a great handicap as the computer can search long numbers easily. The same goes for complexity of notation. Expressiveness for generic relations becomes more desirable if hierarchical searching is envisaged. Relationships may need to be made explicit through notational symbols – as in the example cited by Austin earlier in this chapter.

READINGS

1. VICKERY, B. C. *Faceted classification*. London, Aslib, 1960.
2. VICKERY, B. C. *Classification and indexing in science*. London, Butterworths, 1959. p. 35.
3. RANGANATHAN, S. R. *Elements of library classification*. 3rd ed. London, Asia Publishing House, 1962.
4. MILLS, J. *A modern outline of library classification*. London, Chapman & Hall, 1960. The principle of inversion is discussed on pp. 18–19. See also p. 44 – retroactive notation. Also Mills’ paper at the Dorking Conference (item no. 6 below).
5. For a further account see MILLS, J. op. cit. pp. 25–30.
6. FOSKETT, D. J. ‘Classification and integrative levels’ (*In*

- Classification Research Group. *Sayers memorial volume*. London, Library Association, 1961. pp. 136-50.) See also: VICKERY, B. C. 'Relations between subject fields: problems of constructing a general classification' (*In International Study Conference for Information Retrieval, Dorking, 1957. Proceedings*. London, Aslib, 1957. pp. 43-9).
7. (i) AUSTIN, D. 'Development of a new general classification: a progress report'. *Information Scientist*, 3 (3), November 1969, 93-115.
 (ii) Also the same author's 'Prospects for a new general classification'. *Journal of Librarianship*, 1 (3), July 1969, 149-69. And:
 (iii) LIBRARY ASSOCIATION. *Some problems of a general classification scheme*. London, Library Association, 1964.
8. AUSTIN, D. op. cit. - 7 (i). p. 99.
9. BLISS, H. E. *The organization of knowledge and the system of the sciences*. New York, Holt, 1929. *The organization of knowledge in libraries*. New York, H. W. Wilson, 1939.
10. COATES, E. J. 'Notation in classification' (*In International Study Conference for Information Retrieval, Dorking, 1957. Proceedings*. London, Aslib, 1957. pp. 53-4.)

In addition to the titles mentioned above and those found at the end of Chapter 6, the student will find the following useful:

- Classification research*: proceedings of Second International Study Conference, held at Elsinore, Denmark, 1964. Edited by Pauline Atherton. Copenhagen, Munksgaard, 1965. An advanced work.
- FOSKETT, D. J. *Classification and indexing in the social sciences*. London, Butterworths, 1963.
- PALMER, B. I. *Itself an education: six lectures on classification*. London, Library Association, 1962.
- PALMER, B. I. and WELLS, A. J. *Fundamentals of library classification*. London, Allen and Unwin, 1957.
- RANGANATHAN, S. R. *Prolegomena to library classification*. 3rd ed. London, Asia Publishing House, 1967.
- SAYERS, W. G. B. *Manual of classification for librarians*. 4th ed. by A. Maltby. London, Deutsch, 1967.

Chapter 8

Schemes of Classification

This chapter examines briefly the more important of the general schemes of classification. All were first produced before the theories of facet analysis had been fully established. All – except Colon, in which Ranganathan has experimented with the ideas of facet analysis and which is placed first for this reason – are, to a greater or lesser extent enumerative, and inevitably the schemes suffer from the limitations associated with enumeration; in particular:

(i) the enumeration is incomplete; many compound – and at times even simple subjects (concretes, for example) – are not catered for;

(ii) number-building facilities are limited and therefore many compound subjects cannot be accommodated;

(iii) a preferred combination order is too seldom observed and therefore cross classification is likely;

(iv) schedule order is confused because facets are not clearly distinguished and compounds made up of the same categories (e.g. Operations and Materials), and even simple foci belonging to the same facet, may be found in more than one place.

Clearly it is no reflection on the original producers of these schemes that they were unaware of methods created after their work was finished, but we can legitimately criticize any scheme for not meeting the basic requirements of a classification scheme mentioned at the beginning of Chapter 7. And we can legitimately use the principles outlined in Chapter 7 to indicate more clearly why the faults arise. This is, in fact, the method used in this chapter. Each scheme is examined broadly under such headings as: schedule order, combination order, hospitality of notation, and so on, to try to indicate why the limitations of the schemes arise.

Note: this chapter does not attempt to give a detailed description of the schemes but rather to use the schemes as examples of theoretical points already made.

flexible, 'freely faceted' scheme. A distinction is made between earlier editions in which 'the facets that go with each Basic Subject and their sequence are predetermined, without reference to various possible Compound Subjects capable of going with that Basic Subject', and the 7th edition in which 'predetermination of the facets for all Compound Subjects likely to go with any Basic Subject is ruled out'. The term 'facet' is re-defined as 'a generic term used to denote any component – be it a Basic Subject or an Isolate – of a Compound Subject'; and it is stated that 'above all this Version recognizes that facets belong to Compound Subjects and not to Basic Subjects'.

(v) Notational provision for the increase in the arrays of classes within the Main Subjects is made by using the lower-case alphabet as well as numerals. This will necessitate the use of double inverted commas before anteriorizing common isolates – e.g. B^a for a bibliography of mathematics, not Ba. Similarly, the zero, hitherto reserved to indicate phase relations, is also to be used for the formation of arrays, whilst the ampersand is to be used for phase relations – e.g. s&bT for a book on Psychology for teachers, not sobT.

(vi) Fundamental work on the nature of many foci in the Energy facets – e.g. Anatomy, Physiology, Disease, and Development, in the Biology and Medicine classes – has resulted in their being reconstituted as Matter-Property Isolates. These will be introduced by the semi-colon as the facet indicator. Thus the compound, Morphology of the cow, will have the class number KX,311;2 not KX311:2 as in the 6th edition. (The Personality facet is always to be introduced by a comma, where this was – as already shown – frequently omitted for the first round, first level, of Personality in the 6th edition. The reason given for this change is that the longer notation resulting from depth classification is easier on the eye when broken by a facet indicator.)

Melvil Dewey. Decimal Classification and Relative Index

First published in 1876. The 14th edition (1942) was the

'fullest'; the 15th edition (1952) was virtually an abridged edition. The 16th edition, though still shorter than the 14th, was once again expanded. In 1965 the 17th edition appeared, incorporating a number of new features (see below). There is an abridged edition for use in schools – now in its 9th edition. The scheme has been translated into many languages.

The most extensive revision was in the 15th edition where more than a thousand topics were re-located in the scheme. This represented a change from the traditional revision policy which, whilst admitting the need to accommodate new subjects, had laid stress on the policy, initiated by Dewey, of 'integrity of numbers', i.e. the maintenance of existing class numbers; thus new editions had been concerned mainly with expanding existing numbers rather than structural reorganization, and revision had been uneven, extensions being made to suit pressing demand and interest. Many libraries complained of the re-locations and in the 16th edition about forty-five per cent of them were once again altered – back to their original and often less satisfactory places in the 14th edition. This shows the pressure that can be brought to bear on the editors of an established scheme by librarians who do not wish to go to the trouble of re-classifying sections of their libraries to keep pace with the growth of knowledge.

17TH EDITION

However, a bolder stand has been taken in the 17th edition. Although the total number of re-locations is modest, the changes are often fundamental. In his Introduction the editor states that 'even though the total number of re-locations in this edition is less than half the number in Edition 16, the average effect of each is probably greater...' The new features can be briefly summarized as follows:

(i) *Integrity of subjects*. 'Classification by attraction', i.e. the placing of a subject at the most concrete element represented in it, without regard to the basic discipline concerned, is renounced. Thus, it is affirmed that, for example, a book on the sociology of the Jews should go in class 301 with sociology and not at 296, the number for Judaism. 'Classification by attraction' has always been a tendency, especially among American

librarians, and the practice had been given some support in the 16th edition.

(ii) *Facet analysis and synthesis*. Though still an essentially enumerative scheme, there is in this edition a distinct trend towards basic analysis into simple subjects with provision for combining them in a preferred order. For example, in class 630 Agriculture, it is now made clear that the citation order is to be Crop-Process, and notational devices allow such number-building as 633.155 Harvesting corn (where 633.1 is corn and the 55 has been detached from 631.55, the number for harvesting). Even when provision for compounding has not been made, there has been an attempt to suggest priority orders among facets to achieve consistency and to avoid cross classification. In his Introduction the editor gives certain general rules for combination order. Beyond the broad Subject-Place-Form, he suggests that 'the following precedence formula is a generally reasonable and helpful one to follow, although it may require modification in certain places: class the subject by (1) kinds, (2) parts, (3) materials, (4) properties, (5) processes within it, (6) operations upon it, (7) agents'. The increasing emphasis placed on synthesis can be seen both in the editorial remarks and the table, *Synthesis of Notation*, in Volume Two.

(iii) Allied to this concern with analysis and synthesis is the attention given to the general structure. Scope notes are improved and the use of black arrows to identify 'centered headings' is of great assistance in classifying and indexing. Again, the editor emphasizes the hierarchical structure and the schedules aim to demonstrate clearly the successive degrees of subordination. Where a subject is divided by more than one characteristic of division an effort has been made to spell this out, as for example in 373.2 where Secondary schools are categorized as follows: types as to control, types as to organization, and types as to curriculum.

(iv) Classes that have undergone major re-casting include Psychology, now provided with a new schedule at 150, leaving 130 for 'pseudopsychology, parapsychology, occultism'. The common form divisions have been re-labelled Standard Subdivisions in recognition of the fact that they cover more than forms of material, and several of them have been revised,

notably the divisions for historical periods at 090 which are a marked improvement on the skeletal divisions available in earlier editions. Particularly important is the removal of the place facet from the 900's and its presentation as an Area Table in Volume Two, emphasizing its general applicability throughout the scheme. It will be noted that the divisions in the Area Table have been revised and, in particular, provision is now made for concepts other than political boundaries – e.g. physiographic regions and socio-economic regions and groups.

The 18th edition, due in 1971, will no doubt develop these trends and we are promised some major revision in certain areas, notably in 340 Law and 510 Mathematics.

DC is still the most widely used of the general schemes of classification, though in recent years there has been a distinct movement towards the LC classification, and in the US in particular many librarians are busy re-classifying their stocks. The reasons for this change would seem to lie less in the superiority of LC than in the numerous advantages to be gained through the centralized services offered by the Library of Congress. Indeed, it might be argued that in terms of the basic criteria of classification LC can hardly stand comparison with DC (see Appendix), and that this process of re-classification is a reflection of the low esteem in which classification is generally held in the US where the dictionary catalogue is the main avenue for the subject approach, classification being regarded as little more than a shelving device.

The initial success of DC was immediate and has continued, despite its relatively slight decline, through the hundred years of its existence. The enthusiasm of its first reception was probably due to its uniqueness and to the following features in particular:

(i) relative location: then a new idea to overcome the disadvantages of 'fixed location', i.e. the reservation of certain shelves for set subjects;

(ii) the simplicity and hospitality of the decimal notation;

(iii) the relative index – always a good one and emphasized by Dewey as one of its most important features;

Its continuing use is largely due to:

- (i) its being firmly established in so many libraries;
- (ii) its permanent revision organization established in the Library of Congress;
- (iii) its use in schemes of centralized cataloguing such as BNB and Wilson;
- (iv) the continuing lack of a general scheme sufficiently excellent to convince librarians of the need for re-classifying their stocks.

Clearly there are advantages in adopting a scheme in general use – not only because of the consequent benefits of centralized cataloguing (see Chapter 20) but also for stock revision purposes, cooperative book acquisition, and so on.

Inevitably a scheme planned nearly a hundred years ago is bound to have its limitations. Over such a period of time even the basic structure of knowledge will undergo significant changes and it is hardly to be expected that even the most thorough revision will satisfactorily cope with the unprecedented growth of knowledge during this period and the increasing subject complexity of documents. Despite the improvements in the 17th edition, the scheme has been held back for years by the old policy of 'integrity of numbers' referred to above, the effects of which are not likely to be quickly mitigated. And it has to be recognized that the very popularity of the scheme will always act as a brake on the most radical editorial team.

In the following notes reference is made to both the 16th and 17th editions as both are widely used in libraries today.

MAIN CLASSES

As the notation used is one of decimal numerals, there are ten primary divisions, but in fact the first hundred divisions give a better impression of the major divisions of the scheme. The basic outline is as follows:

000 Generalities (including Bibliographies and Library science); 100 Philosophy (including Psychology); 200 Religion; 300 Social sciences (e.g. Political science, Economics, Law, Education, Commerce); 400 Language; 500 Pure sciences (e.g. Mathematics, Physics, Chemistry, Biology); 600 Technology (including Medical sciences,

Agriculture and agricultural industries, Domestic arts and sciences, Business); 700 Arts (including Photography, Recreation); 800 Literature; 900 Geography and history and related disciplines.

Collocation is limited in certain parts of the scheme, e.g.:

- (i) Language 400 – Literature 800;
- (ii) Technologies in 600 separated from their fundamental sciences in 500 (e.g. Electricity 537 – Electrical engineering 621.3; Chemistry 540 – Chemical technology 660);
- (iii) Economics 330 – Commerce 380 – Business 650;
- (iv) Buildings 690 – Architecture 720.

Within some of these main classes there are further anomalies; e.g. note the placing of Printing and Publishing in the following sequence:

650 Business, 651 Office services, 655 Printing (including publishing and related activities), 658 Management.

There are some examples of dated collocation, a particularly obvious one being the placing of Psychology in Class 100 (Philosophy).

There are a few examples of *alternatives*. Thus Bibliography at 010 allows for collocation of subject bibliographies, or they can be placed with subject; Biography 920 can be used to collect all biographies or they can be placed with relevant subjects associated with the person (the latter being preferred in the 17th edition).

INDEX

The index is detailed and relative, i.e. it concentrates on collocating distributed relatives. The following entry from the index to the 16th edition is typical:

Ships	
Accident prevention	614.864
Canal transportation	386.22
Construction and engineering	623.8
Naval science	359.32
Ocean transportation	387.2

(Incidentally this is a good illustration of the difficulties of placing such concretes as Ships when treated from many viewpoints in one document.) In general the index does not

wastefully repeat divisions found in the scheme itself, e.g. there are few numbers in the 800's to be found at the index entry Literature.

We have seen that two features of the 17th edition are the stress placed on the 'integrity of subjects' and the development of synthesis. Each of these underlines the need (always there, of course, though often ignored) to classify by the schedules rather than the index. In an attempt to force users to adhere to this basic rule the editors of the 17th edition pruned the index by omitting many 'minor' headings, referring the user to relevant broad terms and the schedules. Clearly, the more synthetic the scheme, the simpler the index (as illustrated by CC), but the mixed nature of DC creates difficulties in indexing and some of the simplifications did little more than involve the user in unnecessary searching. For example, the index entry *Mysticism* has the following subdivisions:

Religion

General works 291.14 (a misprint for 291.42?)

see also Religious experience

Under Religious experience we find:

General works 291.42

see also other spec. rel.

The schedules at 291.42, Religious experience, have the scope note: 'Mysticism, conversion, asceticism, self-discipline'. Specific religions can often be divided like 291, so 294.542 stands for Hindu religious experience (including mysticism).

So far so good – though why the user should make the journey from *Mysticism* to Religious experience in the index is not immediately clear as the latter makes no mention of *Mysticism*. However, the point here is that the reference *See also other spec. rel.* will work properly only if there is consistency in the schedules – i.e. if *each* religion is divided like 291.42. This could be assumed in a fully analytico-synthetic scheme, but the fact is that in DC there are exceptions – e.g. *Mysticism* in Judaism is at 298.71. To cover such exceptions additional index entries are required, e.g.:

Religious experience – Judaism 298.71

As a result of the complaints received on publication a new

index based on that of the 16th edition and appropriately modified, was distributed to users.

FACET ANALYSIS

Dewey was always aware of the necessity for number-building for compound specification and he was years ahead of his time in making some notational provision for this (see below). However the complementary measure of preliminary analysis of subjects into simple elements capable of being subsequently combined was generally unrecognized, and the scheme is still fundamentally enumerative of compounds.

As we have seen, the 17th edition moves in the direction of synthesis, and this is evident if we compare the following extracts. In the 16th edition we find levels of education isolated (e.g. 372 Elementary), but there is also the sequence:

- 371 Teaching, school organization, etc. (The Operations facet);
- 371.21 Admission;
- 371.211 Primary school admission standards (a compound due to the introduction of an element from the Level of education facet).

This has been revised in the 17th edition to give precedence to the level of education facet: Admission standards in elementary education is placed with Elementary education at 372.1216 (the final 216 being taken from 371.216 – Admission standards in general).

Again, in the 16th edition we find:

- 371.7 School health;
- 371.73 Physical education in schools (a compound due to the introduction of the Subject taught facet);
- 371.7322 Physical education in elementary schools (a triple compound due to the further introduction of the Level of education facet).

In the 17th edition physical education in elementary schools goes with Elementary schools at 372.86.

The successive application of characteristics of division and consequent enumeration of compounds is still characteristic of much of D.C. One of the most exemplary of the exceptions to

this is to be found in Class 400, Language, where essentially the Problem facet is outlined at 420 (covering such subjects as Etymology, Grammar, etc.) and each problem can be applied to most of the languages in the Language facet 430/499.

The inevitable results of this enumeration and limited analysis are, apart from the length of schedules:

(i) *Many simple subjects and compounds cannot be specified* (because not all of the latter can ever be listed). For example, in Class 800 (16th ed.), in some respects a good example of analysis, the Literary period facet has not been isolated, nor is there any General period facet of any consequence in DC, and therefore it is impossible to specify, for example, the simple subject Literature, 1837-1900, or the compound English literature, 1837-1900. But the compound English drama, 1837-1900 can be specified because it is enumerated (and in fact Period can be specified under any Form, but not elsewhere). There is some improvement in the 17th ed. where, at 809, divisions .01-.04 can be used for historical periods, though specification is limited to the periods listed in the *Table of Standard Subdivisions*. See page 150 for further examples.*

(ii) *Related material is scattered*. In the example quoted above (16th ed.) the teaching of a particular subject (Physical education) in elementary schools is separated from the teaching of other subjects in elementary schools at 372.3-372.8. In the 17th edition, despite the implicit rule that the Level of education facet takes precedence, we find at 376 (Education of women) a place for 'the education of women by level' - e.g. 376.63 Secondary. (However, this could be justified on the grounds that special categories of education take precedence over level.)

(iii) *Cross classification is likely*. For example, at 372.8 (under Elementary schools) in the 16th edition any subject taught in elementary schools can be classified - and therefore the subject Physical education could be placed here, with the risk of

* The premature listing of compounds is frequently found. Thus at 385/388 (16th ed.) there are divisions for Railway transport, Road transport, etc., but there is no place for Transport. At 757.4/.6 (17th ed.) there are divisions for Men's portraits, Women's portraits, etc., but there is no place for Portrait painting, the number 757 (Human figures and their parts) having to stand for this.

having material on the same subject in two places: here and at 371.7322. In the 17th edition there is a place for Public schools (i.e. State schools in the British sense) at 373.224, and a series of divisions at 373.3-.9 for Secondary education and schools by continent, country, and locality - but no clear indication of how to deal with a document on State schools in England. Again, Public libraries are given a place at 027.4 and Cataloguing can be found at 025.4 but what is the classifier to do with a document on Cataloguing in public libraries? Whenever a compound is not listed a document on that compound can be placed at any of the simple subjects that comprise it, and unless the cataloguer creates his rules for this, cross classification is inevitable. However, the scheme often contains instructions to assist here: for example, in Class 630 Agriculture (16th ed.), Harvesting is 631.55 and Wheat 633.11, and though the compound Harvesting of Wheat cannot be specified we are told to place such compounds with Crop rather than Operation. (As indicated above, the 17th edition has provided number-building facilities at this point.)

(iv) *The order of facets (schedule order) is frequently confusing.* For example, in Class 200 Religion (16th ed.), 260 is Christian Church (262 Church government and organization, 264 Ritual, 265 Sacraments, etc.) but Church and parish administration is at 254, Pastor (life, everyday duties of minister, priest, etc.) 253, Preaching 251. (There has been some attempt to clarify this area in the 17th edition, but little basic improvement is to be seen.) In Class 370, 371 accommodates various types of teaching Operation or Problem, 372/374 accommodate Grade of Education (Elementary, etc.), 375 returns to Problems (Curriculum), 376 Education of women (other special groups are at 371.9), 378 returns to Grade (Higher education) and 379 concerns an organization problem (Education and state). Again, confused order is found in Class 900: 900 General history, 910 Geography, 920 Biography, 930 Ancient history, 940/990 Modern history. However, there is a general tendency to place the Problem/Operation facet before the others (e.g. in Agriculture, operations such as Harvesting, Irrigation, etc. precede the Crop facet) and this is generally regarded as resulting in an order of increasing speciality.

Combination order

(a) *In enumerated compounds.* It is impossible to generalize about these as they do not follow a consistent pattern. (It should be remembered, however, that in the Introduction to the 17th edition the editor has suggested a basic citation order to be followed as a general guide.)

(b) *In synthesis.* The amount of number-building possible is much reduced in the 15th and 16th editions – and is in fact, confined to the common subdivisions (form, place, period) and to the instructions in the scheme ‘divide like . . .’ (for an example of the latter see page 151). The common subdivisions precede at 01/09 the special subdivisions of a subject in the approximate order: Form of presentation – Period – Place. For example (notation according to 16th ed.):

360.2	Handbook to social welfare
360.3	Encyclopaedia of social welfare
360.904	Social welfare twentieth century
360.942	Social welfare in Great Britain
360.942082	Social welfare in Great Britain in the twentieth century
361.5	Disaster relief
etc.	

NOTATION

Pure – numerals used decimally, with a point separator after the third place. Three figures at least are used: 100, 200, 250, etc., not 1, 2, etc.

Hospitality. The decimal notation allows virtually complete hospitality to new foci and steps of division, but the notation is less hospitable to compounds. As already pointed out the amount of synthesis possible is much reduced in the later editions. Up to the 14th edition it was always possible to join any two class numbers by the use of 0001, e.g. 532.000162 (Fluid mechanics for engineers) and although this was a cumbersome device it was better than nothing. Now, despite the improvements in the 17th edition, many compounds cannot be specified, e.g. Portrait painting in oils (Oil painting 751.45, Portraits 757).

Standard subdivisions (01/09) can be used at any point in the

scheme (as shown in the above example). They cover common approaches to a subject (philosophy) and forms of presentation (encyclopaedias, periodicals) as well as periods of time. Although the common period divisions have been extended in the 17th edition they allow only limited specification – contrast the hospitality of UDC here. Place divisions have been removed to a separate Area table – see above.

'*Divide like . . .*' is a frequent instruction in the schedules, allowing compound subjects to be specified, e.g.

- 338 Production
 338.47 Specific goods and services
 Divide like 001–999, e.g. Motor vehicle industry:
 338.476292 (where 6292 is taken from the
 Engineering class to represent Motor vehicles).

Again:

- 823 English fiction
 Divide as instructed at 820–890. (Thus 823.91 stands
 for Twentieth-century fiction, the .91 representing
 Twentieth century being taken from the period
 divisions for English literature at 820.9001/.900914.)

The notation is frequently expressive:

- 600 Technology
 620 Engineering
 621 Applied physics (Mechanical engineering)
 621.2 Power derived from liquids
 621.21 Water wheels

Because of this expressiveness, the use of numbers only, the cumbersome synthesis ('divide like . . .'), and the fact that the notation is badly apportioned (e.g. 200 Religion and 400 Language have as much room as the whole of Science and Technology), the notation is lengthier than it need be, e.g. 610.73 Nursing, 150.195 Psychoanalysis, as against 221 Old Testament, 227 Epistles, 419 Non-verbal languages, etc. However its simplicity – through the use of numbers only and the lack of non-ordinal signs – helps to compensate for this defect to some extent.

Mnemonics. Clearly literal mnemonics are not possible but the use of certain numerals fairly regularly, though not always

consistently, throughout the scheme (e.g. -42 for British Isles, -03 for Encyclopaedias, and in special sections, e.g. in Literature: 1 for Poetry, 2 for Drama, etc.) has an important mnemonic value.

Universal Decimal Classification

UDC was first introduced as a result of the Brussels Conference on Bibliography in 1895; the work was carried out by the old IIB and the first edition (in French) appeared in 1905. The aim was to create a scheme more suitable than any existing at the time for the classification of specialized pamphlets, articles, abstracts, and so on – and particularly for such materials in the fields of science and technology. DC was used as a basis for the new scheme.

Since 1905 several versions of the scheme have appeared, e.g. 2nd edition (again in French), 1927–33, 3rd edition (German), 1934–53. The 4th edition (English) was started in 1943 but is still not complete; however, an abridged English version was issued in 1948 (3rd edition of this 1961).* In addition to full and abridged editions (and, more recently, intermediate editions), individual classes have been separately issued – e.g. (in English) education, building, mineralogy. After the 1914–18 war FID became responsible for the central organization of UDC and various national bodies, in this country the British Standards Institution, have assumed responsibility for their own editions. Revision is constantly in progress, though the international nature of the scheme makes it a slow process at times. FID regularly publishes *Extensions and corrections to UDC* and what are known as 'P-notes' – working documents circulated among those actively engaged in the process of revision.

UDC is used by an increasing number of special libraries throughout the world and also in many published bibliographies, e.g. RIBA *Catalogue*, *Index bibliographicus*, *Aslib booklist*, *Electrical engineering abstracts*, etc. Its popularity can be attributed to its hospitality to specialized compound subjects and to the usual advantages arising from standardization. Moreover, the

* This version has been used for the examples in this chapter.

notation permits a variety of combination orders of facets (see below), frequently allowing any library to choose the order most suited to its particular purposes – clearly an advantage and one not found to the same extent in the other general schemes. No general scheme is more suited to the needs of special libraries, especially those in science and technology.

MAIN CLASSES

The basic outline is still that of DC – see the *Outline of main divisions* that prefaces the scheme. Two modifications are immediately apparent: the confinement of Psychology to 159.9 and History to 93/99. Thus to some extent the scheme still has the weaknesses of DC in its main class order (the separation of Economics, Commerce, and Business and of Language and Literature, for example) – though in fact the disadvantages have been minimized by more careful definition of the scope of overlapping classes, and in the case of Language and Literature suggestions are given for their amalgamation. Recently a decision has been taken to vacate Class 4.

INDEX

UDC like DC has a detailed relative index which concentrates on collocating distributed relatives rather than repeating unnecessarily the divisions of the classification (contrast BC index for example).

FACET ANALYSIS

It is in this respect that UDC differs so greatly from DC. The content of most classes has been radically revised. Enumeration has been pruned and many classes are now mainly composed of simple elements, rather than enumerated compounds. This accounts for the comparative brevity of the scheme: the abridged British version has some 119 pages for the scheme itself, yet despite this many more compound subjects can be specified than by the much larger DC. The Literature, Geography, Biography, and History classes are perhaps the best examples of this – three pages in all. Note also the -1/-09