DRAFT: NOT FOR PUBLICATION OR QUOTATION

C O C T A

COMMITTEE ON CONCEPTUAL AND TERMINOLOGICAL ANALYSES
INTERNATIONAL POLITICAL SCIENCE ASSOCIATION

WORKING PAPER NO. 3

Relationship Between Elements of Knowledge

— use of computer systems to facilitate construction, comprehension and comparison of the concept thesauri of different schools of thought

Anthony J. N. Judge
Union of International Associations
1 rue aux Laines, Brussels, Belgium

Social Science Research Institute
University of Hawaii
Honolulu, Hawaii

March 1972

CONTENTS

(contents)

QUOTES

Some views on the complexity of knowledge and its representation

"Handling complexity seems to be the major problem of the age, in the way that handling material substance offered challenge to our forefathers. Computers are the tools we have to use, and their effective use must be directed by a science competent to handle the organization of large, complex, probabilistic systems."

(Stafford Beer. Managing modern complexity. In: The Management of Information and Knowledge, a compilation of papers prepared for the 11th Meeting of the Panel on Science and Technology. Washington, Committee on Science and Astronautics, US House of Representatives, 1970)

***

"Des mots aussi courants que "groupe", "classe", "pouvoir" ou "structure" comptent actuellement non pas deux, ou trois, ou quatre significations fondamentales -- ce qui est est normal -- mais autant d'acceptions que d'auteurs, acceptions parfaitement irréductibles à un commun dénominateur, et même totalement autonomiques."

(M. Moulin. Discours inaugural du Colloque sur le mot structure sous les auspices de l'UNESCO. In: Roger Bastide (Ed.). Sens et usages du terme structure dans les sciences humaines et sociales. The Hague, Mouton, 1962, 165 p.)

***

"I cannot emphasize too strongly the importance of this activity of intellectual synthesis...Any notion that we may have about the nature of science includes the belief that something like an overall pattern is to be discovered and described. What we need is scientific knowledge -- not more and more miscellaneous and unrelated information."

(J.M. Ziman. Information, communication and knowledge. Nature, 224, 25 Oct. 1969, p. 323)

***

"...face à la spécialisation croissante de la pensée et de l'action par la diversification de la recherche et la division du travail...(UNESCO)se doit de favoriser les recherches et les confrontations interdisciplinaires, d'encourager les réflexions d'ensemble, bref de souligner l'importance vitale de l'esprit de synthèse pour l'équilibre de notre civilisation. Je dis bien vitale, car l'homme -- j'entends l'essentiel, à savoir son jugement, sa liberté -- peut aussi bien être asphyxié par son savoir que paralysé par son ignorance, et il peut tout autant se perdre dans la complexité d'un comportement social dévorant que s'atrophier dans la simplicité élémentaire d'une condition dite de sous-développement."

(Allocution de René Maheu, Directeur Général de l'UNESCO au colloque international de l'UNESCO sur le thème "Science et Synthèse". Paris, Gallimard, 1967)

***

"The practical manifestation of this interdependence... (between the social sciences and the human sciences) is interdisciplinary co-operation, which culminates in multidisciplinary research and is embodied in teamwork: an indispensable basis for the knowledge of man, but at the same time an idea which, in general, abstract terms, has a dangerous fascination, and which might remain no more than an empty and unproductive slogan if its foundations and mechanisms are not clearly identified by contact with specific problems propounded to research, and with due regard to the various institutional, financial, technical and human factors on which its development, fruitfulness and capacity for renewal and creation in fact depend."

(René Maheu, Director General of UNESCO. In: UNESCO. Main Trends of Research in the Social and Human Sciences. Paris, UNESCO, 1970, vol. 1, p. xiv-xv)

***

"The most probable assumption is that every single one of the old demarcations, disciplines, and faculties is going to become obsolete and a barrier to learning as well as to understanding. The fact that we are shifting from a Cartesian view of the universe, in which the accent has been on parts and elements, to a configuration view, with the emphasis on wholes and patterns, challenges every single dividing line between areas of study and knowledge."

(P.F. Drucker. The Age of Discontinuity; guidelines to our changing society.)

***

"The possibilities open to thinking are the possibilities of recognizing relationships and the discovery of techniques of operating with relationships on the mental or intellectual plane, such as will in turn lead to eve wider and more penetratingly significant systems of relationships."

(B.L. Whorf. Language, Thought and Reality, p. 84)

***

"One of the most significant results that should naturally emerge from a study such as this, is the presentation of a chart -- admittedly provisional and subject to constant revision -- of the strong points and weak points of interdisciplinary cooperation and of their substratum, and the identification of priority areas to which research scientists should direct their thinking and institutions their activities."

(René Maheu in the preface to: Main Trends of Research in the Social and Human Sciences. Part one: social sciences. Paris, UNESCO, 1970, p. xxv.)

***

"Like the life forms of the physical world, the dreams of men spread and colonize their inner world, clash, excite, modify and destroy each other, or preserve their stability by making strange accommodations with their rivals. So I regard it as a legitimate analogy, though not, of course, an exact one, to speak of our interpretative system -- I call it an appreciative system -- as an ecological system, even though the laws which order and develop a population of ideas (conflicting, competing, and mutually supporting) in communicating minds are different from those which order and develop a population of monkeys in a rain forest or of insects under a paving stone."(p.11-12) and, in the same context, "Every field of activity, politics, law, and not least science, like every society, has its own stability to guard." (p. 182)

(Geoffrey Vickers. Value Systems and Social Process. London, Pelican, 1970.)

***

"L'uniformité croissante de la terminologie éthique que l'on remarque dans les réunions internationales, où elle facilite assurément les communications formelles et les ententes apparentes, ne doit pas nous abuser. Derrière le mur de brouillard des mots, la diversité, voire l'opposition des interprétations, des motivations et des utilisations divisent profondément les esprits et, à la faveur de cette confusion, les droits universels sont bien plus souvent invoqués comme une arme offensive ou défensive contre autrui que reconnus et pratiqués comme la route royale de l'union de soi et d'autrui en une fraternité objective."

(René Maheu, Directeur Général de l'UNESCO, 15e Conférence Générale, Comptes rendus des débats, Paris, 1968)

***

"Thus is affirmed now life and now death, now death and now life; now the admissability of a thing and now its inadmissability, now its inadmissability and now its admissability. The disputants now affirm and now deny, now deny and now affirm. Therefore the sagely man does not pursue this method but views things in the light of his heavenly nature, and hence forms his judgment of what is right.....
Words are like the waves acted on by the wind: the real point of the matters discussed by them is lost. The wind and waves are easily set in motion; the success of the matter of which the real point is lost is easily put in peril. Hence quarrels are occasioned by nothing so much as by artful words and one-sided speeches."

(Chuang Tzu (4th century B.C.) from Chuang Tzu: Genius of the Absurd, arranged from the work of James Legge, by Clae Waltham. N.Y., Ace Books, page 50 and 72.)

***

Preface

This report has been prepared for the Committe on Conceptual and Terminological Analysis (COCTA) of the International Political Science Association. It constitutes a much-expanded version of a set of notes which were discussed in relation to proposals in COCTA Working Paper No. 1 (prepared by Fred Riggs, Secretary of COCTA), and in the COCTA Manifesto (prepared by Giovanni Sartori, Chairman of COCTA,and Fred Riggs) at a meeting sponsored by the International Studies Association and held, on the invitation of the Rockefeller Foundation, at the Villa Serbelloni (Bellagio, Italy, 1-5 September, 1971).

Although COCTA currently derives its main support from the political science field, it is intended that its approach should if possible be made relevant to a broad range of social sciences. This report has therefore been written in such a way as to make the design useful to a variety of disciplines and users.

In order to achieve this, a very simple approach has been adopted which results, from a computer-level perspective, in a means of handling any entities or relationships. In this report, the stress has been placed on concepts, theoretical constructs, etc. It could equally well handle (i) organizations, and other social system entities and relationships, (ii) real "world problems" and their relationships, or possibly (iii) personal beliefs and their interrelationships.

The computer approach suggested here is in fact the simplification and modification of one developed by the author for the Union of International Associations with a view to creating a data bank of entities significant to the international system, based initially on the contents of the Yearbook of International Organizations. (1) The problems of concept-handling arose in the treatment of organization program concerns.

This work also suggested a more systemic approach to education about the interrelationships between fields of knowledge and activity (2) as well as the possibility of handling and analyzing information on interlinking "world problems" -- in

1 A.J.N. Judge. "Visualization of the organizational network." International Associations, 1970, 22, 5, May 1970, p.265-281.

_____."Information Systems and inter-organizational space." Annals of the American Academy of Political and Social Science, 393, January, 1971, p. 47-64.

_____. "Inter-organizational data and data bank design." Paper presented at the annual conference of the International Studies Association, San Juan, 1971.

2 J.W.Clark and A.J.N.Judge. Development of trans-disciplinary conceptual aids. Brussels, Union of International Associations, 1970, UAI Study Paper INF/7.

conjunction with the production of a Yearbook of World Problems (1).

The possibilities of this computer approach may in fact be most quickly recognized and funded in studies of entities in natural environment systems where the representation of complex interlinking "food chains" and "food webs" has,to date, posed an insurmountable problem (2). The use of the interactive graphics techniques suggested here, and for which a demonstration film has been prepared (3), may constitute a breakthrough in handling organized complexity. Hopefully it will be possible for groups interested in the different uses of the same type of network analysis computer program to work together in building up a case for funding -- particularly in the case of the graphics display programs.

This report makes reference to activities and techniques in a wide spread of domains. Clearly the author can claim no special competence in all these domains. It is nevertheless important to juxtapose material from such different sources rather than simply provide a bibliographical citation, particularly as much of it is relatively inaccessible. A number of the Appendices are therefore summaries, partial extracts, or commented extracts from published material. It is hoped that this approach will facilitate the reader's task in appreciating the many facets of this project.

1 A.J.N. Judge."What is a World Problem?" International Associations, 23, 5, 1971, p. 268-273.

2 D. Pimentel. "Complexity of ecological systems and problems in their study and management." in: K.E.F. Webb (Ed.) Systems Analysis in Ecology. New York, Academic, 1966, p. 15-35.

3 A.J.N. Judge." Visualization of Organization." 16mm film, Brussels, Union of International Associations, 1970.

Introduction

This report addresses itself to the practical problems of developing a means of filing concepts and other theoretical constructs in a data bank. Such concepts would be filed in terms of their meaning and not in terms of the word by which they happen to be represented in a particular school of thought. The reason for this approach is that many of the words on which most reliance is placed in the social sciences (e.g. "group", "class", "power", or "structure") have acquired a multiplicity of overlapping meanings (*).

The concept file so created would be used to generate lists, to facilitate classification and interrelation of concepts to produce concept thesauri, and, finally, to facilitate the allocation of "authoritative" terms to permit the production of terminological thesauri.

The object of this project would be to ensure that any qualified person -- with a few safeguards -- would be free to register entities in the file which would then become available for secondary analysis at any interested research centre.

One form such analysis might take would be the construction and comparison of various models or classification schemes for theoretical entities. At a tertiary level, efforts could be made to link such entities with each other, cutting across the boundaries of disciplines, ideologies, epistemological approaches, paradigms or problems. This activity would provide new alternative means of approaching the entities held on the file but would not affect their use for more restricted purposes.

In this report particular attention has been paid to some of the techniques available to analyze complex entity networks or structures. Because of this complexity and the problems of comprehending it, the use of interactive computer graphics has been examined as a powerful means of simplifying the task and making the project more widely significant.

(*) Fred W. Riggs,"Concepts, Words and Terminology."Honolulu, University of Hawaii, Social Sciences Research Institute, 1971, 66 p. (Committee on Conceptual and Terminological Analysis, Working Paper n° 1.)

Giovanni Sartori. "Concept misinformation in comparative politics." American Political Science Review, 64, December 1970, 4, p.1033-1053.

Proposal

1. Project Objectives

A project to handle, structure, and analyze theoretical constructs is proposed which would be operated as three distinct phases:

-- a concept-filing phase leadin to the creation and maintenance of concept inventories

-- a concept classification phase, leading to the production of concept thesauri

-- a term-allocation phase leading to the production of standardized terminological thesauri

A translation phase, to make the project more widely relevant, would sun in parallel with the above three. Each succeeding phase builds on the previous one, but need not necessarily follow it immediately in time for the project as a whole to be of value.

1.1. Concept Inventory Phase

A computer-based concept registration or tagging system should be set up which would allocate sequence numbers to concepts on a continuing basis. The criteria for concept registration should be kept to a minimum to ensure that the system repains "open" to a wide variety of users and contributors.

This approach permits rapid inclusion and organization of the data and rapid production of updated concept lists. These would facilitate the scrutiny of the data in various phases and in terms of the perceptions of different need groups.

1.2. Concept Classification Phase

Evaluation, classification and identification of concept interrelationships would be made independently by a limited number of contributing groups, possibly associated organizationally with the international academic bodies. These groups would be primarily concerned with allocating codes to be fed back to the computer system so that ordered and refined concept thesauri could be produced to reflect the perceptions and needs of the contributing groups. An important aspect of this coding function by groups would be the rejection of those conceptions registered which are considered to be of little value to the group's perspective.

From the computer data handling point of view, each contributing group would be building, refining, and maintaining its own "model". Each such model would be handled as an independent optional qualifier on the sequentially-ordered concept list.

From the point of view of any such group, the computer system would be viewed as holding the concepts in which it is interested in the order of its own preferred classification scheme.

There would of course be the opportunity at any time to look at the same concept list through the classification scheme of any other contributing school of thought. Concepts would be identified by their sequential number plus a number which would identify the model employed.

1.3. Term Allocation Phase

At a later stage users of one model might find it useful to produce an "authoritative" list of terms to be used for those concepts of interest to them. This could also be incorporated into the computer system. Such terms could then be used to produce standard terminological thesauri for the users of one model.

2. Project Organization

2.1. The general organization of the project is outlined in Appendix A1.

2.2. The problems of classification or modelling of conceptual entities and the advantages of a sequence number are examined in Appendix A2. The project is conceived as being of use in a variety of domains. A summary is given of the range of possible conceptual entities (Appendix A4), relationships between entities (Appendix A5), types of entity classification scheme or model (Appendix A3), and data to be incorporated on each entity (Appendix A6). Priorities are suggested in order to limit the scope of the project (Appendix A7). A standard form of concept notation for use in print, but independent of the operation of the system, is suggested in Appendix A8.

Each contributing group may wish to distinguish differently between, or interrelate, the "entities" tagged in the computer sequential register. There is no reason why "concepts" "propositions", "relationships", "problems", etc. should not all be treated as entities and appropriately distinguished and interrelated (or ignored and rejected) at the modelling phase. It might, for example, be particularly valuable to include "theories", "frameworks of inquiry", etc. by first giving each a sequential number (as indicated above) and then (in the modelling phase) relating them to the major variables considered significant and necessary to define the frame of discourse associated with that theoretical viewpoint.

This would permit the same system to handle concept thesauri, inventories of propositions, inventories of problems, etc.

2.3. Once the concept registration system is running smoothly and the professional groups are interacting effectively with

the system to feedback their classification of the concepts within their own models, other groups of different levels of "multi-disciplinarity" may constitute themselves to work on the integration into "meta-models" of two or more of the models already produced (e.g. for political science and sociology into a social science model).

2.4. There is no reason why, for example, a copy (on computer magnetic tape) of the concept list and various models should not be made available to universities for comparative research on the models or as a tool in the educational process. Alternative models could be constructed which could be made generally available.

With respect to research, it is clearly important to enable the user to examine the thesauri at different levels of abstraction by introducing filters. In addition there is the possibility of comparative study of the manner in which different disciplines perceive and interrelate phenomena.

With respect to education, it is possible to develop educational meta-models which would permit selection of concepts by filters corresponding to different educational levels (e.g. an "atom" may be viewed as a billiard-ball type structure in the elementary stages, a miniature solar system, a system of electrically-charged potential clouds or, in the final stage, as something which can only be described with mathematical symbols). At each level a precise definition in the appropriate terms could be provided. In addition the approach could permit individual students to create their own concept thesaurus and to learn from the differences between their own and those of particular disciplines.

3. Computer Techniques

3.1. The outline of the design of a suitable computer record is given in Appendix A6. Suitable record-handling software is discussed in Appendix B1. Particular attention should be paid to the approach used by the team at the M.I.T. Center for International Studies (Appendix B2).

3.2. The graph theory techniques mentioned in the next section (4) suggest the need for more powerful ways of displaying and interacting with the network of theoretical constructs represented in computer memory -- in order to avoid the necessity to generate long, indigestible, and impenetrable lists (however ordered). The use of the interactive computer graphics technique for this purpose is examined in Appendix B3. Suggestions for the design of suitable graphics demonstration programs are made in Appendix B4.

3.3. It should be stressed that the basic programs required to operate the filing and listing functions are very simple and could be produced without making use of any sophisticated techniques or computers.

Some of these more sophisticated techniques have been discussed to give some idea of the analytical possibilities. In fact there is no reason why some institutes should not use the file in its simple form whilst others convert it into a more complex form.

4. Methods of Representation and Analysis

4.1. Particular attention has been given to the use of graph theoretical methods to handle the complex theoretical constructs (Appendix C1)

4.2. Graph theory and related techniques have been used in the fields of artificial intelligence (Appendix C2), personal constructs (Appendix C3), input/output analysis (Appendix C5) and semantic matrices (Appendix C4). These particular uses are closely related to those possible in connection with this project and represent areas from which analytical techniques and computer programs may be obtained and adapted.

5. Earlier Initiatives and Sources of Concepts

5.1. There have of course been many previous initiatives in this field. Some of these are discussed in Appendix D1. Efforts to develop conceptual discionaries are discussed in Appendix D2.

5.2. A number of different techniques and proposals are discussed to establish the special focus of this project in relation to them -- citation indexing (Appendix D3,D7), the Universal Decimal Classification and Dewey systems Appendix D4), the Aligned List of Descriptors (Appendix D5), the UNISIST World Science Information System (Appendix D6),the International Standard Book Numbering Technique (Appendix D8), and the SATCOM recommendations (Appendix D11).

5.3. Sources for the social science concepts required for this project are suggested in Appendix D9. A list of organizations, mainly international, which might be interested in one or more aspects of this project is given in Appendix D10.

6. General Considerations

Most earlier initiatives and proposals examined seem to fall foul on one or more of the following difficulties:

6.1. The simple and unambiguous administrative task of filing entities is merged into the complex intellectual task of coding and classifying them. This makes the whole project lengthy, costly, and complex.

--In this project the identification of entities to be included in a thesaurus and the practical problems of incorporating these entities into an information system are distinguished from the theoretical problems of classifying and interrelating such entities. The first is a relatively fast and unskilled operation and the second is a relatively slow and skilled one. The technique of identifying the entity within the system by a numerical tag derived from a classification scheme is avoided.

The savings in labor associated with this technique are only significant in a system in which all operations are manual. Where computers can be used, the two types of operation can be distinguished in order to save resources, speed up operations and increase the flexibility of reconceptualization of any classification scheme.

6.2. The classification of theoretical constructs may be associated with an intellectual and material investment in a document physical-location system. This opposes any flexibility or major reconecptualization of relationships between entities.

--In this project there is no direct relationship between the classification scheme(s) and the physical problem of locating source documents.

6.3. The classification scheme may be rigid and "final", based upon a high commitment to a particular set of theoretical assumptions of limited comprehensiveness, and therefore unable to adapt to new types of interrelationships.

--In this project both rigid and rapidly evolving classification schemes can be used to interrelate the entities handled.

6.4. The classification scheme may be exclusive or "inhospitable" and therefore of limited use.

--In this project both exclusive and hospitable schemes may be used. This gives it a wide range of uses.

6.5. Some systems are specifically designed with the special problems of a particular field of knowledge in mind. This makes them difficult to use in other areas.

--In this project specialized and general overdesigning the information handling system to meet immediately-perceived needs would reduce its usefulness and relevance to others and therefore increase the difficulty of ensuring adequate funds over a long period. (The degree of "hygiene" introduced may be inversely proportional to the utility of relevance of the system to potential users.)

6.6. Exclusive or rigid schemes, once created, are viewed and defended as unique and "universally applicable" by their proposers, thus eliminating any possibility of more comprehensiveness, better-funded, joint efforts.

--In this project, every effort has been made to ensure that it does not become associated with particular schools of thought, organizations or personalities who might resent criticism of their perspective and alienate potential collaborators. All such individualism is contained within the model building activity which does not jeopardize other models or the project as a whole.

6.7. Even adequate universal schemes may become viewed as authoritarian and a vehicle for some form of conceptual imperialism. Unfortunately the organization of relations between entities is equated with the imposition of a new set of relations. The organizers are perceived as acquiring power. Such systems may give rise to a proliferation of competing alternatives for groups of users with slightly different perspectives on subject areas (e.g. UDC, Dewey and UN/OECD Aligned List of Descriptors) who need a tool with slightly different properties.

--see 6.6.

6.8. The actual procedures for incorporating new entities into any "approved" list within the system may appear bureaucratic and stultifying unless the system is user-oriented. There is therefore the old problem of minimizing the bureaucratic desire for due process and order and maximizing user participation.

--In this project suggestions have been made concerning means of maximizing user participation.

6.9. The system may be designed with only one type of user in mind, e.g. scholars or students. New systems, which compete for the same resources, then have to be created for other users of the same data.

--In this project some consideration has been given to methods of introducing "filters" in conjunction with special models in order to show special relationships between entities in a manner significant to other types of user.

Some of the needs of users not immersed in the Western cultural perspective have also been considered.

6.10. The notation used to indicate the position of an entity in a classification scheme may be very complex. This may make data handling very difficult.

--In this project it is not necessary to use a notation in order to file the entity. Only a simple sequence number is required. To indicate its position in a given model, cross-references are inserted which again take the form of simple sequence numbers. In some models the entity may be defined by its relationship context rather that by any special notation which users are free to add. A standard notation for use in print, but independent of the organization of the system, is suggested in Appendix A8.

6.11. The system may be viewed as a "one-shot" job using all the appropriate specialists. This is the case with some conceptual directories. Even so, non-participants criticize the position taken by the participants, thus suggesting the need for new projects.

--In this project it is not necessary to limit classification to the views of one specialist. A number of competing specialists can participate together or separately without jeopardizing the ability of the system to adapt and respond to new proposals.

6.12. Systems may be slow (up to decades) in responding to proposals for change, to the point of acting as a constraint on innovation to those dependent upon them.

--In this project, modifications and alternatives can be handled without difficulty.

6.13. A system proposal may raise problems of standardization for purposes of handling bibliographical or other data. The system design may then become a pawn in the debate between the different schools of standardization and information handling.

--In this project there are no features which could become a major issue in the ongoing debate, since it is not a conventional documentation system and does not have major bibliographical concerns.

6.14. A system proposal may become a pawn in the debate between different shcools of classification.

--In this project alternative classification systems may be handled.

6.15. A system proposal may constitute a threat to other systems competing for the same resources -- particularly if major changes are proposed for existing systems.

--This project does not appear to compete with other systems. It can be considered complementary to some documentation systems.

6.16. A system may demand, or be designed for, complex computer systems to the point of being unusable in less-richly-endowed environments.

--This project is based on a very simple filing system for entities and relationships between them. The resulting file may however then be subjected to analyses of varying power depending on the computer environment available.

16.17. A system design may raise fundamental theoretical issues, and therefore alienate important potential supporters.

--In this project the accent is on providing a simple technique for filing entities and relationships in a way which permits a number of general analytical and display techniques to be used. Every effort has been made to avoid giving a final and exclusive definition of what is incorporated. Such theoretical debates are carefully confined to the activities of modelling groups which are each free to ignore or accept entities and relationships filed by other modelling groups.

6.18. A system may raise difficulties concerning the status of the entities handled as "knowledge" or in relation to language and semantics.

--The description of this project -- but not its operation-- is paradoxically subject to many of the terminological problems it is intended to solve. An attempt has therefore been made to discuss semantic fields (Appendix E1), problems of language and translation (Appendix E2), the status of a discipline as a language (Appendix E3), the relationship between knowledge, and information (Appendix E4), and knowledge as a evolving structure (Appendix E5). The problems of natural language information processing have been avoided.

7. Future significance.

This report attempts to lay great stress on the distinction between a document-oriented information system and what has been termed a knowledge-representation system. This project is considered to be a step towards more effective knowledge-representation. To clarify the distinction even further and to show the possible future significance of this effort, an attempt has been made in Appendix F1 to compare an ideal document system with an ideal knowledge-representation system.

8. Next Step and Funding Required.

The phasing of the project is discussed in Appendix A1. The next step is to obtain critical comments on the various proposals but forward and to undertake pilot projects in some of the following areas:

--file organization and computer program development or adaptation
--operational and logical problems of classification with models in a few test areas
--computer simulation of file movement, modelling activity

and behavioral complications in a decentralized, minimum organization environment
--computer simulation of different strategies to keep the system "open" without it becoming uncontrollable
--preparation of a graphics demonstration program as a means of generating further interest and showing the power of this technique.

Exactly how much pilot project activity is required will depend upon the speed with which it is desired that the project as a whole should move forward and the range of interests it is desired that the project should serve. These must be decided.

No comments have been made on the funding required since cost estimation depends on decisions taken for the next stage. The computer programs envisaged for the filing of entities and relationships and generation of lists and thesauri are however fairly simple to prepare and cheap to run. The other major costs would be collection of conceptual entities (unless done voluntarily by a team using existing material), administration (unless incorporated within the budget of some existing institute) and travel costs of those concerned with modelling (unless it was decided to switch immediately to the postal modelling concept outlined in Appendix A1).

A list of organizations, mainly international, through which further support might be obtained for this project is given in Appendix D10.

Organization of Project.

The success of a project of this type would be dependent upon the extent to which any central organization can be avoided in favor of a process of catalysis. There is too much to be done to run the risk of jurisdictional, behavioural, and personality problems associated with a central organization. These could rapidly alienate potential support. The problem is therefore to bring into existence a decentralized network of groups working on different aspects of the project, but able to exchange the results of their activities without difficulty.

(It is important to remember that it is probably impossible to "organize a whole area of knowledge because the latter is well subdivided into territories and "stamping grounds" whose incumbents are reasonably content with the current situation. It may, however, be possible to offer them a neutral device by which they can each facilitate and order their own particular approach, and, as a by-product, see more clearly its relationship to that in other "neighboring" territories. Having by this means obtained a decentralized picture of the current situation, it is then possible to lobby the incumbents into participating to some degree in inter-territory efforts at organizing areas of knowledge whilst guaranteeing safeguards for the protection of their "sovereignty".)

A. Launching Phases.

A number of Phases can be envisaged, some of which could overlap.

1. Investigation. During this Phase the project would be investigated in detail by circulating proposals among appropriate specialists. The main object would be to ensure that the proposal is oriented in the right direction, and that funds for pilot projects are obtained. This Phase may be considered to be underway already, through the actions of the COCTA committee.

2. Pilot Projects. During this Phase, efforts would be devoted to the following areas:

   a) computer program development and file organization
   b) operational and logical problems of classification with models in a few test areas.
   c) computer simulation of file movement and modelling activity in a decentralized, minimum-organization environment.
      It would be particularly valuable to gain some insight into the behavioural problems of rivalry and suspicion between model building groups, and efforts to "take over" the system.
   d) computer simulation of different strategies to keep the system "open" to theoretical formulations from as wide a range of sources as possible whilst trying to minimize the inclusion and retention of formulations of dubious value.

3. Agreement on Standard Formats. On the basis of the previous Phase, standard formats for filing new formulations and for holding them on magnetic media would be agreed. Since this is a new type of project, it should not encounter the apparently insurmountable difficulties of those concerned with organizing the computerized exchange of bibliographical information.

4. Production of Standard Solution. Once agreement has been reached, a standard software computer program can be made available to all those bodies which wish to initiate some concept modelling activity, or to act as a central filing point for their particular constituency. It is possible that initially only one body will be active, possibly as an extension of the pilot project stage.

5. Filing Procedure. Once a standard filing form is developed, there should be no difficulty for any group in receiving and filing theoretical formulations. This can of course be done by mail.
   By filing is meant the purely administrative activity of preparing the forms for the computer. There should be a minimum of judgemental effort at this stage, and none with respect to the theoretical problems of the subsequent modelling activity. The object is to get the incoming information into a form which facilitates the activities of the members of the modelling bodies.

   The area of difficulty which does require examination is that of how to decide who should not be permitted to submit concepts for filing into the common data base. This point is considered below.

B. Periodic Operations.

1. Lists of Formulations. Periodically the sequence of theoretical formulations held on magnetic tape should be scanned to produce lists for circulation to the modelling bodies and, if required, their members. Two types of lists can be envisaged.

   a) lists of newly-registered formulations which must be scanned by each modelling body to see whether they are in any way relevent to its concerns
   b) lists of the complete sequence of formulations for newly formed modelling bodies wishing to re-examine all possible formulations and interrelate them in their own way.

2. Modelling or classification. The lists derived from the previous operation can be examined by the modelling bodies in committee or distributed by post to their members. From these (postal) deliberations should emerge

a collective opinion on the place within the classification scheme, of each theoretical formulation reviewed. If necessary, a "provisional" view can be formulated by the use of appropriate coding. In fact this might be a most useful way of submitting a committee's view for wider consideration. Different degrees of "definitiveness" could thus be envisaged.

3. Feedback of Model Information. The details of the place of the formulation within a particular model would be indicated on a standard form which could be returned by post for keypunching and incorporation. A modification of this approach would be to permit individual committee members to each return forms for any new entity under consideration. In this way all the alternatives would be incorporated into the model with some "provisional" code so that each member could see the proposals of the others, and their implications. In some cases, this could even be operated as a means of postal voting on the treatment of controversial formulations. The administrative load of the committee is in this way largely computerized.

4. Input of Model Information. The forms from each modelling body would be handled at the central registery point (for that constituency), keypunched and fed onto the magnetic tape file. Keypunching errors would be corrected there as far as possible.

5. Production of Model Amendment Lists. Whenever required, the formulations incorporated into a given model would be selected and sorted into the thesaurus-type structure appropriate to the model and listed for distribution back to the members of the modelling body. This gives members an updated model with all the formulations coded to different levels of "provisionality".

   Members can then reconsider their views and proceed from Operation 2 above or, alternatively, for those formulations which have been classified to the agreement of all concerned, the term allocation operation may be initiated.

6. Allocation of Model Terms. Working from the formulations structured into a thesaurus-type order, members can allocate terms to each entity in English and whatever other languages are considered necessary. Again, there is no reason why "provisional" coding should not be used to cover various working cycles of term allocation.

7. Feedback of Term Information. As with model information, the alphanumeric terms allocated to each formulation can be indicated on a standard form which could be returned by post for keypunching and incorporation onto magnetic tape.

8. Input of Term Information. The forms from each body allocating terms within a model would be handled at a central registery point, as with the model information itself.

9. Production of Term Lists. Whenever required, the formulations incorporated into the model would be selected and sorted into term lists, either in alphanumeric order or in terms of a thesaurus-type structure. This gives members an updated model expressed in terms coded to different levels of "provisionality".

   Members can then reconsider their views and proceed from Operation 6 above.

It is clear that the above operations permit a quite extensive degree of "de-committeefication". Members of a modelling body can individually register their views and preferences by post on each formulation in the model and in their own time. The resulting lists are circulated and amended to firm up progressively the consensus on each point until final agreement can be reached. Alternatively, if this is a final difference of opinion, then this can be registered as such. Actual discussion need only take place when the accumulation of cases (which cannot be handled by correspondence and a "modelling bulletin" mechanism) merits such contact.

C. Subsequent Phases.

A number of Phases can be envisaged which follow on from those detailed in "A" above. They do not, however, modify the basic operations noted in "B".

1. File Movement. One of the disadvantages of isolated registration points is that formulations common to two or more constituencies will not necessarily be juxtapositioned. In particular, unless each such point is allocated a block of sequence numbers, there are liable to be overlapping sequence numbering systems which would jeopardize the whole project.

   One means of avoiding this, aside from allocating blocks of numbers to each registering point, is to circulate copies of the files between registration points. (Either the tapes themselves could be moved, or data links could be used.) This might be considered a standard procedure by which duplicates in all newly-added formulations could be located and grouped together for consideration by each of the interested modelling bodies, prior to arriving at a "final" decision.

   The circulation of such information can be made very rapid. A courier file can be circulated between the registration points for a particular discipline. Information is copied onto and off each such sub-specialty file. At one point in its movement, such an intra-discipline file could interact with an inter-discipline file (e.g., for disciplines in the same group) to permit a similar two-way transfer to take place. Similarly a higher level courier file moving between groups of disciplines could permit further exchange.

In this way cross-disciplinary confusion could be avoided. Clearly refinements are possible by using mission-oriented files or geographical area files. The system is very flexible. It could even be made to interact with "classified" files by using security, subject matter and evaluation filters to govern the interaction.

The key feature is that it does not require more than a bare minimum of overall organization or funding. It can be extended very loosely in response to the initiative of any highly specialized discipline. Registration points are created wherever (in terms of subject, jurisdiction or geographical level) there is sufficient common interest -- i.e., motivation plus resources. This gets around the current situation in which vain attempts are made to get significant funding for general multi-purpose projects, particularly via any international program.

If cross-jurisdictional problems arise in particular areas, all the administrative work there may be delegated under contract to some party judged to be impartial and uninvolved -- a commercial computer service bureau, a university, a government agency;or a user cooperative point might be organized.

The costs involved at each collecting point are
(a) conversion of information and queries to machine-readable form
(b) processing and output relevant to immediate user contacts
(c) transport costs of the courier file to the next collecting point.

The funds are expended locally in a manner which can be immediately justified and yet this results in making available current information from points very conceptually distant within the system.

D. Accredited Sources.

It is clearly an advantage to allocate responsibility for modelling group activity in a particular domain to the appropriate international professional organization.

The difficulty arises in determining which sources of information should be recognized by such modelling groups. In the earlier phases when the group is working through the standard texts, few problems should arise. But once a model is available for inspection, problems will arise in determining whose suggestions for additions or amendments should be accepted. Within a well-defined profession this difficulty may be avoided by recognizing only accredited

members of the profession. The right to submit amendments then becomes a right accorded by the profession. This procedure will undoubtedly lead to conflict when areas common to a number of disciplines are considered (e.g., the social sciences, in general), unless each discipline is restricted to its own model.

A distinction should also be made between the right to file an entity and the right to suggest amendments to the model. There is some advantage in giving wider access for filing, but limiting the "retention period" of the entity filed according to the profesional standing of the filers.

A later development could be the possibility of retaining entities only if a supporting "vote" was registered by an appropriate number of appropriately accredited persons. The degree of support would be a "real time" measure of the degree of significance to the discipline of a given theoretical formulation.

Whatever procedure is adopted it is essential, for the vitality and general relevance of the project, that a wide range of people and organizations should be in a position to add entities to the file -- given a few simple safeguards. In this way the interests of every relevant discipline, school of thought, problem area, "approach" or paradigm should be protected. The system would therefore be "open" to social scientists writing in any language (for language problems, see Appendix C2), or taking any epistemological or ideological position.

Classification and Modelling

There is a considerable terminological variation in the scientific literature that characterizes the use of the term "classification". Dalenius and Frank, after making this observation (*) define the term as follows.

> "Consider a collective of objects of some kind and a set of mutually disjoint classes. Every object belongs to one, and only one, of these classes. By classification we will denote the act of assigning the objects into these classes.
>
> In taxonomy, classification indicates the act of creating classes according to some principle, the term "identification" is used for classification as used in this paper. By the same token, the term "coding" is rather ambiguous. We refrain from its use here, but mention that classification as used in this paper is referred to as coding in the literature dealing with e.g. population censuses.
>
> Classification is a major operation of such statistical studies as a census of population or a census of business. Thus, the units enumerated in a census of population are assigned into classes defined with respect to sex, age, status of employment, etc. Likewise, the units enumerated in a census of business are assigned into classes defined with respect to e.g. total turnover."

This definition, whilst appearing to be inclusive, in fact only covers one type of classification, namely hierarchic classification where classes are mutually disjoint. Classification of theoretical formulations is one area in which classes may or may not be mutually disjoint.

Alan J. Mayne (**) notes that J.H. Shera has made an excellent general assessment of the problems of general library classification in an article of his, originally published in 1953 and represented in his book, Libraries and the Organization of Knowledge. Mayne adapts Shera's basic requirements for traditional classification to give

1. Linearity of subject arrangement (to permit ease of location of books on library shelves)

(*) T.E. Dalenius and O. Frank. "Control of classification." Review of the International Statistical Institute, 36,3, 1968, 279-295 (includes formal description of classification and introduces various parameters useful for control purposes)

(**) Alan J. Mayne. "Some modern approaches to the classification of knowledge." The Classification Society Bulletin, 1,4,1968, 12-17.

2. Coverage of all knowledge (inclusiveness)
3. Meaningfulness of all terms
4. Meaningful differences between terms
5. Significant arrangement of terms
6. Provision of a specific place for every existent or non-existent, past, present or future, topic or field (infinite hospitality)
7. Unique arrangement of terms
8. Uniquely defined notation for each term.

Mayne then notes that the limitation of the traditional classifications are

1. Inconsistency of organization
2. Excessive complexity
3. Inability to adapt to the advance of knowledge so that these classifications fairly soon become incomplete

He concludes with Shera that the hierarchical form in itself is not a sufficient basis for the classification of knowledge and that what is required is a directed graph, or non-hierarchic representation.

As a consequence, Mayne then revises the traditional requirements for classification:

1. -is considered invalid for the classification of knowledge

2-6 -remain essential

7. -is not possible with a hierarchic classification schema but becomes possible with non-hierarchic or directed graph representations.
8. is considered impracticable for a general classification of knowledge (which permits Mayne to argue for a mnemonic system for certain specific areas)

The relationship between hierarchic and non-hierarchic classification schemas has been the subject of considerable work by Jardine and Sibson (*). They are particularly interested in the stability of the classification produced by a given method as the amount of information (or number of attributes) is increased for the entities being classified. They are looking for measures of distortion introduced by the imposition of a

---

(*) C.J. Jardine, N. Jardine, and R. Sibson. "The structure and construction of taxonomic hierarchies." Mathematical Biosciences, n° 1, 1967, 173-179.

N. Jardine and R. Sibson. "A model for taxonomy." Mathematical Biosciences, n° 2, 1968, 465-482.

N. Jardine and R. Sibson. "The construction of hierarchic and non-hierarchic classifications." Computer Journal, n° 11, 1968, 177-184.

R. Sibson. "Model for taxonomy II", Mathematical Biosciences, n° 6, 1970, 405-430.

N. Jardine and R. Sibson. Mathematical taxonomy. London, Wiley, 1971.

given classification schema.

They develop a set of conditions for evaluating cluster methods used to derive hierarchic classification schema and show that the majority of cluster methods fail to satisfy these conditions, the only exception being the "single-link" or "nearest neighbour" method. The latter can be given a simple graph-theoretical description which makes clear its defects as a method of classification (*). The defect of the single-link method is that it clusters together at a relatively low level entities linked by chains of intermediates. This defect is generally called "chaining". They note, however, that to call it a defect of the single-link method is misleading, because the graph-theoretic description makes it clear that chaining is simply a description of what the method does.

Jardine and Sibson suggest that the limitations of the single-link method are limitations of the hierarchic classification itself, and that these limitations can be overcome using a (mathematically) more general system of classification. In particular, they note that it is possible to develop numerical systems of classification that reveal the kinds of information that are concealed by chaining, for example, information about the relative coherence or homogeneity of clusters of entities.

Where clusters are permitted to overlap one would expect that as the degree of overlap is allowed to increase, so the accuracy of the representation of the entities should increase, although at increased cost of complexity. In the limiting case where arbitrary overlap is allowed an exact representation of the original data should be obtained. These intuitions are precisely expressed in the generalized model developed by Jardine and Sibson (1968), covering both hierarchic and non-hierarchic classificatory systems.

This generalization permits Jardine and Sibson to conclude that classification is a two-stage process. The first stage is the derivation of a dissimilarity coefficient based upon the distribution of states of characteristics (attributes, or properties) among the entities to be classified. The second stage is the derivation of a classification from the dissimilarity coefficient. The single-link method is then regarded as the first term in a sequence of classificatory methods giving successively more accurate but more complex representations of the entities. Within this theoretical framework it is possible to define measures of the distortion imposed by a classificatory system, and of the relative isolation and homogeneity of clusters.

(*) see: M. Wirth, G.F. Estabrook, and D.J. Rogers. "A graph theory model for systematic biology." Systematic Zoology, n° 15, 1966, p. 39.

Comment.

The above work makes it clear that classification can introduce distortion and that this can be avoided by using a directed graph representation. In this project the distinction is made between the filing process, the classification process, and the term allocation. But in the light of the above argument, the classification process could be split into two stages which correspond approximately to the two stages distinguished above.

It is useful to think of the first stage of the classification process as one of "relationship indication", in which the relationships of a given theoretical entity with other entities are inserted. This results in a (directed graph) network of entities which can be searched by computer, particularly to detect clusters with certain properties. This stage corresponds to the determination of similarity or dissimilarity between entities.

In a second stage, the above network can be distorted so that its elements can be fitted into a chosen set of classes with a certain relationship to one another. This is "classification" as opposed to the previous phase which inserts relationships irrespective of any class boundaries. It is convenient to call this activity "modelling". Clearly the modelling activity is a valuable preliminary to "classification". It is particularly valuable in that once completed, different systems of classification can be compared using the entities inter-related by the model, i.e., different degrees of distortion can be imposed upon the network of entities according to the immediate needs of the user. It may be useful to think of modelling in this context as a long-term multi-person activity, whereas a given classification can be selected from the modelled entities in terms of short-term, need-oriented considerations which permit certain relationships in the network to be considered as "irrelevant" -- permitting the isolation of simple, possibly hierarchic, classification schemas. In some cases, it may however be preferred not to distinguish modelling from classification and to blur the two operations into one another.

Control of Classification.

There is ample evidence, according to Dalenius and Frank (see above), that the classification operation may be rather susceptible to errors: objects are not assigned into the proper classes. As a consequence, there is need for control. As a basis for a control of the classification operation, two kinds of schemas for verification have found extensive uses: schemes for dependent and schemes for independent verification, respectively.(*)

(*) For this distinction, see M.H. Hansen et.al. Quality control in the 1960 census. American Society for Quality Control, Proceedings of 1962 Middle Atlantic Conference, Milwaukee, Wisconsin, 1962.

An example of a scheme for dependent verification may be described in the following way. The "production classifier" (that is the classifier carrying out the classification that is to be verified) assigns an object into a class by putting one or more digits on a schedule carrying the information collected for this object. The schedule is then handed over to the verifier, who inspects the outcome of the work of the production classifier, and approves or disapproves of it.

Independent verification calls for having an object assigned into a class by two or more classifiers, who operate independently of each other. The verification procedure exploits, in the one way or the other, a comparison of the outcomes of the works of the classifiers.

Dalenius and Frank have investigated a means of evaluating various alternative procedures for independent classification which might be of use in designing the procedures of modelling groups and testing their consistency. They provide a rational means of attacking the problem of whether and how one should use a small number of highly qualified classifiers, or a larger number of less qualified ones -- although they do not appear to consider how extra weight might be given to the more qualified if a mix were used.

Filing and Classification.

It is a basic feature of this proposal that there are major advantages in separating the filing of new entities from any modelling or classifying activity.

(a) In the case of census data applications, possibility of using computer methods to simplify the relationships between a team of classifiers is touched upon by Dalenius and Frank who make a very clear distinction between the information as filed under an identification record (i.e., the sequence number, in this case) together with a census record giving any verbal data (i.e., the definition and conventional label, in this case), and the class label as decided by the team of experts (i.e., the relationships determined by the modelling activity, in this case).

(b) In the case of document indexing application, no distinction is made between filing and classification. Because of this, the administrative problem of filing and the qualified expert problem of classification combine to create severe problems. Two approaches within this frame of reference are envisaged:

-- sharing the burden of classification (and filing) by transferring this function to the producers of documents
-- reaching agreement on unique, or compatible classification schemes and indexing languages

The UNISIST (**) Study noted that little progress can yet be reported in the way of indexing-at-source and that a serious limiting factor to any form of cooperative indexing is the range of acceptability of the proposed indexes. Even the all-embracing and widely used U.D.C. has adversaries. The Study also noted that it is unlikely that the concept of a universal scheme will even make any practical sense in the realm of deep content analysis (p.46). The reasons are the observed differences in the semantic basis of indexing languages which are the consequence of well-founded differences in outlook and interests on the part of a highly-diversified community of users.

All that can be looked for, according to the UNISIST Study,"is the existence of semantic relations between the different lexical sets (be they called classifications, lists of disc thesauri, automatic dictionaries for converting natural language into information language, etc.). The study of these relations is the subject of ongoing research on the "compatability" of indexing vocabularies .... the subject is now receiving much attention, as an essential part of projects aimed at establishing world-wide interconnections between information systems." (p.46)

It would appear from this, that the distinction between the impracticalities of classification and the practicalities of "relationship identification" (i.e., modelling) is becoming established. But the filing or administrative aspect of "entity capture" is now blurred into the modelling phase. There is as yet no suggestion that work on "compatibility" would be considerably facilitated if similar filing techniques were used prior to the activity at the modelling level at which the "well-founded" theoretical differences arise. Standardization is possible, but at a lower level consistent with user requirements. Until this is realized the relationship between lexical sets cannot be handled systematically by computer methods.

Advantages of Numerical System.

The three major advantages of a sequential, non-significant numbering system for entities are

--facilitation of administrative activity by removing the burden of requiring that the file number receive the "imprimateur" of an overloaded qualified expert
--preparation of the basis for a proper semantic analysis by avoiding "the difficulty encountered in manipulating semantic reality without the assistance of a corresponding concrete reality" (*) and permitting

(*) A. Martinet. "Arbitraire linguistique et double articulation." Cahiers Fernand de Saussure, 15, 1957, 107 (cited by Georges Mounin, Les problèmes théoretiques de la traduction, Paris, Gallimard, 1963, p.122-123).

(**)See Appendix D6.

"semantic facts to be treated independently of their formal (linguistic) supports " (*).

--admission of "artificial" theoretical entities (new concepts, groupings of other concepts) for which no simple term exists or for which a questionable neologism would have to be invented. This is difficult in the case of term oriented systems.

(*) A. Martinet. op.cit.

Types of Model

It is important to keep in mind the many possible uses of the proposed computer-based filing system. Concentration on one set of uses may not necessarily keep the system alive either in terms of funding or value to current research activity. Multiple demands on it would ensure multiplicity of fund sources and many bodies willing to feed in entities and assist in different aspects of the coding.

The following types of model are an illustration of the possible lines of development. The list does not pretend to be exclusive so that other kinds of model could be included. As attempt has been made to group the models into types which in some cases might usefully be treated on the same occasion by the responsible modelling group.

It is important to note that the models are not only simple hierarchies but can also be networks of relationships in cases where categories overlap or one entity can be a component of several other entities (see Appendix A2).

Group I. Current structures.
This is a poor title but refers to all the current and new structures and relationships as made up of:

a. Compositional models
These models would be primarily concerned with the manner in which entities are nested within one another to form hierarchies. Six types of relationship are possible here in three sets of two.

1. Meta-level: reference numbers of all entities of which this entity is a component.

   (This relationship could be split into two sub-types as the computer-level data formats for other types of model require such a split.)

   Examples are: theories in which this concept is used, general class of concepts to which this concept belongs, general problems of which this problem is a part, organizations of which this organizational unit is a member

2. Sub-level: reference numbers of all entities which are components of this entity. (This relationship could be split into two sub-types for the same reasons as above.)

   Examples are: concepts used in this theory, concepts which belong to this class of concepts, properties or attributes of this concept, sub-problems of this problem, organizational units which are members of this organization, etc.

3. Associated reference numbers of all relevent entries which have a horizontal relationship to this entity.

   a. See - also entities, namely those which should also be borne in mind when considering this entity.

      Examples are: cases of insufficient terminological precision.

   b. Use - instead entities, namely those which should be substituted for this entity.

      Examples are: cases where an entity is outmoded for that model.

b. Behavioural models

At the same time that the modelling activity is undertaken on the compositional relationship in Ia, it should be necessary to consider some non-compositional relationships to other entities. Six such types are distinguished in three sets of two.

1. Meta-level: reference numbers of all more-inclusive entities whose value is potentially modified by this entity.

   a. Strengthened entities, namely those more inclusive entities which are perceived by the modelling group to be in some way legitimated or reinforced by the presence of this entity.

      Examples are: theories which are indirectly dependent upon the validity of this concept, problems which are indirectly aggravated by the presence of this problem.

   b. Weakened entities, namely those more inclusive entities which are perceived by the modelling group to be in some way undermined or threatened by the presence of this entity.

      Examples are: theories which are indirectly undermined by the validity of this concept, organizations whose monopoly is weakened by the presence of this organization.

2. Sub-level: reference numbers of all lower level entities whose value is potentially modified by this entity

   a. Strengthened entities (see above)

   b. Weakened entities (see above)

3. Associated: reference numbers of all entities in other disciplines whose value is potentially modified by this entity.

   a. Strengthened entities (see above)

   b. Weakened entities (see above)

      Examples are: a purely economic development theory that is undermined by a social development consideration.

Group II. Contextual structures

Again this is a poor title but refers to the historical and comprehensional relationships which constitute a context for the Group I current situation, and would be used in learning about the Group I situation.

a. Educational models

These models would be produced by those modelling groups primarily concerned with education and making more sophisticated concepts comprehensible. The concept of levels here is therefore associated with "learning order". Six types of relationship are again distinguished in three sets of two.

1. Meta-level: reference numbers of the entities representing more sophisticated, complex or difficult to comprehend versions of the definition represented by this entity.

   a. Correct entities, namely those definitions which are valid, and useful complexifications.

      Examples: the chain from the "billiard ball" concept of an atom, through a "miniature solar system", through Bohr-Summerfeld orbital model, then to a symbolic representation of the "electron cloud" in quantum mechanics.

   b. Incorrect entities, namely those definitional complexifications which are useful educationally as an illustration of less valid or less useful complexifications (i.e., how not to conceive of it)

2. Sub-level: reference numbers of the entities representing simpler and easier to comprehend versions of the definitions represented by this entity.

   a. Correct entities, namely those definitions which are valid and useful simplifications

   b. Incorrect entities, namely those definitions which are less valid and useful simplifications.

3. Analogies: reference numbers of entities representing analogies which may clarify understanding about this entity.

   a. Correct entities, namely those analogies which are valid and useful

   b. Incorrect entities, namely those analogies which are less valid and useful

b. Historical models

These models would be produced by those modelling groups interested either in historical research on the history of ideas or in providing an historical framework to assist education. It is probable that the educational and historical models whould be considered together, which is why they have been grouped. Six types of relationship are again distinguished in three sets of two.

1. Meta-level: reference numbers of the entities which succeeded or replaced this entity in the history of the evolution of the discipline (organization, problem, etc.)

   a. Correct entities, namely those definitions which were valid developments from that represented by this entity.

      Examples are: the chain of concepts of the structure of the solar system.

   b. Incorrect entities, namely those definitions which were invalid and fruitless developments from that represented by this entity.

2. Sub-level: reference numbers of the entities which preceeded or were replaced by this entity in the history of the evolution of the discipline (organization, problem, etc.)

   a. Correct entities, namely those definitions from which this entity developed directly.

   b. Incorrect entities, namely those "competing" definitions in the same conceptual milieu which did not constitute a step in the formulation of this entity.

3. Associated: reference numbers of entities contemporary with this one but insulated from it within another culture or school of thought, such that it did not then affect the history of the evolution of the discipline (organization, problem, etc.), namely a parallel historical evolution.

   a. Correct entities, namely those which corresponded closely to the definition represented by this entity.

   b. Incorrect entities, namely those which contradicted or undermined the definition represented by this entity.

Group III. Real world systemic relationships

The previous groups of models deal with the relationship between conceptual entities in anthropocentric terms or within the logic of particular disciplines. There is little concern for the effects of the presence of one conceptual entity on another in the ecosystem of ideas, but especially the effects on one another, across disciplinary lines, of components of real world systems.

The best example of this distinction is the interdisciplinary nature of environmental problems, when for example, it is the real world interaction of chemicals in food chains which cause egg shells to become thin -- leading to high chick mortality rate of some bird species. For a social example, the relationship shown between the entities, represented by boxes in Figure 1, give a schematic representation of the factors binding a Canadian Indian to a pattern of problems.

The same approach as above may be used to handle models of such real world systems.

1. Meta-level: reference numbers of entities representing systems of which this system is a subsystem.

   a. Positive entities, namely those systems to which this sub-system contributes in some positive manner.

   b. Negative entities, namely those systems to which this sub-system contributes in some negative manner.

2. Sub-level: reference numbers of entities representing systems which are sub-systems of this system.

   a. Positive entities, namely those sub-systems which contribute to this system in a positive manner.

   b. Negative entities, namely those sub-systems which contribute to this system in a negative manner.

3. Associated: reference numbers of entities representing systems which are part of the same system (but unrelated to this entity in any of the above ways).

   a. Positive entities, namely those sub-systems which contribute to this system in a positive manner.

   b. Negative entities, namely those sub-systems which contribute to the system in a negative manner.

Group IV. Term-oriented models.

In some cases where classification is rudimentary or non-existant, the emphasis is placed immediately on the terms. This is the case when:

(i) official terms are used and the definitions are conventional or undefined as in many library or descriptor lists. The entity is defined by the term.

(ii) a particular official definition exists for a particular term as in official dictionaries (e.g. the Larousse Litré as reflecting the decisions of the Académie Française)

(iii) terms are related in a thesaurus without definitions (e.g. as in Roget's Thesaurus). Such thesauri may have many levels of classification.

There is no reason why each such set of terms should not be treated as a model as in the other groups. Where appropriate, the classification code position would be omitted and only the term positions used.

Group V. Administrative models.

The assumption made in discussing the earlier groups of models was that the model was in some way a definitive structure on which new work would build. It is however possible to use the model building code

to facilitate the administrative work on the definitive model.

(i) Preliminary models

Each modelling group can use one (restricted) model to store features of the model which are not yet finalized. Or else several preliminary models could be experimented with.

(ii) Entity handling models

When entities are first registered it may be useful to use one or more rough classification models to ensure that the entity, if inadequately coded, is drawn to the attention of the modelling groups liable to be interested in it.

Group VI. Mission-oriented models.

An assumption made in earlier groups is that the modelling bodies would all be discipline-oriented. There is however no reason why mission-oriented models should not be used where appropriate (e.g. in connection with development, environmental problems, etc.)

Group VII. Interdisciplinary models.

Clearly it is most important to avoid a "babel of models". A second level operation of model reconciliation to form a set of interdisciplinary or inter-model models could therefore be instituted when required.

These could either (i) be constructed (automatically by computer) from all the entities common to the models from which it is desired to produce an interdisciplinary model, or (ii) be constructed by selection based on judgement of the best from each.

Group VIII. Future-oriented models.

A final assumption made in dealing with the earlier groups was that only the current or historical situations would be modelled. There is however no reason why speculative models should not be produced showing the relationships between entities at different points in the future. The modelling activity might then in some ways represent the Delphi method of forecasting.

Group IX. Personal models.

Perhaps a long term ideal is for a person to be able to "look at" (or interfere) with the basic list of entities in terms of his own model which is his personal "thought file". Each new idea he gets could be usefully reflected in the structure of this file.

It would not of course correspond exactly to the models for his discipline but the precise points of difference could be established by computer analysis.

Clearly such personal models allow the person to think in the categories most meaningful to him with the labels he finds most mnemonic. The translation into the terms of any other model can be done automatically whenever required. Such models would not be widely circulated but might be very useful if held on the file of a particular faculty of a university.

Group X. Sub-models.

In some cases a particular sub-branch of knowledge may be fragmented by reinterpretation, reconceptualization and redefinition of the same entities. It is then appropriate to use a "sub-modelling" strategy. In other words, instead of requiring "dissident" groups to conform or to divert their energies into a parallel model with differences in a minor area, a sub-model could be used to redefine that area in the dissident group's terms. The sub-model would therefore offer an alternative interpretation.

Group XI. Languages as sub-models.

It may be convenient, for some purposes, to consider the relationships between theoretical formulations used in a particular language as a sub-model. The differences between the concepts encountered in Indo-European languages are relatively minor, so that term equivalents pose no great problems, but should it be necessary to enrich the system by incorporating theoretical formulations from other language groups, problems could arise. (see Appendix E2).

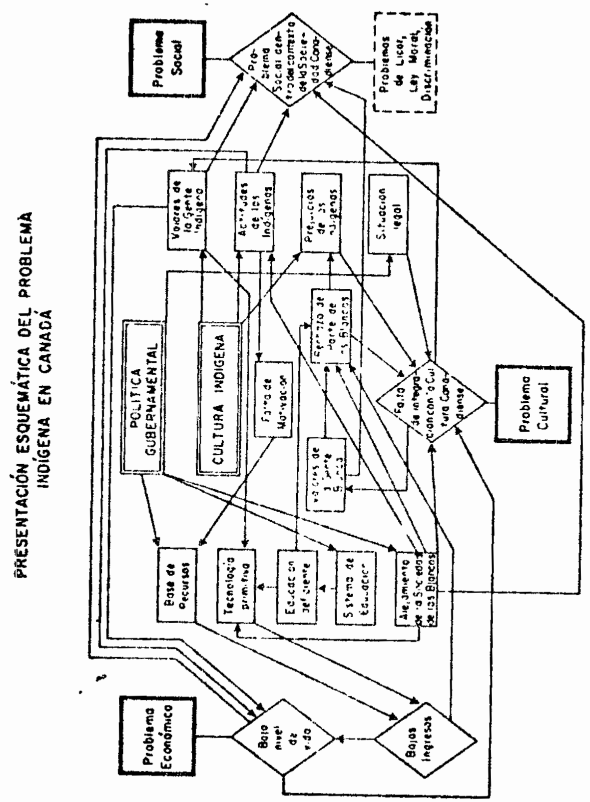

Figure . Illustration of cybernetic interactions

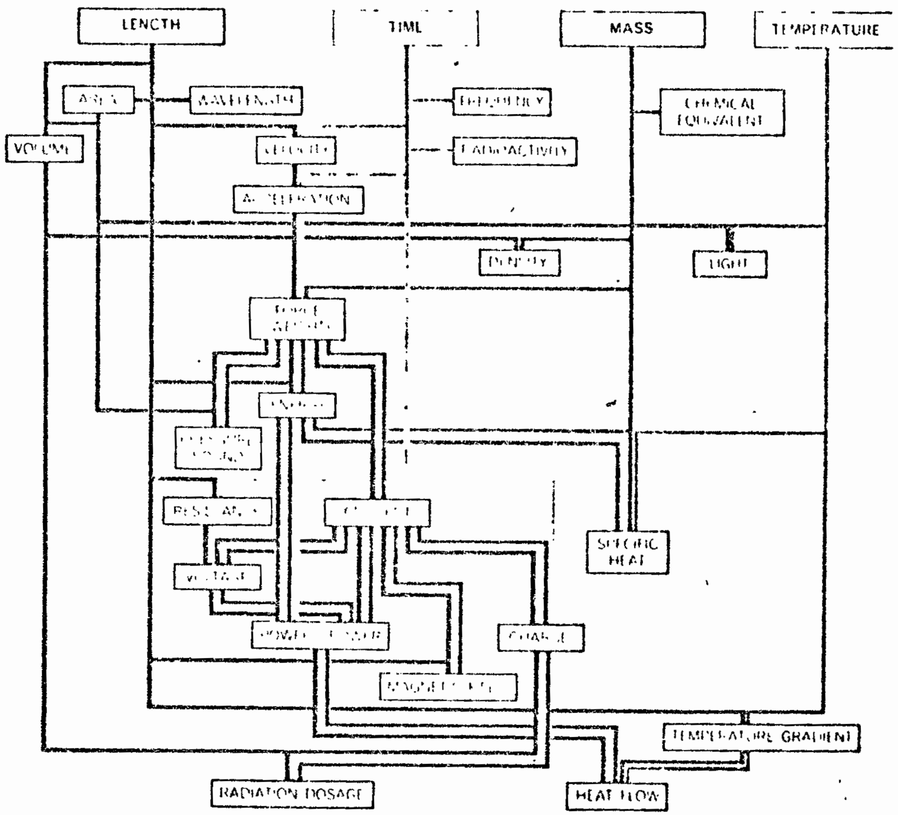

Figure . Interrelationship of measuring standards

Types of Entity Included.

There is a very varied terminology currently is use to characterize theoretical products. Gunnar Sjöblom notes the use of conceptual (analytical, theoretical) frameworks, analytical schemes, paradigms, orientations, frameworks for inquiry, theory sketches, pre-theories,etc.(*)The same is true for the components of the scientific process: problems, observations, empirical generalizations, models, derived propositions, hypotheses, theories, etc. It is unlikely that any immediate agreement could be achieved or a standard terminology, even if this was in fact beneficient.

Each of the conceptual constructs represented by the above terms may be treated as an "entity" which could be incorporated into a computer file. Once incorporated, efforts could be made to attach an appropriate distinguishing code to them within the framework of a given model. It is highly probable, for example, that under different models the same entity may be coded differently, or alternatively that distinctions important within one model will be insignificant in another (e.g., theory and model; hypothesis and proposition).

The same system could well handle other types of entity at any time if required, such as: organizations, problems, schools of thought, theoretical viewpoints, individual authors. With respect to real world substantive problems, which might be considered as irrelevant in this context, it is interesting to note one author's comment:

> "The fact is that most of the problems (in society) that we stand ready to consider are bogus problems. They are generated by theories about technological progress, and theories about the way society works. Theory is often the only reality countenanced by our culture."
> (Stafford Beer. "Managing modern complexity." In: The Management of Information and Knowledge; a compilation of papers prepared for the eleventh meeting of the Panel on Science and Technology, Committee on Science and Astronautics, U.S. House of Representatives, 1970, p.43)

As a summary,the above entities are numbered below to facilitate discussion on possible groups of entities:

A. Concepts

B. Meta-concepts

1. theories
2. propositions
3. hypotheses
4. models
5. analyses
6. conceptual frameworks
7. analytical schemes
8. theory sketches

(*) G. Sjöblom. Theoretical testing of approaches in political science (Paper presented at a conference of the International Studies Association, Bellagio, 1971)

C. General
   1. paradigms
   2. viewpoints
   3. schools of thought

D. Assumptions
   1. assumptions
   2. criteria
   3. values

E. Methods

F. Problems
   1. substantive
   2. methodological
   3. problem formulation

G. Hierarchies
   1. taxonomy
   2. typology
   3. classification

H. Operationalization
   1. indicators
   2. indexes

I. Data
   1. bodies of data
   2. interpretations of data
   3. observations

J. Social
   1. organizations

There is some advantage in a two-level coding here, because it might be possible to arrive more easily at agreement on the more general level coding, even if there are differences between models on the coding within that level. There is of course the possibility that within a particular model the grouping would be done differently, in which case the coding scheme would be peculiar to that model.

Types of Relationship Included

It is not the intention of this project to set up a single rigid classification of permissible relationships between entities. Just as no effort was made to limit the types of entities that could be handled (Appendix A4), it should not be necessary to make the futile attempt to resolve the intellectual problem of how many types of relationship are significant. That the attempt would be futile on the part of any one group is shown by Eric de Grolier's excellent chapters on the expression of relationships in generalized and specialized coding systems, in natural languages, and in experimental languages (*). He concludes, in his UNESCO/FID supported study, that it proved impossible to produce a systematization that was "sufficiently satisfactory to warrant even preliminary publication".

This conclusion should not however lead to a decision to adopt some hypothetical "best existing scheme" or to the formulation of a new scheme. It should be recognized that the project should be capable of handling as many different schemes as possible. In fact the evolution of knowledge (Appendix C5) is partly represented by attempts to produce new schemes of relationship and categorization.

Without recommending any particular scheme it is useful to attempt to list out some of the relationships to give an idea of the variety that has been envisaged.

1. De Grolier (**)
   A suggested clarification of the sign ":" in the UDC (rejected by the FID Central Committee on Classification for the UDC) consists of the following relationships:

   1.1. Appurtenance (belonging)
      11 Inclusion, implication
      12 Parts, organs
      13 Components, constituents
      14 Properties, attributes
      141 " " physical
      142 " " chemical
      143 " " biological
      15 Aptitudes, predispositions

(*) Eric de Grolier. A Study of General Categories applicable to Classification and Coding in Documentation, Paris, UNESCO, 1963, p. 17-60, 61-142, 143-158.

(**)Relationship sets 1-4 were obtained from de Grolier, op. cit.

1.2. Process
21 Action: acting on (subject), affected by (object)
211 Favourable (stimulation; increase)
212 Unfavourable
2121 Delay
2122 Inhibition
2123 Destruction
21 Interaction
211 Favourable (symbiosis)
212 Unfavourable (antagonism, competition)
22 Operation, means used: process (subject),product, result (object)

1.3. Dependence
3 Causality, origin, etc.
31 Causality; cause (subject), effect (object)
32 Origin: originating (subject), arising from (object)
33 Conditioning, requirement: conditioning (subject), conditioned (object)
3 Interdependence
31 Correlation
32 Association
33 Combination, synthesis

1.4. Orientation
41 Aspect, particular case
42 Application
43 Use

1.5. Comparison
51 Resemblance, likeness, similarity
511 Analogy
512 Equality, identity
52 Dissimilarity, unlikeness
521 Difference
522 Opposition (of character)

The negation of a relation may be represented either by putting a zero before the number used, or by putting over it the sign used for this purpose in logistics, the dash.

2. Gardin
Four relationships are enough to record the chief situations of a term or class in a dictionary in its relationship to any other term or class:

2.1. Predicative, attaching to a term indicating an element or entity practically autonomous as an object of study, an essentially dependent property which describes its state, quality or function, i.e. predicate.

2.2. Consecutive, or, of causality, finality, etc., joining two elements of which the presence or action of one affects the presence, state or status of the other. This relationship can assume various meanings, according to the nature of the two elements concerned: the opposition of active and passive, a genetic relation, causal relation, conditional relation, functional or factorial relation, bond of finality, mediatory relation.

2.3. Associative, defined, in opposition to the consecutive relationship, as joining two elements of which the mention of one implies the simultaneous mention of the other. This relationship also assumes different meanings according to the the context: relation of the part to the whole, of place, of hierarchy, of appurtenance, of specification, semiological relation. It is parallel to the predicative relationship; the difference being that in this an element is considered as dependent ('predicate'), whereas the associative relationship joins two elements both considered -- in the dictionary -- as being independent.

2.4. Comparative, indicating an extrinsic bond between two elements (independent, as for the associative relationship) of the dictionary: the author weighing up any two characteristics, in order to differentiate them as to their nature or their function in the same context.

3. Farradane
A set of nine relationships obtained by correlating two series of three characteristics derived from the psychology literature:

3.1. Concurrent, Non-time: co-presence of two otherwise unrelated concepts.

3.2. Concurrent, Temporary: comparison, or relation agent/activity.

3.3. Concurrent, Permanent: association, also subjective properties.

3.4. Non-distinct, Non-time: equivalence, synonymity.

3.5. Non-distinct, Temporary:dimensional relation; properties derived from environment.

3.6. Non-distinct, Permanent:belonging to; genitive.

3.7. Distinct, Non-time: distinction from, or substitution for, imitation.

3.8. Distinct, Temporary: action upon

3.9. Distinct, Permanent: causation or functional dependence.

4. Perry and Kent

A system of ten analytical relationships:

4.1. Categoric: A is a member of the class B

4.2. Intrinsic: A is composed of B

4.3. Inclusive: A is a component of B

4.4. Aggregate: A groups (is made up of) several members of class B

4.5. Productive: A produces the object B or is used for the action B

4.6. Affected: A makes a use of, is determined or influenced by B

4.7. Instrumental: A is produced by, acts upon, or upon which B acts.

4.8. Negative: A is characterized in an important manner by the absence of B.

4.9. Attributive: A possesses B as one of its most important characteristics.

4.10. Simulative: A has certain properties of, but is not B.

5. Juilland (*)

5.1. Relations of occurrence (**)

5.1.1. Functional or part/whole relations; each part may be characterized by the relations of occurrence it contracts with similar wholes.

5.1.2. Distributional or part/part relations; each part may be characterized by the relations of occurrence it contracts with similar parts of similar wholes.

5.2. Relations of constituency

5.2.1. Analytic; relations between the smaller entities of a lower order and the larger entities of a higher order which contain them i.e. from constituent to constitute.

(*)A. Juilland. An outline of a general theory of structural relations. The Hague, Mouton, 1961.

(**)"In regard to relations of occurrence, the main weakness of many modern studies is due to relying too heavily on the more specific part/part relations, to the practical exclusion of the more general part/whole relations. A truly scientific model, capable of satisfying the dual requirement of specificity and generality, must combine both types of relations: its functional roots fulfill the condition of generality required in comparative and typologic investigations, its distributional roots the condition of specificity required in the analysis of particular structures."

5.2.2. Synthetic; relations which obtain between the larger entities of a higher order and the smaller entities of a lower order they contain i.e. from constitute to constituent

Though often coextensive with, they should be distinguished from functional or intra-level relations, since they are inter-level relations which hold between entities of different orders.

5.3. Relations of presupposition

5.3.1. Syntagmatic; essentially relations of agreement which hold between different invariants in a group.

5.3.2. Paradigmatic; relations of presupposition that hold among variants of the same invariant or among related invariants, within paradigms of variants, or within paradigms of invariants. Relations of this order are used to assign entities to categories.

5.4. Physical relations

5.4.1. Relations of length; by "length" is understood not some strictly physical measure of length but a structural or functional length measured in number of parts per whole, or of entities per domain, or of constituents per constitute.

5.4.2. Relations of prominence; relations are also structured to some extent by any stress consistently placed upon the importance of one entity in relation to others.

5.5. Statistical relations

5.5.1. Relations of frequency; the relative frequency of occurrence of entities in a system can be used to accomplish a quasi-mechanical segmentation of entities into classes or domains. In this way a statistical definition for categories may be elaborated.

5.5.2. Relations of dispersion; these are established with reference to the coefficient of dispersion that ought to accompany each entity subject to a frequency analysis.

6. Others

Other types of relationship are noted in some of the appendices. The different types of model (Appendix A3) cover compositional, behavioural, didactic, historical, cybernetic (i.e. input/output of Appendix C5), problem oriented, etc. relationships. There is a citation relationship (Appendix D3, D7), and relationships between beliefs (Appendix C2, C3).The M.I.T. ADMINS approach is particularly interesting (Appendix B2).

Data to be Included on each Entity

A. Concept inventory (Filing, dentification or Registration Phase)

1. Entity sequence number.(*)

Each new theoretical formulation, of whatever type (see Appendix A4), receives a unique number which is the next available in a sequential list. The number therefore contains no significant digits or codes and has no meaning for classification purposes. (It may be an advantage to use the check digit technique described in Appendix D8)

For practical purposes it may be convenient to pre-allocate blocks of numbers to different filing centres whenever required. This avoids problems of duplication and speeds up administration. Where duplication does occur, this is eliminated at the modelling stage.

One advantage of this sequence number as a concept identifier is that it is not necessary to file a definition or conventional term at the same time. This is convenient if a new theoretical formulation has been tentatively conceived with known relationships to other concepts but with no clear definition or label yet. It avoids the need to coin doubtful neologisms in order to register the concept. In some cases it may even be an advantage to leave the term defined by its context of relationships, and not to bother attempting to find a suitable term. In which case the sequence number would be used as the only identifier until a suitable terminology for concepts in that domain can be elaborated more systematically.

2. Model description

2.1. Model number

The act of filing an entity is distinct from the later modelling activity. The "model number" in this case is "0". This artifice permits the definitions and the conventional terms or labels in different languages to be handled within the computer record framework as well as the modelling and term allocation activity.

(*) Paragraph numbers refer to columns in Figure 1 of the computer record layout. No attempt has been made at this preliminary stage to indicate how many character positions would be required for each zone in the record.

2.2 Sub-model number

Again, since entity filing is distinct from the later modelling activity, this zone is "free". It is therefore used to distinguish between

-- entity definitions (for which it is "0")
-- entity conventional labels or terms (for which it is "1")

2.3. Language

Since the definitions or the label may be given in several languages, a language code is used. (e.g. English "1", French "2", etc.)

2.4. Alternatives

There are bound to be cases, for a given language, in which alternatively worded definitions (with the same meaning) are put forward. Similarly, where several conventional terms or labels referring to the same entity exist, these may also have to be filed. A simple sequential code ("1", "2", etc.) is therefore used to distinguish between successive alternatives.

3. Cross-reference.

Cross-references are used during the modelling phase so that this zone is "free". It is, however, used in this phase to identify the sequence number of

-- other entities which use the same conventional labels as this entity (i.e. where the same label is used with a different meaning)

-- other entities which are defined using the same verbal definition (but for which the definition has a different meaning). This may be a low-frequency or trivial case.

4. Source code

There are several possible ways of handling information about the source of information on the entity.

1. Ignore. In a simplified system it is not necessary to include it since such information can be found in a backup card file.

2. Abbreviate. Some general code, indicating the country, the publication, or the filing group can be used.

3. Name. The name of the person, or filing organization, may be given in some abridged form (e.g. "DEUTKW" for Karl W. Deutsch)

4.4. Name and Support. In a more elaborate system, in which members of a discipline are expected to indicate any strong "support" or "opposition" to any new theoretical formulation, a "voting" technique may be envisaged (see Accredited sources in Appendix A1). This option could be confined to the "elders" of the profession -- or left open to all members of a profession. As "professional" activity, this might be restricted to the modelling phase.

A given member of the profession, if sufficiently aroused, could then file his support or opposition in the form "DEUTKW +" or "DEUTKW -".

4.5. Name and Reference. It might be thought more valuable to give not only the name but the reference to the document in which the theoretical formulation is discussed and justified.

On the question of abbreviations to document reference, one is immediately in the jungle of dispute amongst librarians, documentalists, etc. Several possibilities exist.

4.5.1. Use an extended bibliographical "standard" reference. This uses a lot of space and is mainly pleasing to librarians.

4.5.2. Use an abbreviated reference as in "Science Citation Index" (e.g. the first four letters of the first two significant words of the title, plus the year date, issue or volume number within which pagination is consecutive, and the first significant page number -- "DEUT KW -- NERV GOVE-- 1963 -- 0 -- 192")

4.5.3. Use a sequence number code. To avoid getting bogged down in documentation problems, a simple sequence number could be used for each publication:

either: i) referred to by the system (e.g. a complete sequence across all authors)
or: ii) referred to by the system for a given author (e.g. starting from zero for each new author)

A parallel "documentation" system would be required to decode the codes used in the approach but it might prove much tidier and practical in the long run (e.g."DEUTKW 509") *. The precise

---

(*) For a very useful discussion of this approach to documentation, see Jacques E.J. Halkin, "Proposal and wishes for an open structure in the communication of information." Scheduled for publication in: A.I. Mikhailov (Ed.) The Theoretical Problems of Information Retrieval Systems. (The Hague, International Federation for Documentation, 1971)

page numbers might be an additional requirement (e.g. "DEUTKW 509-192"). Again, as a "professional" activity, this might form part of the modelling phase.

5. Model descriptor.

This is not used during this phase.

6. Relationship descriptor.

This is not used during this phase.

7. Date codes.

7.1. Date first used

The date on which a theoretical formulation was first used is inserted here. If this is not supplied, the computer can automatically insert the date on which the entity was filed.

7.2. Date last used

This date is supplied as a result of general consensus by all modelling groups and is therefore not dealt with during this phase.

7.3. Retention period

It may be an advantage in this phase to tag some entities of unknown value so that they will automatically be dropped from the system after a certain period unless some contrary instruction is received in the meantime. Different retention periods can be used according to the status of the source.

8. Status.

For administrative purposes it is convenient to have a zone in which codes may be used to indicate that the entity is "under consideration", "of doubtful value", "no longer used", etc.

9. Text.

The words or text used for:

-- the conventional terms or labels
-- the definitions

would be inserted into this zone. This zone could also be used for any special comments which might be usefully added.

B. Concept coding (Modelling or Classification Phase)

Many of the zones discussed above are used in this phase but for a different purpose or in order to establish computer records distinct from those created during the earlier phase or by other modelling groups.

1. Entity sequence number.

This is repeated for each new relationship established within a model and is of course the same as that used in filing the identity in A.1.

2. Model descriptor.

2.1. Model number

As discussed elsewhere (see Appendix A1), each modelling group receives a unique number (e.g. "362") which identifies the system of relationships which are elaborated and filed, while at the same time distinguishing it (at computer level) from any other systems.

There is some argument for attaching special significance to particular digits of the model number with a view to clarifying a hierarchy of models or, at least, showing a relationship between models. In other words, at this level a U.D.C.-type approach might be used so that "political science" models are all identified by "32N" and "anthropology" models by "39N". This is probably a temptation to be resisted however, since it has some theoretical implications which are better contained within models. In which case a simple sequential list should be established from which the next available model number could be taken.

2.2. Sub-model number

This is a zone to be used by a modelling group whenever a level of dissent is encountered so that alternative sub-models within the general model can be satisfactorily handled and identified. Normally, in the absence of sub-models, this would be "0".

2.3. Language

Since the relationship between concepts is supposedly language independent, this zone should normally be "0".

There are, however, cases where relationships are identifiable in one language but absent, ridiculous, or ambiguous in a second. In such cases it may be convenient to use this zone for a form of language-dependent sub-model.

2.4. Alternatives

This zone is not used in this phase and must be "0"

(to permit identification of the term records in the next phase where it is non-zero)

3. Cross-references.

This zone supplies the main means by which the relationship of this entity to other entities is indicated for the particular model indicated in 2.1. The sequence number of the other entity is indicated here. In effect, every such "relationship" gives rise to a new computer record (see Figure 1).

The type of relationship is either implicit, because of the model used, or is described in 6 and 7.

4. Source code.

Depending on the method chosen (see A.4.1., A.4.2., A.4.3., A.4.4., and A.4.5.), the source coding would probably either be allocated during the concept filing phase with nothing in this phase, or in this phase with nothing in the previous phase. In the most sophisticated system, it might however be desirable to give:

-- source coding for the entity in the concept filing phase

-- source coding for individual relationships within a model, during the modelling phase.

Source coding during the modelling phase might be particularly helpful in the administrative work of elaborating a model, since it permits members of a modelling group, working independently and in isolation, to "vote" on the insertion or deletion of particular relationships (see A.4.5.). Such a postal vote system would be particularly helpful in clarifying with precision just what was under discussion at any point in time.

5. Model descriptor.

This zone is used to indicate which model is to be considered at the entity cross-references in 3.

In a simplified system this zone would not be required because the assumption would be made that each model was totally isolated from other models.

In a more sophisticated system however, there is need for a means of expressing relationships between parts of models. For example, it may be that in a certain domain two models are identical or that one forms a subset of the other. In such a case there is little need to duplicate all the relationships in the second model, provided cross-reference between the models is possible.

5.1. Model number

As for 2.1., but the model is only to be entered at the entity to which the cross-reference in 3 refers.

5.2. Sub-model number

As for 2.2., but again is only to be entered at the entity to which the cross-reference in 3 refers.

5.3. Language

As for 2.3., but again is only to be entered at the entity to which the cross-reference in 3 refers

5.4. Alternatives

Not used. (This zone may even be omitted entirely.)

6. Relationship descriptor

This zone is used to describe the relationship constituted by the link between this entity and that cross-referenced in 3. Two basic types of relationship descriptors may be distinguished.

6.1. Relationship descriptor A

This is used to give an indication of the relative levels of the two entities related (e.g. class and member), directions of flow (e.g. from or to), etc. These are used, for example, to indicate any hierarchical relationships. These codes and the cross-reference in 3 are all that is required for a graph-theoretical analysis of the network of concepts.

It is here that any "see other" code would be inserted.

It is also important to indicate the type of relationship between two entities, for example:

- -- logical (i.e., B includes A, etc.)
- -- consistency (contradicion/support) (see Appendix D7)
- -- time (precedes/follows)
- -- cybernetic (information exchange)
- -- responsibility (flow of decisions)
- -- etc. (see Appendix A5)

This is an indication of what is flowing or the nature of the relationship. It does not seem feasible to predetermine the possible types of relationship which might be required (see Appendix A5). The technique which can be adopted is therefore to use a simple numeric code -- the next available in a sequential list -- for each new type of relationship with which a modelling group wishes to work.

The arrangement of this zone could be left up to the modelling group. It is desirable that standard codes should be developed to facilitate graph-theoretical analyses and that a standard code system should be used to denote types of relationships (e.g. "321" where the numbers have no special significance).

6.2. Relationship descriptor B

This is used for evaluation descriptors. In other words the codes used here supply some form of ranking to the relationship described in 6.1. (e.g. some measure of relative importance (within the model), some measure of degree of relativity, etc.)

It is in this zone that the degree of consensus on the characterization of the concept by the discipline could be coded.

The zone may even be used to carry quantitative information on the size of flow represented by the relationship and also its periodicity, if relevant.

Again, the arrangement of this zone could be left up to the modelling group. It is however desirable that a standard form should be developed -- even if exceptions to it are frequent.

7. Date codes

7.1. Date first used

This may be used to indicate the date each relationship between entities was first noted, or alternatively the computer can automatically insert the date on which the relationship was first filed.

7.2. Date last used

This date may be used when the relationship is finally rejected as invalid or unacceptable.

7.3. Retention period

This zone may be used by members of a modelling group to communicate with one another. A member may submit "trial balloon" relationships, with a very short (one-cycle) retention period so that others can "see how it looks". Once agreed, the retention period can be set so that relationships periodically come up for review.

8. Status code.

For modelling group administration purposes, it is convenient to have a zone which may be used to indicate that the relationship is "under consideration", "a tentative proposal", "a firm proposal", "agreed by the group", "required priority attention", etc.

9. Text

Normally a relationship record should require no text. There is however no reason why this zone should not be used for any text comments on a relationship which may seem significant to the modelling group.

C. Term Allocation Phase

1. Entity sequence number

Required as before

2. Model descriptor A

2.1. Model number

Required as before. A term can only be authoritatively allocated within the modelling group. It is utopian to expect that consensus can be consistently achieved between modelling groups on a unique authoritative term for the entity to which they all refer in their different ways.

2.2. Sub-model number

This should normally be zero, since it will probably be easier to achieve consensus on a term between model and sub-model than between model and model.

2.3. Language

Required as before for each language version of the authoritative term.

2.4. Alternatives

This must be "1" or greater to distinguish the term records from the relationship records. If alternative authoritative terms are required in a given language the zone would be used to distinguish between them.

3. Cross-reference.

Normally this would be "0". It may however be necessary to indicate other entities using the same term (but obviously with a different meaning).

4. Source code.

There may be some cases where it is important to indicate the document in which the justification for the unique authoritative term is urged.

5. Model descriptor B.

May be required if the cross-reference to a use of the same term in a different model is needed.

6. Relationship descriptor.

Not required.

7. Date codes.

7.1. Date first used

This may be used to indicate the date the term was first used, or alternatively the computer can automatically insert the date on which the term was first filed.

7.2. Date last used

Terms fall from favour. The last date of use can be indicated here.

7.3. Retention period

May be used as in B.7.3.

8. Status code.

May be used as in B.8.

9. Text.

The words used in the authoritative term are inserted into this zone. Alternatively the equivalent decimal coding could be inserted, if desired.

Seq. No.	Model Descrip. A: Model	Model Descrip. A: Sub-model	Model Descrip. A: Lang.	Alt.	Cross-ref. No.	Sour.	Model Descrip. B: Model	Model Descrip. B: Sub-model	Model Descrip. B: Lang.	Relat. type A codes	Relat. type B codes	Date first used	Date last used	Reten. period	Status codes	Label/Defin./Text/Term
1	2.1	2.2	2.3	2.4	3	4	5.1	5.2	5.3	6.1	6.2	7.1	7.2	7.3	8	9
987	0	0	1	0	0	0	0	0	0					different		(DEMOCRACY
987	0	0	2	0	0	0	0	0	0					language		(DEMOCRATIE
987	0	0	3	0	0	0	0	0	0					labels		(DEMOKRATIE
987	0	0	1	0	849	0	0	0	0							)entities
987	0	0	1	0	279	0	0	0	0							)with same
987	0	0	1	0	988	0	0	0	0							)English label
987	0	0	2	0	434	0	0	0	0							)idem - same
987	0	0	2	0	9321	0	0	0	0							)French label
987	0	0	1	1	0	0	0	0	0							)alter. Engl.
987	0	0	1	2	0	0	0	0	0							)definitions
987	0	0	1	3	0	0	0	0	0							)same concept
987	0	0	2	1	0	0	0	0	0							)idem for
987	0	0	2	2	0	0	0	0	0							)French defs.
987	1	0	0	0	8914	0	1	0	0							)Model 1
987	1	0	0	0	3256	0	1	0	0							)cross-links
987	1	0	0	0	232	0	1	0	0							)
987	25	0	0	0	6754	0	25	0	0							)Model 25
987	25	0	0	0	121	0	25	0	0							)cross-links
987	25	0	0	0	4132	0	25	0	0							)
987	1	0	1	0	0	0	0	0	0					different		(DEMOCRACY
987	1	0	2	0	0	0	0	0	0					language		(DEMOCRATIE
987	1	0	3	0	0	0	0	0	0					terms		(DEMOKRATIE
987	1	0	1	1	849	0	0	0	0							)entities
987	1	0	1	2	279	0	0	0	0							)with same
																)English term
988	0	0	1	0	0	0	0	0	0							POWER
...																
...																

Figure 1. Outline of computer record layout (Column heading numbers refer to text section headings)

Limitation of Scope

The design of the system is sufficiently general that it could be used to order theoretical formulations in any area of knowledge. Such broad coverage would clearly be impracticable, and probably even undesirable, in the foreseeable future. In considering the problem of scope, it is useful to take into account what the UNISIST project intends to cover, bearing in mind that it is designed to order a vast number of documents and not the more limited number of original theoretical formulations which they may contain. Originally (1967-68), it was intended that UNISIST should cover the basic natural sciences but arguments were put forward for the inclusion of technology "or at least some of its branches, especially medicine, agriculture, building and construction." Ultimately, "the position of the ICSU/Unesco Central Committee was that UNISIST should devote its primary effort to the basic sciences...and at the same time be sympathetic to a progressive inclusion of the applied and engineering sciences -- and eventually the social sciences -- on an equal footing with the former" (UNISIST Report, p. 135-6) No time scale was given, however. (In some respects the priorities seem strange, in that it is not scientific knowledge which is lacking but rather the social science knowledge on how best to use and contain it.)

But the prime interest of the UNISIST system, science and technology, suggests a convenient dividing line. Basically the UNISIST system is concerned with documents about material objects, attributes of objects, theoretical formulations about objects, and processes involving the manipulation of objects. A term-oriented documentation information system will undoubtedly be satisfactory for this domain, since any term is probably discreet as far as its meaning to this domain is concerned.

The domain which could be better handled by the project proposed here covers the non-material psycho-social entities, theoretical formulations about psycho-social systems, with the possible inclusion of relationships between objects in environmental eco-systems which interact with social environment systems (*). Clearly even this is vast and ambitious, and the scope can be narrowed even further.

(*) Even in the case of the proposed "global environmental monitoring" information system, which will be integrated into UNISIST, there is no concern for interaction between different pollutants and other environmental factors or with the impact on social systems. The latter is considered "important" but "subjective" and not "included as an operative part of the system" (Global Environmental Monitoring; a report submitted to the UN Conference on the Human Environment, Stockholm, 1972, by the Commission on Monitoring of the Scientific Committee on Problems of the Environment (SCOPE) of the International Council of Scientific Unions (ICSU). Rome, ICSU, 1971.

It is useful to re-emphasize that the proposal is not concerned with the areas covered by social science documentation as there are many such documentation projects. The UNISIST report mentions the parallel programs proposed by such bodies as the International Council of Social Sciences and the International Committee for Social Sciences Documentation. There are numerous equivalent projects at the national level. The object is to concentrate on theoretical formulations which may or may not be mentioned in a given collection of documents.

It is possible to allocate a tentative order of priority for the formulations which should be included. Needless to say this order is governed primarily by the interests of those prepared to allocate resources. Any group with its own funding could modify the priorities and ensure the early incorporation of those formulations which it considered significant.

The priorities proposed would be based on three dimensions:

i) commencing with the more abstract formulations and then moving to the more specific or concrete

ii) commencing with formulations of interest to several social science disciplines and then moving to those common to several schools of thought, and finally to those current within one school of thought only. (The suggestion is that an effort should be made to elaborate the significance of "inter-", "multi-" or "trans-disciplinary" concepts as a priority area of study with respect to knowledge analogous to the focus on international relations as opposed to national level activities. Clearly "inter-disciplinary" can be defined to include "regional" groupings of disciplines down to a "bilateral" interdisciplinary focus. The degree of interdisciplinarity of a concept is a valuable means of determining priorities.)

iii) commencing with theoretical formulations before going on eventually to methods and supporting data

This proposal does not of course preclude any modelling group from concentrating solely on the formulations of its own school of thought. However, since the object is to improve communications between schools of thought using the same terms differently, it does seem that the suggested priorities should focus on the more difficult areas first and establish whatever common ground there is. Clearly once disagreement arises between disciplines or schools of thought over a formulation, more specialized models are required to reflect the subtleties defended by each side.

The main concern should be to ensure that the system reflects the general framework of theoretical formulations. Highly specialized formulations should not clutter up the modelling activity. Little effort should be made to include minutiae

about particular social entities which have not been reflected in more general formulations -- unless such minutiae represent unique evidence of the need for new formulations. The system should be compact and easy to use rather than large and unwieldy as are most documentation systems.

Concept notation in documents

It has been stressed that this project does not require a complex notation system since each concept is represented by a single sequence number, plus an indication of the model number in question, if required. Nevertheless, since one object of this approach is to permit scholars to refer with precision to a particular concept in their papers, a standard method of indicating such a concept in print is required.

A similar problem arises in the natural sciences in distinguishing between different isotopes of the same atom (i.e., cases where slightly different versions of the same atom exist due to differences in atomic weight), where the same symbol does not distinguish between isotopes. The solution adopted is to indicate the atomic weight as a superscript to the standard symbol.

In the case of concepts, represented in print by the same word, one solution would be to use the sequence number of the concept as superscript to the word:

e.g. democracy^{*251} democracy^{*942}

To avoid confusion with bibliographical references, the number could perhaps be preceded by an asterisk.

Computer Record Handling Software

In order to carry out the initial stages, very simple computer programs are quite sufficient. These may be used to accept records of new entities, produce entity lists, accept model coding, produce model lists of concept inter-relationships, accept authoritative terms and produce term thesauri.

At a later stage, which should however be kept in view in the design of the first stage program, it is possible to switch from sequentially ordered processing to processing networks of concepts. Sequential processing is highly convenient in computer terms in order to maximize the efficiency of the administrative aspects of record handling, sorting and list production. It does not however give direct access to networks of concepts (and other entities) radiating out from the central entity in which the user happens to be interested, nor does it allow him to switch rapidly from model to model for comparisons. In other words, for day to day operations resulting in the production of standard check lists and thesaurus updating, sequential processing is probably essential, whereas when the information stored is to be used via a direct access terminal or on a query by query basis, then some form of network processing is essential. It is a relatively simple matter to convert from one to the other provided this is planned for. The record handling could in fact be done (centrally) on a sequentially ordered file and institutions wanting copies could convert the file into a network order for direct access work within their institute.

Software already exists to handle "networks". A frequent application is the computer processing of Critical Path and PERT networks. These are networks over time and are less applicable then programs developed to handle parts listing and assembly and stock problems in manufacturing companies. One of these programs PLUTO (Parts Listing/Used-on Technique) developed by International Computers (UK) will be described as an illustration. (N.B. UNESCO (Paris) has installed an ICL computer which could use this software.)

PLUTO disc files record structures. That is, data about the entities that form structures and the relationships between them. Entities can be a person, an organization, a concept, a problem or any nameable thing. Many types of entity and relationship may be handled simultaneously. Information is held in the form of multiple inter-linked hierarchies of entities which greatly simplifies retrieval and presentation.

A distinction is made between master files (denoted by rectangles in Figure 1) which carry data about the entities within a structure and structure files (denoted by diamonds) which carry data about the relatipnship between entities. The files are linked together by cross-references to form a total information system or data base which can comprise a number of master and structure files. Information is retrieved under program control by following links from record to record and from file to file.

Each master file contains the entity records and each structure file contains the cross-references either between the entities in the same master file or to those in a second master file. The network of files may be added to or modified as new applications are envisaged.

In the management situations (for which the program was originally conceived) structure information is complex and affects many parts of an organization (see Figure 1). The same is even more true of structure information in relation to concepts, theories, assumptions, methods, etc. The special diagrammatic notation (shown in Figure 1) has been developed to facilitate thinking about the sort of interlinked system of files which is necessary in a given case. This notation is used in Figure 2 to illustrate the power of this approach as a means of handling the different conceptual entities of interest to social science. Clearly this can only be an illustration for much thought is required to obtain the correct file design.

Special computer programs are used to explore at user request the structures created by complex file interlinkage of this kind. Searches down a hierarchy are termed explosions and searches up a hierarchy are termed implosions. These may be requested from any starting entity or file and can be governed by examination of qualifiers in link or entity records at each level encountered. It is this sort of feature which would be vital to obtaining full benefit from the graphics display (see Appendix D3).

It should be apparent that this, if not the software itself, is a very useful method of handling and exploring data on the relationship between concepts and other entities. In fact, the full power of the PLUTO software would not be required (although it has the advantage of being available) and it is possible to envisage a very much abridged version of it which would perform all the structural inter-linking required and be more easily related to the sequentially ordered file.

Computer programs relevant to this project have been produced for work on sociometric data. Programs are also mentioned in connection with citation indexing (Appendix D3), analysis of belief structures (Appendix C2), and personal construct theory (Appendix C3).

The ADMINS Computer System (*)

Work has been in progress for some years at the M.I.T. Center for International Studies on the development of very general systems for time shared computer data management. The key to the ADMINS data management concept is the maintenance of data content directories at the systems programming level normally reserved for computer operating systems and disk access control codes. The programming language allows the specification of relationships between named characteristics of entities and allows these relationships to be manipulated in several useful ways. Great stress is placed on using the computer to function as "officer manager" in handling and checking incoming information to be inserted into the system.

An item of data is perceived to be a sequence of categories of information in n-adic relations applied to a specific entity. Relations may be:

a) monadic concerning one category (e.g. something exists);

b) dyadic concerning two categories (e.g. an entity has an attribute, an entity precedes another entity in time; an entity includes another entity, an entity receives information from another entity, etc.);

c) triadic concerning three categories (e.g. an entity sends a certain type of information to some other entity, an entity includes one entity which is related to a third entity, etc.)

d) etc. for four or more categories (**)

N-adic data descriptions for social science propositional inventories are noted as being quite complicated e.g. 'violence' is 'power' over 'power' over 'well-being'.

The ADMINS system makes use of a "calculus of relations" for stating the derivation of a new relation that draws on those already existing, and which yields a new relational record between particular entities. It is in the structuring of the programming language around the relational record and in achieving intimate interaction with many storage levels that this system differs from most procedure languages.

(*) This Appendix is based upon the following material:

Stuart D. McIntosh and D.M. Griffel. The requirements for a computer-based information system. M.I.T., Center for International Studies, 1968, (c/68-14c), 82 p.

- The current ADMINS System for non-textual data. M.I.T., Center for International Studies, 1967 (Rev. Oct 1968; c/67-26), 39 p.
- Computers and categorization (Paper presented to the Classification Research Conference, Bangalore, 1969). M.I.T., Center for International Studies, 1969 (C/69-28), 41 p.
- ADMINS Implementation. (M.I.T., Center for International Studies), 1971.
- Large disparate data basis. In: E.L. Jacks (Ed.). Associative Information Techniques. N.Y., American Elsevier, 1971, p. 71-113.

(**) In the Mark III version of the system only dyadic relations were possible. The Mark V version will permit four or more category relations.

One of the forms of analysis possible, which is relevant to this project, is that of cross-reference analysis. The system is designed to handle sociometric data, citation relationships, thesaurus structures, and "maps". The two main features are measurement of flows and logical operations on the cross-reference relations.

In responding to problems, including the non-hierarchic classification schemes noted with respect to this project, a simple matrix structure is ruled out (e.g. categories are columns, rows are items, cells hold entries). An "extended" type of complex matrix is used as the basis for the data structure which has "both vertical and horizontal pointers scaffolding small arrays."

The system is designed to facilitate "model building", particularly with the use of social statistics. It would also appear that the multiple model technique suggested for this project could be easily handled together with some of the problems of conversion between models. The system is of course specially designed to permit many researchers, each at different computer terminals, to experiment simultaneously with and redefine their own sets of categories from a common data base. Such "experimental models" can either be deleted when completed or stored for further use. The aim is to provide an environment where the researcher is really interacting with his data, so that he can make effective intellectual decisions in response to substantive results from the terminal at the pace at which he is able intellectually to deal with his problem.

The ADMINS system is a very ambitious one. It is designed at a high level of generality to handle many applications which are of little interest to this project. But it is quite evident that this project could be run at quite a high level of sophistication on the ADMINS system -- even, possibly, to the point of permitting a modelling group (with each member at a terminal) to interact with one another, and the model on which they are working, as a "computerized committee".

It is also obvious that the "large scale ADMINS" approach is too sophisticated and too dependent on access to large third generation installations. A "limited ADMINS installation" is possible however. The early success of this project, however, depends more on the ability to use much simpler installations for the filing and listing operations, whilst always permitting a switch to a more complex mode, possibly a subset of ADMINS, for network analysis, graphics display, etc., for specific research projects on the data base for which resources can be obtained. Much interesting research can however be undertaken using low cost programs, many of which already exist.

Use of Interactive Graphic Display Techniques.

Description.

The suggestion has been made (see Appendix C1) that structuring the relationship between theoretical entities (concepts, propositions, problems, etc) could best be accomplished using graph theory methods. There are three disadvantages to this approach:

- -- graphic relationships are tiresome and time-consuming to draw (and are costly if budgeted as "art work").
- -- once drawn, there is a strong resistance to updating them (because of the previous point) and therefore they quickly become useless.
- -- when the graph is complex, multidimensional, and carries much information, it is difficult to draw satisfactorily in two dimensions. The mass of information cannot be filtered to highlight particular features -- unless yet another diagram is prepared.

These three difficulties can be overcome by making use of what is known as "interactive graphics" (*). This is basically a TV screen attached to a computer. The user sits at a keyboard in front of the screen and has at his disposal what is known as a light-pen (or some equivalent device) which allows him to point to elements of the network of concepts displayed on the screen and instruct the computer to manipulate them in useful ways. In other words the user can interact with the representation of the conceptual network using the full power of the computer to take care of the drudgery of

- -- drawing in neat lines
- -- making amendments
- -- displaying only part of the network so that the user is not overloaded with "relevant" information

In effect the graphics device provides the user with a window or viewport onto the network of concepts. He can instruct the

(*) This term is used widely to cover both the more common "alphascopes", which can display letters and numbers on predetermined lines, and the "vector displays" with light-pen facility, which can also generate lines and curves. It is the latter device which is discussed here. See, for example:

See: Ivan Sutherland. Computer displays. Scientific American, 222, June 1970, p. 56-8.
Interactive graphics in data processing. IBM Systems Journal, 7, 3 and 4, 1968, whole double issue.
Computer Graphics 1970; and international symposium. Brunel University, 1970, 3 vols.
Michael S. Wolfberg. An interactive graph theory system. Moore School of Electrical Engineering Report 69-25, University of Pennsylvania, 1969.

computer, via the keyboard, to:

1. move the window to give himc, effectively, a view onto a different part of the network -- another conceptual domain
2. introduce a magnification so that he can examine (or "zoom in" on) some detailed sections of the network
3. introduce diminution so that he can gain an overall view of the structure of the conceptual domain in which he is interested
4. introduce filters so that only certain types of relationships and entities are displayed --either he can switch between models or he can impose restrictions on the relationships displayed within a model, i.e. he has a hierarchy of filters at his disposal
5. modify parts of the network displayed to him by inserting or deleting entities and relationships. Security codes can be arranged so that (a) he can modify the display for his own immediate use without permanently affecting the basic store of data, (b) he can permanently modify features of the model for which he is a member of the responsible body, (c) and so on.
6. supply text labels to features of the network which are unfamiliar to him. If necessary he can split his viewport into two(or more)parts and have the parts of the network displayed in one (or more) part(s). He can then use the light pen to point to each entity or relationship on which he wants a longer text discription (e.g. the justifying argument for an entity or the mathematical function, if applicable, governing a relationship, and have it displayed in an adjoining viewport.)
7. track along the relationships between one entity and the next by moving the viewport to focus on each new entity. In this way the user moves through a representation of "semantic space" with each move, changing the constellation of entities displayed and bringing new entities and relationships into view.
8. move up or down levels or "ladders of abstraction". The user can demand that the computer track the display (see point 7) between levels of abstraction, moving from subsystem to system,at each move bringing into view the semantic context of the system displayed.
9. distinguish between entities and relationships on the basis of user-selected characteristics. The user can have the "relevant" (to him) entities displayed with more prominent symbols,and the relevant relationships with heavier lines.
10. select an alternative form of presentation. Some users may prefer block diagram flow charts, others may prefer a matrix display, others may prefer Venn diagrams (or "Venn spheres" in 3 dimensions) to illustrate the relationship between entities. These are all interconvertible (e.g. the Venn circles are computed taking each network node as a centre and giving a radius to include all the sub-branches of the network from that node.)
11. copy a particular display currently on the screen. A user may want to keep a personal record of parts of the network which are of interest to him. (He can either arrange for a dump onto a tape which can drive a graph plotter, a microfilm plotter, or copy onto a videocassette, or, in the future, obtain a direct photocopy.)
12. arrange for a simultaneous search through a coded microfilm to provide appropriate slide images or lengthy text (which can in its turn be photocopied).
13. simulate a three-dimensional presentation of the network by introducing an extra coordinate axis.
14. rotate a three-dimensional structure (about the X or Y axis) in order to heighten the 3-D effect and obtain a better overall view "around" the structure.
15. simulate a four-dimensional presentation of the network by using various techniques for distinguishing entities and relationships (e.g. "flashing" relationships at frequencies corresponding to their importance in terms of the fourth dimension.)
16. change the speed at which the magnification from the viewport is modified as a particular structure is rotated.
17. simulate the consequences of various changes introduced by the user in terms of his conditions. This is particularly useful for cybernetic displays.
18. perform various topological analyses on particular parts of the network and display the results in a secondary viewport (e.g. the user might point a light-pen at an entity and request its centrality or request an indication of the interconnectedness of a particular domain delimited with the light pen.)

Graphics and communication.

In order to understand the value of interactive computer graphics, a few basic principles of communication should be considered. Languages are used to convey thoughts. Languages may be gestural, verbal, written, notational, or graphic. The effectiveness of a language depends upon its ability to retain and transfer meaning and this in turn depends upon the complexity of the language. One can conceive of a spectrum of "language and medium" from primitive gestures through to sophisticated computer environments. At each point in the spectrum there are disadvantages and advantages for communication. An attempt has been made to list these out in Figures 1 and 2. These should be considered as very tentative schemas only. (*)

(*) Figure 1 was inspired by a similar tentative effort by Colin Cherry to relate communication equipment (radio, TV, press, etc.) to psycho-social qualities. See: World Communication, threat or promise? New York, Wiley, 1971, p. 53.

These Figures suggest that most of the advantages of the early portions of the spectrum are combined together in the later protions where interactive graphics is used in various ways. The question in why do graphics help to convey more information than words. One reason is that as concepts become more complex they do not lend themselves to easy encapsulation in words and phrases. Many objects, processes, or abstractions can be portrayed for discussion using a few simple graphical symbols much more easily than they can be described verbally (cf. the classic example of the spiral staircase). The other pressure is of course that many subtle invariants and relationships currently displayed in statistical tables, are ignored unless they can be represented in meaningful graphical form (*).

Some current interactive graphics uses include, for example, calculation and analysis of electronic circuits, design of aerodynamic shapes and other mechanical pieces, design of optical systems and plasma chambers, simulation of prototype aircraft and rocket flight, visualization of complex molecules in 3 dimensions, air traffic control, chemical plant control, factory design and space allocation, project control, primary, secondary and university education and educational simulations.

In every case above there is some notion of geometry and space, but the geometry is always the three-dimensional conventional space. There is no reason why "non-physical spaces" should not be displayed instead -- and this is the domain of topology. The argument has been developed by Dean Brown and Joan Lewis (**).

> "Both geometry and topology deal with the notion of space, but geometry's preoccupation with shapes and measure is replaced in topology by more abstract, less restrictive ideas of the qualities of things...Being more abstract and less insistent on fine points such as size, topology gives a richer formalism to adapt as a tool for the contemplation of ideas....
>
> Concepts can be viewed as manifolds in the multidimensional variate space spanned by the parameters describing the situation. If a correspondance is extablished that represents our incomplete knowledge by altitude functions,

(*) See, for example: R. Bach. Graphical national patterns; a new approach to graphical presentation of statistics. London, Humphrey, 1968.

( ) Brown, Dean and Lewis, Joan. The process of conceptualization; some fundamental principles of learning useful in teaching with or without the participation of computers. Educational Policy Research Center, Stanford Research Institute, Menlo, Park, California. p. 16-18.

> we can seek the terrae incognitae, plateaus, enclaves of knowledge, cusps, peaks, and saddles by a conceptual photogrammetry. Exploring the face of a new concept would be comparable to exploring the topography of the back of the moon. Commonly heard remarks such as "Now I'm beginning to get the picture" are perhaps an indication that these processes already play an unsuspected role in conceptualization.
>
> Topology is thus a generalization of the idea of diagram. Traditional in teaching (grammatical diagrams of sentences, geneologies of kings, whirling models of solar systems), it extends easily to the machine. By sketching tentative three-dimensional perspectives on the screen and "rotating them on the tips of his fingers", one internalizes ideas nonverbally and acquires a sensation of sailing through structures of concepts much as a cosmonaut sailing through constellations of stars.
>
> Such new ways of creating representations break ingrained thought patterns and force re-examination of preconceived notions. A mapping is a correspondence is an analogy. Teaching by analogy, always a fertile device, can be carried out beautifully by topological means....Topological techniques are useful at even the most advanced levels of scientific conceptualization....
>
> Most traditionally educated humans are brought up with the belief that thinking is synonymous with verbal thinking. The time seems ripe to make a break with this limiting concept. Psychologists and educators are coming to the realization that man often has to get away from speech to think clearly. Scientists and creative artists have testified that to create they had to regress at times from the word to the picture, from verbal symbolism to visual symbolism....
>
> Whether the concept seems spontaneous or belabored, there comes a point in its evolution when the mind transcends its accounting for the elements of information one by one, and begins to form an integrated impression. The whole is quantitatively differentiated from the sum of its parts to become conceptually quite different."

It is useful to introduce C.S. Peirce's term "iconic", namely "a diagram ought to be as iconic as possible, that is, it should represent (logical) relations by visible relations analogous to them."(*) Iconics is therefore connected with the degree to which features of the graphics display contribute towards (or

(*) C.S. Peirce. Collected Papers; edited by C. Hartshorne and P. Weiss. Harvard Univ., 1933, vol. iv. cited by: A. Battersby. "The Application of iconic principles to the design of a problem solving system for psychological research and of value for management training." In: Computer Graphics '70. Brunel University, 1970.

There is, however, a question of "iconicity for whom". Philip Hendren (*) cites a well-known survey by Anne Rowe (The Making of the Scientist) in which a high correltaion was found between (1) visual imagery and experimental inclination, (2) non-visual imagery and preference for theoretical science. Many theoretical scientists prefer not to use visual imagery -- which may explain their difficulty in communicating with other sectors of society. Don Fabun (**) points out in the following that non-Americans may not find the display of concepts and their relations by newtork structures very meaningful (***).

> "Americans tend to see the edges of things and the intersection points of crossing lines, and to attach importance to them. Thus our streets are normally laid out in a grid pattern and we identify places by their proximity to intersections. Europeans and Orientals, however, are inclined to attach importance to an area; thus a French street or avenue may change its name every few blocks; and houses in Japan may not have street numbers but be identified by name and area or the time at which they were built."

The fundamental importance of interactive graphics, in whatever form, is its ability to facilitate understanding. Progress in understanding is made through the development of mental models or symbolic notations that permit a simple representation of a mass of complexities not previously understood. There is nothing new in the use of models to represent psycho-social abstractions. Jay Forrester (****), making this same point with respect to social systems, states

> "Every person in his private life and in his community life uses models for decision making. The mental image of the world around one, carried in each individual's head, is a model. One does not have a family, a business, a city, a government, or a country in his head. He has only selected concepts and relationships which he uses to represent the real system. The human mind selects a few perceptions, which may be right or wrong, and uses them as a description of the world around us. On the basis of these assumptions a person estimates the system behaviour that he believes is implied....The human mind is excellent in its ability to observe the elementary forces and actions of which a system is composed. The human mind is effective in identifying the structure into which separate scraps of information can be fitted.

(*) Philip Hendren. Computer graphics and color for design and communications in architecture and urban design. In: Computer Graphics 70. Brunel University, 1970.

(**) Don Fabun. Communications; the transfer of meaning. Glencoe Press, 1968, p. 25.

(***) Although, in the case of France, there has been a marked increase in attention to "communication networks" and "organizational networks"

(****) Jay Forrester. World Dynamics. Cambridge, Mass. Wright-Allen, 1971, p.14-15.

> But when the pieces of the system have been assembled, the mind is nearly useless for anticipating the dynamic behavior that the system implies. Here the computer is ideal. It will trace the interactions of any specified set of relationships without doubt or error. The mental model is fuzzy. It is incomplete. It is imprecisely stated. Furthermore, even within one individual, the mental model changes with time and with the flow of conversation. The human mind assembles a few relationships to fit the context of a discussion. As the subject shifts, so does the model. Even as a single topic is being discussed, each participant in a conversation is using a different mental model through which to interpret the subject. And it is not surprising that consensus leads to actions which produce unintended results. Fundamental assumptions differ but are never brought out into the open."

These structured models have to be applied to any serially ordered data in card files, computer printout or reference books to make sense of that data. Is there any reason why these invisible structural models should not be made visible to clarify differences and build a more comprehensive visible model? The greater the complexity, however, the more difficult it is to use mental models. For example, in discussing his examination of an electronic circuit diagram, Ivan Sutherland writes: (*)

> "Unfortunately, my abstract model tends to fade out when I get a circuit that is a little bit too complex. I can't remember what is happening in one place long enough to see what is going to happen somewhere else. My model evaporates. If I could somehow represent that abstract model in the computer to see a circuit in animation, my abstraction wouldn't evaporate. I could take the vague notion that "fades out at the edges" and solidify it. I could analyze bigger circuits. In all fields there are such abstractions. We haven't yet made any use of the computer's capability to "firm up" these abstractions. The scientist of today is limited by his pencil and paper and mind. He can draw abstractions, or he can think about them. If he draws them, they will be static, and if he just visualizes them they won't have very good mathematical properties and will fade out. With a computer, we could give him a great deal more. We could give him drawings that move, drawings in three or four dimensions which he can rotate, and drawings with great mathematical accuracy. We could let him work with them in a way that he has never been able to do before. I think that really big gains in the substantive scientific areas are going to come when somebody invents new abstractions which can only be represented in computer graphical form."

Graphics and education.

A visual display unit linked to a computer has considerable advantages as a technique for the communication of new concepts (*). As the multiplicity of conceptual frameworks increase in complexity, new techniques must be sought to simplify education concerning them. The problems posed by the time currently required to communicate even a superficial knowledge of the existing frameworks , and the difficulty of building up an integrated picture of their complexity, suggest that a visual display unit with computer mass memory support may have many possibilities.

An important reason for using this approach is the tendency to consider the recognized complexity of the system of disciplines to be too great to lend itself to any form of unified treatment. Normal instruction methods, in the case of such complexity, would have to cross so many discipline boundaries that they necessitate concentration on one particular feature of the system at the expense of the others, and any integrated picture of the whole.

An important possibility in building understanding is the ability to manipulate part of a multidimensional network, via the visual display unit, so as to portray the system of conceptual networks from an origin chosen anywhere within the network. Thus a concept(or even organization, known and understood by a particular user, may be used as visual origin and all other concepts (or organizations) displayed in terms of their relationships to it -- according to a variety of models helpful to differing personality types. Entities distant in communication terms can be reduced in visual importance, whereas "nearby" concepts of relatively little "absolute" importance can be made of greater significance (approximating the recognition normally accorded them by the user).

The newcomer to a conceptual framework has a known system base from which to start his exploration of the neighbouring systems which interact with it. In a programmed learning mode, he is able to understand how his known systems are "nested" within any larger system. He can work from his base system by requesting a restructuring of the display in terms of other system viewpoints as he builds up knowledge of, and a "feel" for, those originally conceptually distant from his starting point. Text can be displayed concerning the new system, interaction or perspective, before any new "jump" is made. In this way, he can progress toward the more fundamental levels of any conceptual framework or into other areas of detail.

A valuable feature of an interactive system is the possibility given to a student of simulating the result on the system of "wiping out" a single sub-system or class of systems which he believes to be of little value. Of greatest importance, the student can work out and locate which conceptual frameworks (or organizations) offer the best avenue of fulfillment for him, or, alternatively, precisely in what way he must initiate

(*)See, for example, Douglas Engelbart. Augmenting Human Intellect; a conceptual framework. Stanford Research Institute, 1962.

some new activity to achieve such a measure of satisfaction. By exploring the network the student is, in a sense, engaging in a parallel exploration of "semantic space". This is of some value according to some perspectives.

Just as world unity is a long way off in organizational terms, and yet a multidimensional network of organizations can be "held" in computer memory for exploration, so unification of knowledge can be simulated by holding and linking concepts in different frameworks between which links have been suggested during the course of research. In both cases the dynamic collection of data stands as a symbol of the goal. Built up empirically, the systems must be explored by research workers and students alike in order to improve their concepts of the more general systems. The details of interactions can be provided in considerable amounts, but the problem for both is to build up more integrative concepts.

The process of interaction between display and person is really one in which the display is used as a crutch until the mind can hold a more integrative concept. The mind is the most potent display device. The problem is how to 'pump' it (in laser terms) to an optimum operating frequency with the aid of interactive displays. It may be possible to use the graphics display unit as a focusing device when "hunting" intuitively for a creative solution. At a certain point the user's mental display abilities can be "launched" from the sophisticated back-up or steadying platform provided by the display unit. If necessary the display can be improved prior to a "relaunch". It is with this sort of approach that the speed of convergence on unifying concepts can be increased.

In order to improve the rate of generation of more integrative concepts, it may be valuable to examine the validity of some of the following assumptions:

1. "Highly general and integrative concepts can be adequately communicated through symbols on paper." It may be that the more abstract concepts required cannot be adequately grounded in symbols on paper (that is without merely using the symbol as an aide-memoire). It is possible to conceive of an equlibrium diagram which would indicate in what communication media, or combination of media, a given concept could be "held", and in which it was metastable or unstable.

2. "Once a concept is "discovered", we ourselves can remember it and hold it effectively in our own minds." To hold a concept however, requires a constant stream of appropriate environmental stimuli to reinforce it. This is particularly the case if the concept is highly sophisticated and "delicate" (even if, and perhaps particularly if, very simple). Conventional media may be associated with a characteristic reinforcement rate which may be too low to permit certain concepts to be held for long.

It may be that sufficiently rapid reinforcement can only be provided with interactive graphics devices. One author emphasizes their importance for maintaining "thinking momentum". We may have to keep using such devices to aid us in focusing our thinking to recover the concept "which we have already discovered", until we have built up an attitude which permits us to pick out sufficient reinforcing evidence from the environment unaided.

3. "Highly integrative concepts can be developed by interaction between specialists using conventional communication and storage media and traditional academic interaction procedures." It may be that for purely technical reasons (despite the possible wishes of the people concerned) interaction may not permit the generation of unifying concepts of great generality.

Interactive devices create a man-machine environment with properties which differ from those of the traditional concept generating environments. Skillfully used it may be possible to ensure the interaction of specialists, manipulating related concepts through interactive devices, in such a way that progressive convergence towards increasingly more general concepts is built into the interaction process.

4. "The discovery of general integrative concepts crossing discipline boundaries would of itself lead to solutions to the problems of modern society." It may be that the place allocated to such concepts in modern society is such that their value is effectively negated. They may be "contained" in a position in culture space in such a way that they are prevented from having any marked effect on society -- even those which have not yet been developed. This is a reason for studying the system in which such concepts are developed.

It may be a question of the speed with which the concept can be got over (and "anchored") relative to that of the reaction of compensating social mechanisms coming into play to counteract any implied changes.

5. "Old or primitive unifying concepts are irrelevant in the 20th century." It may be the case that for some groups of personality types certain 'outmoded' integrative concepts are the most useful in terms of the problems to which they are exposed in the light of their conceptual apparatus. Similarly, it may not necessarily be true that the learning path for some students and schoolchildren is optimized if the latest theories are stressed at the expense of their historical predecessors.

A major function of systems thinking could be to determine the inter-relationship between historical viewpoints -- particularly since many of them are still held in some parts of the world system.. With appropriate techniques

a student could locate the unifying concepts most in sympathy with his current understanding of his environment. Related techniques could then be used to expose him as rapidly as possible to the evidence which outdates his view. The system framework could then speed him conceptually through the succession of systems perspectives up to the present. Should he "stick" at any point, then it would be in a context which for him possessed lower entropy than that which he perceives in later points. He thus sticks at the point which most reinforces his concept of himself as a whole person in a unified conceptual environment.

Further graphics possibilities

1. It is technically feasible to copy a displayed conceptual network onto a videocassette. These can be recopied for distribution and are played back over normal television sets. This gives a non-interactive, low-cost access to the same information. This technique has considerable potential for education, briefings, and research.

2. Microfilm plotters are currently used to copy the contents of a display directly onto microfilm. They have the advantage of being extremely fast. In addition, unlike current display screens, they can handle very complicated diagrams with several thousands of lines and symbols. The microfilm can then be processed automatically and mounted on aperture cards or enlarged to hard copy.

This gives an excellent method of building up low cost "maps" of the conceptual domains in which one is interested. Alternatively, the film itself can be used for preparing demonstration movies. It would also be technically feasible to arrange for the microfilm frames to be coded under computer control so that the film can be optically scanned to permit later display of a user-specified frame (as on the Kodak Miracode system).

3. Colour graphics units are in use (some up to 150 x 150 cm in size). These permit entities and relationships to be coded so that even more information can be held in one image. The use of colour is however more applicable to displays of areas, such as might be used with a Venn diagram, rather than a network, presentation. (*)

4. It is possible to plot any diagram using drum (simple graphs) or flatbed (complex diagrams) plotters. The latter occur in sizes up to 150 x 1000 cm.

5. Helmets fitted with display screens for each eye have been developed (to train pilots in landing expensive fighters on aircraft carriers). The wearer is provided with a perspective on displayed structures which changes

(*) (Philip Hendren) An ordering principle for perspective drawings on a color display. Unpublished Technical Report, Rice University, School of Architecture.

as he moves his head(*). It could be used to fill a semantic space with structures through which the individual could move i.e. he is completely surrounded by computer generated structures (with which he could interact).

6. A number (up to hundreds) of display terminals may use a common data base. This permits users to interact with, and explore, each other's "semantic space" in a very intimate manner (**). A team can work together on the additions to some complex structure - users from different disciplines each contributing elements and linkages. This technique is currently used in the allocation of structures in three-dimensional space in the design of complex factories, where ventilation, electrical, piping, chemical, and many other engineers have to interlink the structures with which they are concerned (***).

7. Very suggestive of new approaches to experimenting with concept, problem or organizational structures is the work underway using graphics to detect all the different possible ways of constructing a specified chemical structure, given a set of specified possible sub-units and restrictions on the ways they can be combined (****). One possible application in this context, is the charting of possible sequences of concepts leading to the understanding of some more general concepts. This would be of interest in programmed learning work.

8. There is much parallel interest in interactive graphics for art. A definite convergence of interest in the handling of structures and relations is now evident (*****). Hopefully this will lead to the development of even more sensitive interactive devices which could be used to contain and reflect even subtler concepts - a sort of

---

(*) Ivan Sutherland. Computer displays. Scientific American, 222, June 1970, p. 56-61

(**) Nilo Lundgren. Toward the decentralized intellectual workshop. Innovation, Technology Communication, 1971, 24, p. 50-60 (reporting on D. Engelbart's work at Stanford).

(***) H.J. Genthner. Interactive computer graphics. Computer and Automation. November 1968, p. 14-17.

(****) E.J. Corey and W. Todd Wipke. Computer-assisted design of complex organic syntheses. Science, 166, 10 October 1969, p. 178-192 (see also the third footnote on next page)

(*****) See: Papers on computer graphic art (In: Computer Graphics 70. Brunel University, 1970, 3 vols.); papers on art and decision-making information (In: H.S. Brinkers (Ed.). Decision-making - creativity, judgement and systems in press).

dynamic interactive ideograph (*).

Perhaps it will only be such devices which will ensure the adequate utilization of theoretical knowledge. As Harold Lasswell points out:

> "Why do we put so much emphasis on audio-visual means of portraying goal, trend, condition, projection, and alternative? Partly because so many valuable participants in decision-making have dramatizing imaginations ....They are not enamoured of numbers or of analytic abstractions. They are at their best in deliberations that encourage contextuality by a varied repertory of means, and where an immediate sense of time, space, and figure is retained." (**)

(*) Some interesting theoretical and technical suggestions towards such a device have been made by Gordon Hude (A device for generating a universal binary metalanguage for computer operation. London, Prov.Dat.Spec. 69.212; also other unpublished documents from Gordon Hude, 11 The Close, Dunmow, Essex CM6 1EW, England.)

(**) Harold D. Lasswell. The transition toward more sophisticated procedures. In: Davis B. Bobrow and J.L. Schwartz (Ed.) Computers and the Policy-making Community; applications to international relations. N.Y., Prentice-Hall, 1968, p.309.

(***) For suggestive uses of computers to construct potential fields around interacting entities (in this case atoms), see:

Arnold C. Wahl. Chemistry by computer. Scientific American, April 1970, p. 54-70.

Arnold C. Wahl, et.al. BISON; a new instrument for the experimentalist. International Journal of Quantum Chemistry, Symposium 3, Part 2, p. 499-512, 1970.

A.C. Wahl. Chemistry from computers. Argonne National Laboratory Reviews, 5,1, April 1969, p. 43-69.

Figure 1 Comparison of different methods of communicating concepts

Method	Advantages	Disadvantages
Gesture	direct and to the point; dramatic impact	no abstraction possible
Speech	personalized, subtle, poetic, imageful, analogy-full, adjusted to audience	no permanent record, meanings and models shift from phrase to phrase
Writing	permanent record; words weighed and compared in context; document forms an intelligible whole	meaning of words undefined or differ between documents; definitions become concretized and language dependent; complexity of abstractions limited by syntax of language; problem of jargon
Image	provides context in physical terms; involving, highly complex, high information content, high interrelationship	superficial and unstructured
Maths	handles very complex abstractions and relations and a multiplicity of dimensions	loss of intuitive appreciation of the concepts involved; impenetrable without lengthy initiation; system of notation becomes more complex than the concepts described; impersonal
Diagram (exhibit charts)	structured to make a specific point	over-simplification; exaggeration of some features at expense of others; processes only displayed statically
Artistic mobiles	complex, new and unpredictable relationships	experience primarily incommunicable
Diagram (flow charts/ graphs)	portray all detectable inter-relationships in precise manner; panoramic view of system	visually complex to the point of impenetrability; processes still conveyed statically; difficult to modify
Interactive graphics (alphascope)	precise messages; responsive; contents can be oriented to suit user	no structured overview; bounded by language mode of program; processes conveyed as a sequence of isolated messages (or as a game experience)
Psychedelic environment	very subtle and complex imagery and relationships; process oriented; integration of visual and audio; psychologically involving	no scientific content; no significant invariants; experience primarily incommunicable
Interactive graphics (structured image)	greater user selectivity and control on content and form of presentation; complex abstractions held on display; processes displayed as flows; dynamic; enhanced creativity; 2-4 dimensions.	highly structured without the subtle relationships characteristic of arts; user still centred "outside" the structure "looking in"
Computer graphics art	generation of new and unpredictable dynamic imagery	no scientific or "real world" predictive value
Interactive graphics (multi-terminal)	teams working simultaneously on same ideas; access to each others "semantic space"; interactive thinking	fundamental distinction remains between artistic use of the display or surface volume and scientific interest in structure and data base;
Interactive graphics (coloured & image)	higher information content; visually more intriguing; closer to artistic media; more powerful presentation of processes	still only reflects a portion of the subtleties of all invariants and processes known to psychologists, diplomats, etc.
Interactive graphics (3D - helmet)	user psychologically centred within the structure	
Interactive ideograph (hypothetical)	continuous gradation and interaction between scientifically structured and aesthetically structured display; enhanced creativity; reflects subtleties of psychologists, diplomats, etc. able to convert to and from a "field theory" presentation of structures	still only a scaffolding for disciplined thought

Figure 2: Cross-comparison of different methods of communicating concepts

(H=high; M=medium; L=low)

	visible aesthetic quality	psychologically involving imagery	person-oriented	adjustment to user's conceptual skills	complexity of interrelationships displayed	multi-level structured interrelationships	display of field-based interrelationships	level of abstraction handled	limitation to superficial imagery	subtlety of invariants & interrelationships	iconicity (approx. of formal presentation to concept structure	ability to display process	dynamism of display	unpredictability of imagery and output
Gesture	M	H	H	H	L	L	0	L	L	L	L	L	H	M
Speech	H	H	H	H	H	H	L	H	L	H	L	L	H	H
Writing	L	L	L	M	H	L	L	H	M	H	L	0	0	H
Image	H	M	H	0	H	0	L	L	0	H	M	M	M	H
Maths	0	0	0	0	H	L	0	H	H	H	M	0	0	H
Charts	L	L	0	H	L	L	0	L	L	0	M	0	0	L
Mobiles	H	H	M	L	M	L	M	L	L	M	H	M	M	M
Graphs	L	L	L	L	H	M	L	M	M	M	M	M	L	L
Alphascope	L	L	0	L	M	L	0	L	H	L	L	L	M	L
Psychedelic	H	H	M	0	H	L	H	L	0	H	H	H	H	H
Interactive graphics	M	M	L	H	H	H	L	M	L	M	M	M	H	M
Graphics art	H	H	L	L	M	L	H	L	M	M	H	M	M	M
Multi-terminal	M	M	L	H	H	H	L	M	M	M	M	M	H	H
Colour Graphics	H	H	L	H	H	H	M	M	M	M	M	M	H	H
3D-helmet	H	H	H	H	H	H	M	M	M	M	M	M	H	H
Ideograph	H	H	H	H	H	H	H	H	M	H	H	H	H	H

contd. Figure 2

	display of hierarchy of abstraction	ordered movement between levels of abstractions	scientific and predictive significance of display	associated access to computing power	ability to convert between disciplinary models	associated access to reinforcing didactic material	analogy-full imagery-full	facilitation of comprehension	interactive stimulus to thinking momentum	enhancement of creativity	stabilization of elusive abstractions	insight decay rate reduction	dramatic impact	reproducible permanent record
Gesture	0	0	0	0	0	0	M	M	L	L	M	L	H	0
Speech	L	L	L	0	M	H	H	H	M	M	L	L	H	0
Writing	L	L	L	0	L	L	H	L	0	L	M	L	L	H
Image	L	0	0	0	0	0	H	M	0	L	L	M	H	H
Maths	0	H	H	0	L	0	0	0	0	0	H	0	0	H
Charts	L	L	L	0	0	0	L	M	0	0	L	L	L	H
Mobiles	L	0	0	0	0	0	L	L	L	L	L	0	M	L
Graphs	L	L	H	0	L	0	0	M	0	0	M	L	L	[illegible]
Alphascope	0	0	L	H	L	M	L	M	M	M	L	L	L	[illegible]
Psychedelic	0	L	0	0	L	0	H	L	L	0	0	0	H	L
Interactive graphics	M	M	H	H	M	M	H	H	H	H	M	M	H	[illegible]
Graphics art	L	L	0	0	0	0	M	L	L	0	L	0	H	[illegible]
Multi-terminal	M	M	H	H	M	H	H	H	H	H	M	M	H	[illegible]
Colour Graphics	M	M	H	H	M	H	H	H	H	H	M	H	H	[illegible]
3D-helmet	M	M	H	H	M	H	H	H	H	H	M	H	H	
Ideograph	H	H	H	H	H	H	H	H	H	H	H	H	H	

Outline Specification of Possible Graphics Demonstration Programs

A computer program is required, for use at an interactive graphics display console, which would allow an uninitiated person to sit down at the console and "associate into" the screen on the basis

(i) firstly of a field of knowledge (organizations, problems, etc.) well known to him (i.e. of which he has a "model" in his head)

(ii) and, later, of the structures already built up by him on the screen which can be amended or completed.

From computer data processing point of view, it is obviously immaterial what meaning the user attaches to the entities and the relationships which he inserts -- in each case, the meaning is represented by a user selected label. There is therefore a clear advantage, in designing the demonstration program, to make it of use for entities as diverse as

(i) concepts and theoretical formulations. It can then be used with groups interested in relationships between concepts -- in knowledge structure.

(ii) organizations. It can then be used with groups interested in inter-organizational systems and in social systems in general.

(iii) problems. It can then be used with groups interested in relationships between problems e.g. in environmental systems.

(iv) personal beliefs. Though less relevant to the immediate concerns of this report, the program could also be of great use to psychologists working on the visualization of an individual's belief system i.e. a medium into which the subject can subjectively associate.

For example, in the organization case, the programs are to be used to illustrate the importance of visual display units as a means of clarifying the relationships between complex groups of organizations. Examples of such groups are (a) networks of international agencies, such as the United Nations, which has an unknown number of commissions and sub-commissions whose interconnection it is currently impossible to handle on conventional media; (b) networks of governmental agencies within any given country, where the same situation applies, particularly with regard to the difficulty of making evident cases of duplication and overlap between lower levels of different ministeries; (c) networks of business corporations and holding companies which take much study to unravel but which even then are difficult to make comprehensible.

The programs will therefore draw the attention to a new management tool for examining data bank held information to determine activities of distant departments of an organizational network and the extent of their interconnection. Specified in this way such networks can be "explored" from the known to the unknown. Where conventional retrieval systems require a key to be specific the proposed system would draw onto the screen the required item plus any other organizational units which had been directly or indirectly related to it, thus

drawing management's attention to unsuspected links. This facility becomes increasingly important as organization groups become more complex.

Exactly the same technique can be used in library systems to explore the manner in which concepts are linked in indexing systems and thus detect new key ords under hich relevant material may be held. To date, no solution has been found to the problem of showing the interrelationship of organizational, activity, geographical or conceptual entities. The visual display unit could prove to be the significant breakthrough in this area.

Demonstration program A

Specific features required are:

(i) insert entity by pointing to a position on the screen where it is to be placed.

(ii) label entity with mnemonic code which can be called onto the screen against the node. Insert explanatory paragraph or phrase of text hich can be called onto a windo on the screen by pointing at the node in question.

(iii) means of coding entity type so that entities of a given type can be called onto the screen or erased.

It should be possible to type code entities at two levels. Firstly; a "major" range of types should be selectable (e.g. A, B, C. etc. where each refers to a different coding dimension, such that A might be "organization category" and B "organization budget"). Within each major range, it should be possible to provide detailed coding

e.g. Range A = "organization category"

A1 = governmental
A2 = enterprise
A3 = academic
A4 = etc.

Range B = "$ budget"

B1 = "10,000,000"
B2 = ""1,000,000"
B3 = "100,000"
B4 = etc.

Diachronic change can be shown by arranging that one of the type ranges is a time period

i.e. C1 = "1900 - 1910"
C2 = "1910 - 1920"
C3 = "1920 - 1930"
C4 = etc.

Thus by simply pushing the "type" button, the user advances a period and has the new entities and new links added.

(iv) insert link (as per (i))

(v) insert link label (as per (ii))

(vi) insert link type code (as per (iii))

(vii) some means of building up the structure in a simulated 3-dimensional coordinate system

- either by rotating the structure and adding elements to the 2-dimensional plane so exposed
- or by using a program routine to "rearrange" the network of entities periodically so that they are clustered such that certain parts of the display are not unduly crowded.

(viii) converting to the distance matrix of which the network is a representation. (It may be possible to avoid computing this by simply editing the data structure held in memory.)

(ix) "redefine" the display so that sub-network making up any node in the main network may also be inserted.

(x) "integration calculation" by computing the inter-connectedness of the partial network constituted by entities specified with the light-pen.

Demonstration program B

Data input could be from tape onto disc or perhaps directly into memory. Format is:

	reference number	record type	cross-ref.	geographic code	numeric value descriptors
(entity 1)	54361	header	0	U.K.	A B C D E F G
	"	title	. 0		Commission on.
	"	address	0		2 Broad Street.
	"	contact P	3219	U.S.A.	S T U V W X Y
	"	"	4219	France	S T U V W X Y
	"	etc.			
	"	contact Q	1298	India	S T U V W X Y
	"	etc.			
	"	contact R	4990	Algeria	S T U V W X Y
	"	etc.			
(entity 2)	31094	header	0	U.S.A.	A B C D E F G
	etc.				
(entity 3)	52001				
	etc.				

To be effective the number of entities displayable and the number of their inter-connections (i.e. the P, Q, R type records) should be large enough to show the value of the visual display unit as a means of ordering a complex situation for convenient examination. Clearly they should not all be on the screen at the same time, but it should be possible to draw them onto the screen from memory.

Program B requirements

1. Show menu of descriptors A to G on screen. User picks three to be used as the three coordinate dimensions and specifies the range of values to be covered. Program then displays all entities on the screen according to these coordinates, giving reference numbers as identification. (This bears some resemblance to the IDIIOM molecular rotation display.)

Descriptors A to G may be verbally described (i.e. in presenting the program) as representing such means of describing an organization as: size of budget, size of personnel, assets, type (governmental, business, academic, etc.)

2. Show menu of link types P, Q and R of which one may be chosen by user. Program now draws in links between entities displayed for the type chosen (i.e. from reference to cross-reference in each case for all P types, for example). (Again this bears some resemblance to the molecular rotation program).

3. User now is able to manipulate the structure shown to clarify "hidden features". Program must permit rotation of the structure, increase in size and reduction in size. Parts going outside the screen must be chopped off. (Again this bears resemblance to the molecular rotation program.)

   This procedure allows the user to see how different organizational units are linked in terms of, for example: P, here P is interpreted as meaning a flow of funds; Q, where Q is interpreted as meaning a flow of decisions; or R, where R is interpreted as meaning a flow of information. Other such flows could be envisaged, different flo s being more significant for the four types of body likely to be interested in this application. Thus commercial organizations are likely to be more interested in share allocation, voting power, and the manner in which funds are allocated. Others will be interested in the flow of information, memberships links, etc.

4. This link display can now be refined by displaying a menu of numeric value descriptors S to Y. Users may select one of these. This causes the links to be re-displayed in terms of their significance as given by the numeric values. Two techniques may be envisaged. The links may be increased in brightness according to the values given in each record in each case. Or the links may be blinked according to a frequency governed by the value. A third possibility exists that the links might be made dotted.

   This approach enables the user to determine which are the active or important links according to particular criteria which he defines. Clearly it would be of benefit to him to alternate fairly rapidly between different descriptors S to Y, and even between different link types P to R. This might show how the funds flow was related to the decision or information flow pattern, for example.

5. As a refinement on point 1, the entities could be displayed such that the values of one of the descriptors A to G governed the number of rings (concentric) around the point, thus giving a size indication of importance.

6. One argument against this type of display for management purposes is that it lacks descriptive detail. There is however no reason why the light-pen cannot be used to indicate nodes or links on which textual comment is required. This can be retrieved and displayed on the whole screen or in an appropriate window.

7. The user should be able to work with the display to explore parts of the network not held on the screen:

(a) point to given node, program re-displays with that node at centre/origin, drawing in and pushing out parts of the network.

This is used to focus on an organization previously on the periphery of the display and to determine its contacts.

(b) Nodes as concepts or organizations, may be envisaged as having other nodes (i.e. sub-concepts or subsidiary organizations) nested within them. Such nesting could be made evident by instructing the program to "explore" a given node identified by the light-pen. This new information fills the whole screen.

(c) The converse of (b) may also be envisaged. The whole of a display may be considered as an organizational or concept system which can be considered as a node. The display may therefore be "imploded"(using a key or part of the menu) to re-display the network in terms of that node taken at the origin.

Techniques (b) and (c) can be used as means of exploring organizational hierarchies in a "vertical" direction, whilst permitting the program to remove all information from a higher or lower level of the hierarchy. Technique (a) permits horizontal exploration of hierarchies and organizational networks.

8. Visual examination of a network is not sufficient. The eye cannot always focus on or detect significant features of the network. Programs could be envisaged to perform the following:

(a) Examine a displayed network and then display a list of nodes, ranked in order of the number of links to them (i.e. most linked to organization first). This is a means of focussing on key organizations in a network. A very practical follow up is to then select those titles for which name and address list outs are required on the terminal. This permits rapid transfer from deciding that a given organization is in a key position and making arrangements to write or send something to it.

(b) Similarly, a list out could be envisaged of organizations in terms of the extent to which they function as "bottle-neck" for flows through them.

(c) If some of the descriptors are considered to cover information transfer and processing rates at and between nodes, then node pairs can be ranked in terms of the time taken for information to travel along the most direct route between them.

These techniques are extremely useful for the analysis of organizational networks as information transfer systems, prior to recommending the creation of a new organizational unit to improve the performance of the system. Much time and money is currently spent on this in a very ad hoc manner. A future development might for example permit the user to add in an organizational unit at a particular point in the network, defining its characteristics, and then recompute the characteristics of the system.

9. Additional possibilities in summary are

(a) listing of "opposite numbers" in organizational hierarchies i.e. who is concerned with a given subject in another part of the network.

(b) listing of projects undertaken by an organization (held as text), or products sold by an organization

(c) development of techniques to compute cases of overlap and duplication

(d) use of such displays in educational environment to permit exploration of national or international organizational structures in conjunction with a programmed learning type environment where a particular step in the exploration was not understood.

(e) linking use of the system to EVR for educational purposes.

Representation of Concept Networks using Graph Theory

This project is concerned with the collection of entities and the indication of relationships, if any, between those entities. Expressed in these general terms, the techniques of graph theory may be used in this project. Graph theory is concerned with the "arcs" (links or relationships) between "nodes" (entities) and the various structural properties of the network so constituted.

It can be of great assistance in dealing with a broad range of combinatorial problems which occur in various economic, sociological or technological fields. It is, perhaps, that aspect of the theory of sets which can produce the most fruitful results not only for the pure mathematician, the engineer, and the organizer, but also for the biologist, the psychologist, the sociologist and many others. Graphs can be used to represent structures such as: a network of roads, an electrical circuit, communication in a group, a complex chemical molecule, circulation of documents in an organization, kinship structures, etc. (*)

Its use in connection with relations between more abstract social entities such as organizations and nations is much less frequent (**).

(*) A. Kaufman. Graphs, dynamic programming and finite games. N.Y., Academic, 1967.

Claude Berge. Theorie des graphes et ses applications. Paris, Dunod, 1958, 277 p.

(**) Claude Flament. Theorie des graphes et structures sociales. Paris, Mouton, 1965 (English edition, Prentice-Hall)

J. Clyde Mitchell (Ed.). Social Networks in Urban Situations. Manchester U.P., 1969

Norman Schofield. A topological model of international relations. (Papter presented to Piece Research International meeting, London, 1971).

George M. Beal et al. System linkages among women's organizations. Department of Sociology and Anthropology, Iowa State University, 1967.

Robert O. Anderson. A sociometric approach to the analysis of inter-organizational relationships. Institute for Community Development and Services, Michigan State University, 1969

D. Cartwright. The potential contributions of graph theory to organization theory. In: M. Haire (Ed.) Modern Organization Theory, Wiley, 1959.

Its use for handling psycho-social abstractions appears to be even rarer (*).

The image of a 'network or web of ideas' to represent a complex set of inter-relationships in a sphere of knowledge, and particularly culture, is a fairly familiar one (**). This use of 'network', however, is purely metaphorical and is very different from the notion of a network of concepts as a specific set of linkages among a defined set of concepts, with the additional property that the characteristics of these linkages as a whole may be used to interpret the semantic significance of the concepts involved.

Some features of concept networks

Points 1 to 3 below are concerned with the shape of the network, 4 to 8 with interactions within the network.

1. Centrality. A measure (in topological not quantitative terms) of the extent to which a given theoretical entity (e.g. a concept) is directly or indirectly "related" via links to other entities i.e. the extent to which it is "distant" from another entity. One can speak of a "key" concept or of a concept being "central" to the concerns of a particular discipline. It may also be considered a measure of the degree of "isolation" of the entity. A systematic analysis of the centrality of theoretical entities could indicate where new concepts are necessary to bridge conceptual gaps and link isolated domains.

(*)Belief systems (see Appendix C2), Social science data management (see Appendix B2). In the field of documentation a thesaurus may be represented "graphically" but more for the visual presentation facility (see Appendix D3) than for any graph theoretic possibilities. For example: the "genetic maps" of the U.S. Armed Services Technical Information Agency (ASTIA), the concentric circle diagram of the Technische Dokumentatie - en Informatie Centrum voor de Krijgsmacht (TDCK, The Hague), the arrow diagrams used by EURATOM and the Bureau d'etudes van Dijk in Brussels (see Figure 1). See also the computer established "association maps" of Lauren B. Doyle. (Indexing and abstracting by association. American Documentation, October, 1962). See also: Kurt Lewin. The Principles of Topological Psychology. N.Y., McGraw-Hill, 1936; E. Zierer. The theory of graphs in linguistics. The Hague, Mouton, 1970, 62 p.; R. Quillan. Semantic memory. In: M. Minsky (Ed.). Semantic I formation Processing. Cambridge, M.I.T., 1968, p. 225-270; R.B. Banerji. A language for the description of concepts. Unpublished paper, System Research Center, Case Institute of Technology, 1964.

(**)For example, Ruth N. Anshen. "What World Perspectives means". Epilogue to Lewis Mumford. The Transformations of Man. N.Y. Collies 1962.

> "Man's situation is new and his response must be new. For the nature of man is knowable in many different ways and all of these paths of knowledge are interconnectable and some are interconnected, like a great network, a great network of people, between ideas, between systems of knowledge, a rationalized kind of structure which is human culture and human society."

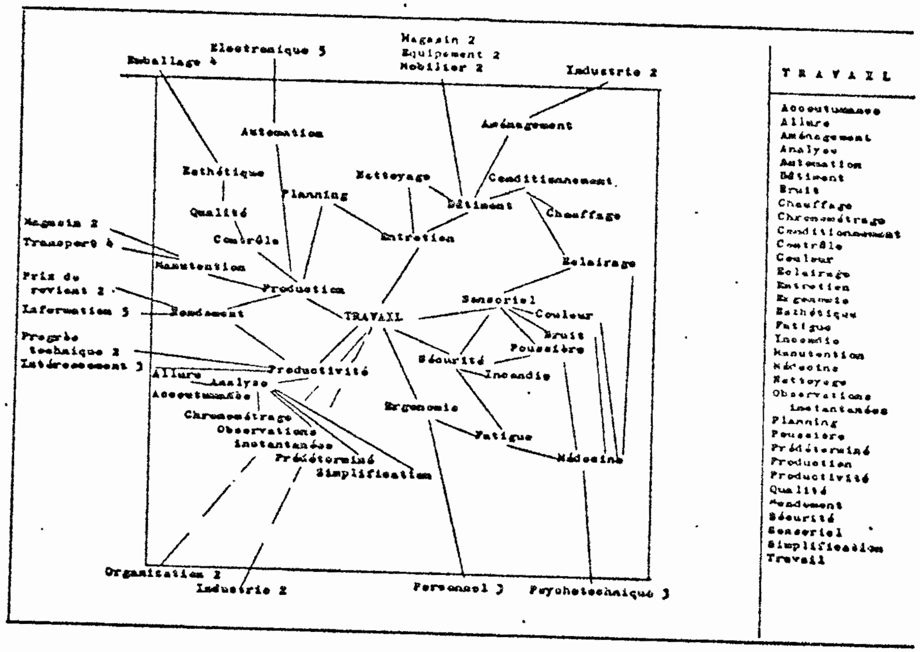

Figure 1. Arrow diagram used in special indexes

2. Coherence. A measure of the degree of "interconnectedness" or "density" of a group of concepts. This may be considered as the degree to which a system of concepts is "complete". Differ ences in density would reflect the tendency for more highly coherent concept systems to appear more self-reinforcing in comparison to less organized parts of the network. In some respects this is an indication of the degree of "development" of a group of concepts.

3. Range. Some concepts are directly related to many other concepts, others to very few. The range of a concept is a measure of the number of other entities to which it is directly related.

Range could be considered an indication of the "vulnerability" of a concept, to the extent that a high range concept would be less vulnerable to attack than a low range concept, since it has more bonds anchoring it to its semantic environment. High range points are therefore either key points in resistance to conceptual change or else key points in terms of which orderly change can be introduced.

4. Content. The "content" of a relationship between entities is the nature or reason for existence of that relationship. The different types of relationship are covered in Appendix A5 . In general; different relationship contents are required for each model (see Appendix A3).

Simple graphs have only one link between any two entities; multigraphs have two or more links, each of different content.

5. Directedness. A relationship between two entities may have some "direction" i.e. A to B, or B to A. The different types of directedness for different models is described in Appendix A3. The most important for this project is probably: A "is a subset of" B, i.e. directedness points to the more fundamental concept of a pair. In a multigraph, one link may point from A to B and the other from B to A -- where each is more significant in terms of different content.

6. Durability. A measure of the period over which a certain rela- ship between entities is activated and used. In fact, any entity may be considered, by someone, to be linked to any other. At one extreme, there are the links activated only on a "one-shot" basis (e.g. a "trial balloon" idea), at the other there are links, and sets of links, which are considered stable over centuries (e.g. the concepts associated with "property"). Links may be considered stable and durable, unstable and short-lived, and metastable. Metastable links are those which would disappear if the appropriate arguments were brought to bear -- but otherwise persist as a localized abnormality. Durability is clearly important for historical models (see Appendix A3).

7. Intensity. A measure of the strength of the link or bond between two entities. Two concepts may be said to be "strongly bound together". In some models, the intensity is a measure of the amount of the "flow" or "transaction" between the entities (see Appendix A3).

   The link from A to B may be strong, and that from B to A, weak.

8. Frequency. A link between two entities may only be established intermittently. This measure is less significant to this project (except perhaps in cyclic approaches to the history of ideas or to the activation of concepts over a 24 hour period.)

9. Rearrangeability and blocking. A connecting network is an arrangement of entities and relationships allowing a certain set of entities to be connected together in various possible combinations. Two suggestive properties of such networks, which are extensively analyzed in telephone communications (*), are:

   rearrangeability: a network is rearrangeable, if alternative paths can be found to link any pair of entities by re-arranging the links between other entities.

   blocking: a network is in a blocking state if some pair of entities cannot be connected.

Examples of types of network patterns

Some of the above features of networks of concepts (or other entities) may be illustrated by the set of diagrams in Figure 1. Each entity is represented by a letter of the alphabet. Four simple types of entity groups are shown. Each type is further distinguished if the relationships between entities are directed.

a) In the non-directed examples of group (1), A is the central concept in (1.3), A and D in (1.4), A and F in (1.2). In (1.1), there is no central concept.

b) In group (1), peripheral concepts are D and C in (1.2); B, C, E and F in (1.3); B, C and F in (1.4). There are no peripheral concepts in (1.1).

c) In group (1), the range of A in (1.3) is 4, in (1.4) it is 3.

d) In group (1), the reachability of A in (1.1) and (1.2) is 3, in (1.3) it is 1, and in (1.4) it is 2.

(*)V.E. Benes. Mathematical Theory of Connecting Networks and Telephone Traffic. N.Y. Academic, 1965, p. 53

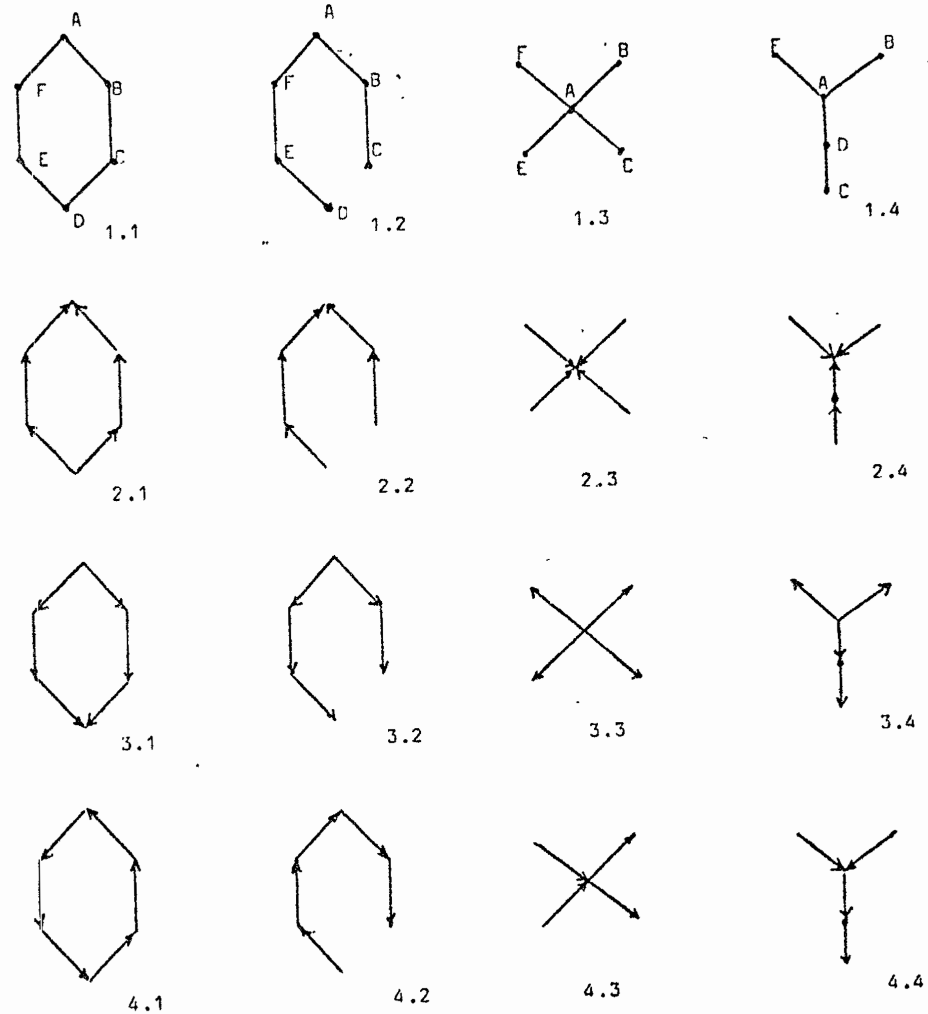

Figure 2 Examples of simple networks of entities

e) In all the directed examples of group (2), A is the central concept with at least B and F as direct component concepts (*). In all except (2.3), there are even sub-sub-components of A.

f) In all the directed examples of group (3), A is the central concept but only as a common sub-component. D is also a common sub-component in (2.1).

g) In all the directed examples of group (4), there is a chain of component/sub-component links. In (4.1), this is continuously forming a loop. In (4.2) and (4.4), C is the major concept. In (4.3), A is the central concept but only by having F and E as sub-components and being itself a common sub-component to B and C.

The above features are all evident, almost to the point of being trivial. But most cases of interest are likely to be much more complex, with many nested levels of concepts and cross-linking relationships. These may be examined by matrix analysis techniques, particularly using computers (to which the proposed record layout is suited) (**). Computer programs exist to detect properties of networks.

Specific reference is made to the use of network techniques in domains related to this project. Mention has been made of citation indexing (Appendix D3), artificial intelligence (Appendix C2), personal construct theory (Appendix C3).

(*)The set/subset relationship is used to illustrate these directed examples but other meanings are also possible (see Appendix A3). In particular, time order of formulation of concepts, and cybernetic information flows between problem areas or within organizational networks.

(**)C. Berge. The Theory of Graphs and its Applications. London, Methuen, 1962.

C. Flament. Applications of Graph Theory to Group Structure. Englewood-Cliffs, Prentice-Hall, 1963.

F. Harary and R.Z. Norman. Graph Theory as a Mathematical Model in Social Sciences. Ann Arbor, University of Michigan, 1953

F. Harary, R.Z. Norman and D. Cartwright. Structural Models: an introduction to the theory of directed graphs. N.Y., Wiley, 1965

Relationship to Artificial Intelligence Projects

In considering the possibility of coding definitions of concepts, propositions and like entities, it is important to benefit as much as possible from related work on artificial intelligence, and possibly pattern recognition. Artificial intelligence and projects to simulate human personality or belief systems have had to develop methods and computer techniques which can handle and interrelate entities such as concepts and propositions. Clearly the object of such projects is not attained once an inventory of entities can be examined, even if it is highly structured in the form of a thesaurus. It is therefore interesting to look at both the techniques used to handle concepts and the types of computer-based interrogations that are then possible.

The suggestion that techniques of handling individuals' "beliefs" about interpersonal relations should have some parallel to a community of scholars' attitudes towards the concepts, propositions, etc., which constitute its territory, may appear somewhat provocative. Does a school of thought constitute a belief system?

T.S. Kuhn (*) uses the terms "belief","metaphysic","commitment", and "conversion" in connection with a scientific community's attitude towards a paradigm and paradigm change. He mentions a non-scientific relationship between the community and its current paradigms. In addition, the direct structural relations between belief and thought are conveniently summarized by Milton Rokeach (**).

It is not necessary to go into this point, however, because it is only the insights concerning the approach to handling highly-interrelated "entities" which are of immediate interest.

The points made in this section are a summary, mostly direct extracts, of a paper by Kenneth M. Colby and colleagues (***). Colby is a psychiatrist working on the Stanford University Artificial Intelligence Project.

There exists a class of problems in the behavioral sciences that

---

(*) T.S. Kuhn. The Structure of Scientific Revolutions. Chicago, University of Chicago Press, 1962.

(**) Milton Rokeach. The Open and Closed Mind; investigations into the nature of belief systems and personality systems. New York, Basic Books, 1960. (see pages 16-19, and Part III)

(***) L. Terler, H. Enea and K.M. Colby. "A directed graph representation for computer simulation of belief systems." Mathematical Biosciences, 2, 1/2, Feb. 68, 19-40.

has been difficult to manage satisfactorily with information-processing methods because of a lack of a good computer representation for very large memory structures. One example is the abstract representation linguists term a "deep structure" into which natural language is translated and to which transformational grammar is applied in generating natural language sentences. Another example consists of the large data bases required in computer simulation of human belief systems.

The paper describes a directed graph used for the representation of the data base of a computer model that simulates the formation and processing of an actual person's or an artificial system's beliefs about interpersonal relations. The graph constitutes a formal structure capable of abstractly representing the great variety of semantic relationships found in human concept and belief systems.

The basic component of the data base is the abstract entity "concept". Examples of concepts are: parents, fear of women, old men, hating authority, John, hatred.

Kinds of concepts used are: sets, individuals, and propositions.

A "belief" in the model is an attitude toward a proposition about concepts. It is convenient to regard a proposition as a special case of a concept. A proposition is considered to have one of two functions -- to represent a "fact" or to impart a "rule".

Not all propositions need to be beliefs in the model. The degree to which the model is willing to accept a proposition is called the "credibility" of the proposition, while the degree to which evidence substantiates a proposition iscalled its "foundation". Credibility and foundation are useful criteria determining whether a proposition is rejected or retained. They are measured on arbitrary scales from 0 to 100. In addition, without regard for their status as beliefs two propositions can still differ only in "intensity"; for example, "John strongly believes in 'x'", and "John weakly believes in 'x'".

Different concepts in a model can vary in thir importance to the train of thought and a single concept can vary in its importance from time to time. The attribute measuring their differences is called "charge", e.g., "sex" may be a permanently charged concept, whereas "washing the dishes" might be temporarily charged.

Other distinctions can be made between concepts on the basis of their "longevity" (how long ago formed) and "inhibition" (tendency to be avoided in communication and reasoning).

In addition to quantitative measures, concepts have qualitative aspects. The "kind" (set, individual, or proposition) and the "origin" (a priori, observed, or reasoned).

Concepts are represented in the model by nodes of a directed graph. Simple relationships are represented by directed arcs between pairs of nodes. Each arc is labelled e, s or p, depending on the type of relationship. The types of arcs are distinguished by their formal properties, but notions of their approximate meanings can be outlined:

A e B = individual A is a member of set B
A s B = set A is a subset of set B
or proposition A is a consequence of proposition B
B p A = A has B, or A has property B, or B belongs to A, or B is part of A, or the idea of A suggests the idea of B, or A does B

The same mode can be an individual, set, or proposition in different "contexts". The formal properties distinguishing the three types of arcs are given by seven axioms of valid graph enlargement.

Axiom 1	A s A	s is reflexive
Axiom 2	A s B BsC — AsC	s is transitive
Axiom 3	AeB BsC — AeC	A member of a subject is a member of the set
Axiom 4	AsB CpB — CpA	A property of a set is had by its subsets
Axiom 5	AeB CpB — CpA	""
Axiom 6	BpA BsC — CpA	Having a specific property implies having the more general property
Axiom 7	BpA BeC — CpA	""

There are two ways to look at a directed graph: locally and globally. Local examination implies that examination begins at some node and proceeds only by following the arcs (in either direction) that touch that node. Global examination requires "stepping back" from the graph and looking for patterns. Three basic methods of graph (or tree) searching are available in sequential processing: depth-before-breadth, breadth-before-depth and a partially random approach. Special computer programs have been developed to do this.

Comment.

It is clear that with the introduction of such additional features as "credibility", "foundation", "charge", "longevity", "inhibition", "mode", and "origin", a much more dynamic picture of the belief system emerges. For each of them, an equivalent exists within a discipline's conceptual world, but whether it would be possible or useful to attempt to incorporate all such information is another matter. At first sight, it would be particularly appropriate to attempt to do so for the educational or historical model types.

Use of computer models of belief systems.

Once a data base exists, it is possible to interrogate it and discover what its beliefs are on particular topics (*). More ambitiously, it is possible to enter into dialogue with it, such that it will examine and accept or reject propositions made, and thus extend or modify the data base on the basis of the credibility of the informant. (**))

Work in this area is relatively advanced, although bound by important constraints due to simplifying assumptions. (As an illustration, a "paranoid" model permitting natural language dialogue has been constructed by altering the sensitivity of the model to statements on certain topics in terms of three scales of "fear","anger", and "mistrust". (***)

Comment.

There is no technical reason why the concepts and propositions of a given discipline should not be handled in this way. The fact that they do not all tie together into a consistent, coherent whole represented by a monolithic hierarchy of concepts is no obstacle. Individual belief systems are not consistent or coherent, either. (The standard sentence forms used in such models can be taken not as crude approximations to English verbalization, but as quite general representations of properties and relations among objects. Such relations and properties can be expressed in a variety of symbolic forms other than verbal. Thus the forms in which the model's beliefs are cast can be seen as general and powerful cognitive schemas, and not merely as exercises in Dick-and-Jane prose. (****)

(*) K.M. Colby, L. Tesler, H. Enea."Experiments with a Search Algorithm on the Data Base of a Human Belief Structure. " Stanford University, Artificial Intelligence Project, 1969, (Memo AI-94).
John C. Loehlin. Computer Models of Personality. New York, Random House, 1968.

(**) K.M. Colby and D.C. Smith. "Dialogue Between Humans and an Artificial Belief System." Stanford University, Artificial Intelligence Project, 1969. (Memo AI-97)

(***) K.M. Colby, S. Weber and F.D. Hilf. "Artificial Paranoia." Stanford University, Artificial Intelligence Project, 1970, (Memo AIM-125).
K.M. Colby, and J.P. Gilbert. "Programming a Computer Model of Neurosis". Journal of Mathematical Psychology, 1964, 1, 405-417.

(****) J.C. Loehlin, op.cit. p. 111.

In thinking of the application to schools of thought, it is interesting to note the comment made by Colby and his colleagues on the status of the model held by the computer.

> Our problem -- how to construct a good model of the informant's belief process. The criteria for "good" can be varied --are we getting at what the informant "really" believes? What "really" means here is obscure, but it is common knowledge that people have limited accessibility to their beliefs at a given moment. Even worse, they have the capacity to deceive themselves, to rationalize, and to distort their own beliefs ...In worrying about what is "really" believed, we found it useful to keep in mind that we were constructing a model of a model. A belief structure is a representation,and in giving information about himself, an informant tells us what he believes he believes. He simulated himself and it is his accessible model of himself that becomes the data base of a computer model. Humans' ability to simulate themselves and to make models of other models is of course a most interesting property for a sumbolic system to have. (*)

Elsewhere he notes that a belief system (like a school of thought) is itself a model (if only partial) of the universe. He considers that reasoning processes are aided in the individual by his simulation of his own mind -- by "autosimulation". It might be useful for disciplines to examine their own conceptual structures in the same way as an aid to the development of the discipline. It could be particularly important as a means of highlighting tensions within the conceptual structures which lead up to Kuhn's paradigmatic changes.

This approach suggests a number of stages of sophistication in the possible development of this project.

1. A static inventory of concepts and propositions
2. A static network of interrelated concepts and propositions
3. "Activation" of propositions as rules governing the relationships between entities
4. Treatment of a school of thought as a belief system
5. Extension to natural language interaction

On this last point, it may be possible to allow a (non-computer-oriented) specialist in a particular field to "dialogue" with the concept data base to permit him to discover and indicate where he differs from its contents and what new he thinks should be included (**) This approach might be a useful method of getting

---

(*) K.M. Colby, L. Tesler, H. Enea, op.cit., page 9.

(**) See: K.M. Colby and H. Enea. Heuristic method for computer understanding of natural language in context-restricted on-time dialogue. Mathematical Biosciences, 1,1-25, 1967.

around the behavioural problems associated with the power position of official classifiers in committees.

It is even possible to have many such people interacting in natural language with the data base via terminals to facilitate communication (e.g., at a special seminar).

This last stage raises the serious problems of "deep structure" in linguistic analysis, the formalization of natural language semantics within the limits of particular scientific dialects, the need to relate such languages to one another through semantic and symbolic manipulations and the question of mechanized translation. These difficult problems are avoided or by-passed in the early stages because it is not the "words" in the "surface structures" which are coded but the meanings of these words as "terms". In other words, this project is significant because it attempts to code the elements of the deep structure directly and in a manner which avoids the verbiage (and of course much richness) of the "carrier wave" functions of natural language. Efforts at mechanized translation seem to be attempting to translate natural language into terms. This approach treats terms as conventional labels but not as the goal of translation. Consultation at an early stage with specialists in these areas would be vital.

Future

As a future development of the application of techniques of simulating belief systems to simulating schools of thought, one can envisage the possibility of an individual being able to interrogate, or dialogue with, different schools of thought, each represented in a model. The individual could compare responses and examine their incompatabilities. This might have some application in the policy sciences where experts from different disciplines in effect each submit different models of a problem situation and its solution. It would however be particularly useful as an educational tool and for interdisciplinary research. (The individual is here an active participant in the dialogue. Of possible interest would be a "dialogue" between computer models of two or more disciplines sparked off by a topic selected by the individual to be educated. This would probably be of more value as an investigation of belief system behaviour under threat to conceptual territory.)

Other references:

C.M. Eastman. In: H.S. Brinkers (Ed). Decision-making; creativity; judgment, and systems. (in press).

F.H. George. "Formation and analysis of concepts and hypotheses on a digital computer." Mathematical Biosciences, 3, 91-113,1968.

K.M. Colby. "Computer simulation of change in personal belief systems." Behavioral Science, 12, 248-253, 1967.

R.P. Abelson and J.D. Carroll. "Computer simulation of individual belief systems." American Behavioral Scientist, 1965, 8, 24-30.

M.R. Quillan. Semantic Memory. Bolt, Berenek and Newman. DDC Report AD 641-671, Oct.1966. Also published in:M. Minsky (Ed.) Semantic Information Processing. M.I.T.,1968.

Relationship to Personal Construct Evaluation Techniques.

There exists a school of thought in psychology concerned with the evaluation of "personal constructs" (*). The arguments in this Appendix are based on extracts from the most recent book summarizing the field (**). A "construct" is the basic contrast between two conceptual groups. When it is imposed, it serves both to distinguish between its elements and to group them. Thus the construct refers to the nature of the distinction an individual attempts to make between events, not to the array in which his events appear to stand when he gets through applying the distinction between each of them and all the others. A construct system is made up of nothing but constructs, and its organization is based on constructs about constructs, which may be set up in concretistic pyramids or abstractly cross-referenced in a hierarchical set of relationships.

The systep of constructs which a person establishes for himself represents the network of pathways along which he is free to move. When a person must move, he is confronted by a series of dichotomous choices -- each choice being channelled by a construct. Each construct represents a pair of rival hypotheses, either of which may be applied to a new element which the person seeks to construe. The construct system sets the limit beyond which it is impossible for the person to perceive. Many constructs have no word labels and represent nonverbal and preverbal bases of discrimination and organization, and these may occupy important and even central places in the economy of a person's orientation towards himself and the world.

The construct system is evaluated using a grid-based interview method, which results in a matrix giving the interrelationships between the elements of the system. This matrix can then be scanned by computer to highlight clusters. A number of computer programmes have been developed for this purpose.(***) Slater (****) has prepared a program which accepts grids cast in

(*) G.A. Kelly. The Psychology of Personal Constructs. New York, Norton, 2 vols., 1955.

(**) D. Bannister and J.M.M. Mair. The Evaluation of Personal Constructs. London, Academic Press, 1968.

(***) For a summary see: J.C.J. Bonarius. "Research in the Personal Construct Theory of George A. Kelly." In: B.A. Maher (Ed.) Progress in Experimental Personality Research. London, Academic Press, 1965, vol. 2.

(****) P. Slater. The Principal Components of a Repertory Grid. London, Vincent Andrews, 1965.

________. "Notes on Ingrid 67." London, Biometrics Unit, Maudsley Hospital, 1967.

any form and is a form of principal component analysis. This analysis delineates significant orthogonal structure both of constructs in relation to elements, and of elements in relation to constructs. Thus a fairly detailed overview can be obtained in mathematical terms, and this can be examined visually in terms of a hypersphere which represents a person's psychological space as subsumed by grid method.

Comment

The authors of the volume, from which the above extracts were obtained, point out that "...we have presented grids with only one type of element -- namely, people. The limitation was accepted for the sake of simplicity in presentation. We cannot too strongly emphasize that the content of grids (the constructs and elements) is, for practical purposes, very variable." (p. 72)

In their discussion of uses other than interpersonal relationships, they mention only relations between individuals and such elements as films, paintings, inanimate objects, emotions, problem situations in a person's life. But Kelly himself points out that "not only can the grid notion be generalized to all conceptualizations, but this mathematical notion can also be generalized. The incidents and voids which populate a grid of intersects provide the binary numerical basis for a mathematics of psychological space... Thus we may have a mathematical basis for expressing and measuring the perceptual relationship between the events which are uniquely interwoven in any person's psychological space." (op.cit. 301-2)

On this basis, therefore, it would seem that this technique could be applied to determine the constructs used by a school of thought or a discipline in ordering its own perception of significant elements in its world view.

One advantage of this approach is that it does not necessarily impose any content dimensions on the subject (in this case a school of thought). Each subject can be encouraged to express the constructs which he uses to make sense of areas of his life. In a sense the resulting picture is culture-free and sub-culture free in that the subject has been allowed to work in terms of his own preferred language -- rather than in terms of an "alien tongue" chosen by the investigator. This lack of an imposed language is most essential to any proposed effort to handle the concepts of different schools of thought -- if only to avoid any form of conceptual imperialism.

It is interesting to see that these same authors reflect some of the preoccupations of Fred Riggs and G. Sartori (*).

(*) Fred Riggs. Words, concepts and terminology. Hawaii, Social Science Research Institute, University of Hawaii, 1971 (COCTA Working Paper n° 1.

G. Sartori. Concept misformation in comparative politics. American Political Science Review, 64, n° 4, December 1970, p. 1033-1053.

"Many constructs are symbolized by verbal labels -- a word or a group of words. All words in general usage in any language have commonly-agreed dictionary meanings; but individuals may often use similar words to describe different experiences or ideas, or different words to describe similar experiences or ideas. Almost all psychological measures dependent on words have relied heavily on the assumption that different people will understand broadly the same thing where a standard set of words is used (e.g.,in a questionnaire) and will mean the same thing when they reply in some standard form. Grid method does not assume that the subject means what the experimentor means by particular verbal labels involved in the test -- on the contrary, the method is designed to help ascertain what the subject means by particular verbal labels....It is then possible to compare their personal "meanings" for words either with their public meaning (the construct interrelationship implied by dictionary definition, or normative relationships yielded by grid administration to groups), or with the experimenter's meanings (either by having the experimentor complete a similar grid, or by having him predict the construct relationships which would reflect his particular explanatory stance)."
p.143)

Clearly this approach could be used by a modelling body:
(a) to obtain a systematic check on the degree of consensus amongst its members
(b) to interview members of the school of thought on particular concepts, propositions, etc., their interrelationship and their importance. In this case, the individual grids are averaged by computer to obtain the dominant clusters (which would seem to be the beginnings of a fairly "democratic"modelling system).

The authors note a major disadvantage of the grid method: "...it is already apparent that the original binary grid and its more recent variations cannot adequately subsume all the ingenious and sometimes contorted forms of constructing which men have undertaken. Not least among its limitations is its fixity in expressing only one type of linkage between constructs (the reciprocal linkage represented by a unitary index of associations), and its failure to incorporate some important aspects of construct theory." (p.74)

Finally, it is very instructive to examine the formal description of personal constructs theory replacing "person" and "user" by "school of thought" or "discipline" (see Figure , reproduced from Bannister and Mair, op.cit.)

Use of Input/Output analysis (*)

Network analysis is closely related mathematically to input-output analysis which has been used for some years by economists to analyze the trading transfers between different sectors of industry.

> "It has long been recognized that in the economy of any town, city, state or nation, each business depends on products and services of other industries in order to produce products or services of its own. This interdependence of industries within an economy is entirely obvious, but difficult to measure, and becomes more difficult as the economy becomes more complex and more mature. The "square matrix" of interindustry transactions -- which shows these interdependencies and measures them for a given period -- is, in combination with electronic data processing, becoming a valuable basis for future economic planning for business, industrial firms, and governments -- local, regional or national. For both sudden and gradual changes in industrial, government of consumer areas of supply and demand alter all other relationships, and individual companies stand to profit or suffer in the transition....Application of Input/Output to marketing problems assures improved information generated through the use of a systems approach: analyzing a problem in relation to the whole economy, rather than as a series of unrelated cases." (**)

It is quite clear from this that interdependence of industry sectors and the constituent enterprises has been widely recognized. This recognition is of course limited to interactions detectable from an economic perspective. The same principle applies, however, to all interactions (funds,information, goods, etc.) concerned with all subject areas (development, environment, education, etc.). This is not generally recognized.

It is interesting to note that Wassily Leontief, who developed the input-output technique, now foresees that input-output tables might be expanded to quantify the byproducts with which the various industries pollute the atmosphere. He considers this would lead to a sharp understanding of the connections between economic processes and the environment and thus help to solve this major problem in the developed countries, namely the rapid deterioration in the quality of life (Business Week, 22 November 1969, p.126).

> "The unique service of input-output analysis is its ability to give a detailed picture of the industrial structure by putting numbers on all the complex interconnections that link the various sectors of the economy." (*)

(*) This appendix is based on extracts from: Jere W. Clark and A.J.N. Judge. Development of transdisciplinary conceptual aids. Brussels, Union of International Associations, 1970.

(**) Facts on Fortune's 1966 Input/Output Matrix -- Computer-age Tool. p.2-5. See also: Wassily Leontieff. Input-output analysis. Oxford U.P., 1967.

The ability of this technique to highlight interdependencies and weaknesses in a system of producers and consumers of goods (represented by their funds' equivalent) suggests a similar use to highlight interdependencies between different sectors of the psycho-social system. In this case it is necessary to deal in terms of producers and consumers of information --in its broadest sense. This technique could then be used with the cybernetic models (Appendix).

The situation becomes very complex since the table or network becomes multidimensional. There are many methods of avoiding these problems and obtaining new insights. As an example, an "information map" in input-output table form was developed for the State of California by concentrating on information flows. A survey was carried out to indicate "every instance where information was exchanged between a particular organization and the State government and the local government." These interchanges were shown by means of a code on an input/output table covering all of the State organizations, cities, counties, Federal Government agencies, and private enterprises. Aside from giving an overview of the State information network, the table highlighted cases where one group of organizations needed information from another group but could not obtain it because it was not available. (**)

It might be possible to employ the same technique to handle information between disciplines and thus provide one aspect of the interdisciplinary chart mentioned by René Maheu, Director-General, UNESCO:

> "One of the most significant results that should naturally emerge from a study such as this, is the preparation of a chart -- admittedly provisional and subject to constant revision -- of the strong points and weak points of interdisciplinary cooperation and of their substratum, and the identification of priority areas to which research scientists should direct their thinking and institutions their activities." (***)

(*) Business Week, 22 November 1969, p. 125.

(**) Hearings before the Special Subcommittee on the Utilization of Scientific Manpower of the Committee on Labor and Public Welfare, United States Senate, 89th Congress,S.2662, 1965-66, p. 35-38.

(***) René Maheu in the preface to: Main Trends of Research in the Social and Human Sciences. Part one: social sciences. Paris, Unesco, 1970, p. xv.

Appendix D1

Related and Earlier Attempts at Concept Coding

There have been many attempts at isolating and classifying elements of meaning at the root of complex concepts. De Grolier (*) notes that methods and the need for them have been regularly discovered and rediscovered since the time of Leibniz or even earlier. He then states: "We draw attention to these 'anteriorities', not in order to underrate the work performed by the various researchers or teams of researchers -- who, in most cases, truly believed that they had discovered a 'new method' -- but to persuade them, rather than to advocate unilaterially any one 'exclusive' process, to agree that they are all engaged in work on common basic principles, whatever may be the differences (at times very minor) in the coding method or the particular type of machine adopted."

De Grolier has summarized the work on classification around the world but only a few initiatives seem to be directly related to this project. Usually the work has been directed towards solving a classification problem in some particular field which strongly influences the design of the scheme (see Appendix A2). The following, noted by de Grolier, is of more direct relevance:

1. Perry and Kent (Western Reserve University)
   Developed a coding method for the field of metallurgy based on 'semantic analysis' of complex terms into 'individual terms'. 30,000 terms were assembled from a variety of sources. The notation is however very cumbersome.

2. S.M. Newman (U.S. Patent Office)
   A "vast attempt at defining or redefining concepts, which could perhaps be entitled -- to paraphrase a famous title -- 'In search of lost simplicity': to discover or rediscover non-equivocal terms beyond the complications of natural language, which 'unfortunately' does not have "uniform or logical rules for the denomination of devices or things". In effect this is an attempt at creating a metalanguage -- but again results in a cumbersome notation.

3. C.G. Smith (U.S. Patent Office)
   Suggested a system which would isolate "ultimate concepts... required in the definition of more specific concepts....There is a basic layer of concepts which do not require definition. It is the use of such elemental concepts which is contemplated in the present system....A fundamental feature is to seek beneath composite words the basic organization of elemental concepts which they represent, and to develop the essential combination for the definition of these words."(**) This was conceived mainly for patentable contrivances on the US Patent Office Interrelated Logic Accumulating Scanner. It does however permit chains of related concepts to be handled.

4. Cordonnier
   Worked on methods "to symbolize the elementary points of view of the classification of ideas and...to study the grouping of these symbols in order to obtain composite symbols representing the structure of complex concepts". He also suggests that "intuition permits the representation in an intellectual space of a logical figure, to n dimensions, a synthesis of the relationships between a group of ideas into the different classes which arrange them naturally according to the various possible individual viewpoints".

5. M.E. Stevens
   Worked on use of computers to handle interrelationships betgw terms and to 'define', by supplying the generic and descriptive terms related to the term of which the definition is sought; 'develop', by furnishing specific examples of a generic term; 'localize', by indicating the place which can be associated with the proposed concept; 'match', by comparing several proposed terms together, in order to find a 'common point' making it possible to relate to these terms another term possessing the same characteristic; and carry out other logical operations.(*)

6. Others
   Other initiatives and their relationship to this project are reported in separate appendixes. Of particular interest is the highly general approach adopted by the M.I.T. ADMINS system (Appendix B2).

Conclusion

Most of these attempts appear, from the perspective of this project, to fall foul of one of the following difficulties: cumbersome notation, rigid and exclusive category or relationship structure, focus on one specialized field of knowledge, difficult to implement because the administrative and intellectual tasks are not distinguished.

(*)Eric de Grolier. A Study of General Categories applicable to Classification and Coding in Documentation. Paris, UNESCO, 1963.

(**)C.G. Smith. Descriptive documentation, International Conference on Scientific Information, 1958; Proceedings. Washington, National Academy of Sciences, 1959, p. 1103

(*) Mary E. Stevens. A machine model of recall. Paris, UNESCO, NS/ICIP/J.5.4, 1959. See also: T.Kilburn, R.L. Grimsdale and F.H. Summer. Experiments in machine learning and thinking. Paris, UNESCO, NS/ICIP/5/6/15, 1959.

Conceptual dictionaries (*)

The outstanding importance of dictionaries in the modern world explains why some lexicographers are dissatisfied with the mechanical method of arranging words in alphabetical order, and would prefer to classify them according to the concepts which they express.

One would be mistaken in believing that this is a recent trend, since, as Franz Dornseiff reminds us in his remarkable 'Vorrede' (Preamble), Der Deutsche Wortschatz nach Sachgruppen (The German vocabulary by subject groups), Berlin, W. de Gruyter, 1934, one finds tentative systematic vocabularies at Babylon in the third millennium before Christ. In modern Europe, the most important work on systematic lexicology was that of Peter Mark Roget, in the nineteenth century, the Thesaurus of English words and phrases, which he was in process of preparing as early as 1806, and of which the first edition appeared in 1852; see also among numerous contemporary editions that of Penguin Books, London, 1953. Concerning Roget, see Henry Sweet, "Words, Logic, and Grammar", Trans. Philological Soc., 1875-76, p. 470-503, reproduced in his Collected Papers, p. 1-33. Oxford, Clarendon Press, 1913.

The Seventh International Congress of Linguists put "conceptual dictionaries" on its agenda (point A5): see F. Mezger's report, p.77-85, the contributions, p.86-9, and the discussion p.443-73. One of the most remarkable ideological dictionaries was that of J. Casares, Diccionario ideologico de la lengua espanola (Ideological dictionary of the Spanish language), Barcelona, 1942; see also his Introduccion a la lexigrafia moderna, Madrid, 1950. Under the impulsion of Antoine Thomas, a certain number of French dialectological studies were made on a systematic basis, e.g. L. Lhermet, Contribution à la lexicologie du dialecte aurillacios, Paris, 1931. It would obviously be a great convenience if conceptual dictionaries of different languages, periods, or single authors could conform to the same general pattern so that they could be readily compared with one another. To this end one would require a conceptual framework so comprehensive and yet so elastic that the most diverse languages and the most idiosyncratic writers would fit smoothly into it. Such a broad classification of concepts was put forward by R. Hallig and W. von Wartburg in 1952.

Walther von Wartburg constituted himself the protagonist of the general application of such a method, see his report "Das Ineinandergreifen von deskriptiver und historischer Sprachwissenschaft (1931); "Betrachtungen über die Gliederung des

(*) This Appendix consists of extracts from S. Ullmann, Semantics: an introduction to the science of meaning. Oxford, Blockwell, p.254-5; but especially from Eric de Grolier. A Study of General Categories Applicable to Classification and Coding in Documentation. Paris, Unesco, 1962, p.226-228 (Note 89).

Wortschatzes und die Gestaltung des Wörterbuchs' (The interaction of descriptive and historical linguistics (1931); views on the structure of the vocabulary and the formation of the dictionary). Mélanges Bally, 1939; Problèmes et méthodes de la linguistique, p.159-67, Paris, Presses Universitaires de France, 1946; Rudolf Hallig and W. von Wartburg, "Begriffssystem als Grudlage für die Lexikographie; Versuch eines Ordnungsschemas" (The system of concepts as the foundation for lexicography; a tentative system of arrangement), Abhandlungen der deutschen Akademie der Wissenschaft zu Berlin, Klasse für Sprache, Literatur und Kunst (Proceedings of the German Academy of Sciences in Berlin, Class for language, literature, and art), 1952, n° 4. See on this subject W. Runkewits, "Kritische Betrachtungen zum Begriffssystem von Hallig v. Wartburg im Zusammenhang mit den Arbeiten am Altgaskognischen Wörterbuch" (Critical remarks on the system of concepts of H.v. Wartburg in connection with the work on the Old Gascon dictionary), Monatsberichte der Deutschen Akademie der Wissenschaften zu Berlin, 1946-56, p. 379-88. See also recently F. de Tollenaere, "Lexicographie alphabétique ou idéologique," Cahiers de lexicologie, n°2, p. 19-20.

See especially K. Baldinger Die Gestaltund der wissenschaftlichen Wörterbuchs", Romantisches Jahrbuch, v (1952), pp.65-94, K. Baldinger, "Grundsätzlicher zur Gestaltung des wissenschaftlichen Wörterbuchs", Deutsche Akademie der Wissenschaften zu Berlin, 1946-1956, Berlin, 1956, pp. 379-88; K. Baldinger, "Alphabetischer oder begrifflich gegliederter Wörterbuch?". Also Julius Gould and W.L. Kolb. A Dictionary of the Social Sciences (compiled under the auspices of Unesco). New York, Free Press of Glencoe, 1964. (discussed in Appendix D9).

The German research on "semantic fields", which later inspired Georges Matoré, La méthode en lexicologie, domaine français, (Method in lexicology in the field of the French language), Paris, Didier, 1953; he offers here (p.70-4) a diagram of a "comprehensive classification of lexicon facts" different from that of Hallig and Wartburg, and, moreover, less satisfactory. It will be noted that Matoré, in defining lexicology as a sociological discipline using words as its linguistic material, tries to make of it an "autonomous discipline" the field of which partly covers that of linguistics, but independently of it: for reasons other than those of certain American structuralists this position results in a dismemberment of linguistics in a way which does not seem to be any longer justifiable (p.50-1) Concerning structural semantics in general, and its (desirable) relationships with other parts of structural linguistics, see S. Ullmann in the second edition of his Principles of semantics, p.307-21, op.cit., with numerous references, and Uriel Weinreich, "On Semantic Universals" (duplicated, 71 pp. March 1961, with an important bibliography), and also his programme of studies,"Semantic structure of natural languages" (duplicated memorandum, 5 May 1961). At the eighth International Congress of Linguists (Oslo, 1957) there was a (rather disappointing) discussion on the subject "To what extent can meaning be said to be structured? (p. 636-704 of the Proceedings), of which the most interesting item was the paper by Hjelmslev, which we have already mentioned. See also Hans Pollak, "Gibt es Wortklassen vom Standpunkt das Hedcutting?" (Are there word-classes from the point of view of meaning?), Beiträge zur Geschichte der deutschen Sprache und Literatur (Tübingen) 80, 1958, p. 33-47.

There is also Andrew Paul Ushenko, The field theory of meaning. Ann Arbor, University of Michigan Press, 1958; F.F. Nesbit, Language, meaning, and reality, New York, 1955; H.R. Walpole, Semantics, Norton, 1941. On basic concepts, see Eric H. Lenneberg and John M. Roberts, The language of experience, supplement to the International journal of American linguistics, 22, (2) Charles E. Osgood and his collaborators have tried to apply a method called the method of "semantic differential", to obtain a "measurement" of meaning: see C.E. Osgood, George J. Suci and Percy H. Tannenbaum, The measurement of meaning, Urbana, University of Illinois Press, 1957; but Uriel Weinreich has rightly observed in "Travels through semantic space", Word, 14 (2-3), 1958, p.346-66) that the "semantic differential" measures "meaning" only in a psychological sense, from the point of view of the emotional reactions of the subjects studied to such and such a word (cf. especially p. 358-60 of his article).

Needless to say, the Hallig-Wartburg system is only one of various possible ways in which concepts could be classified; the aim was not so much to devise an ideal scheme as to have a unique basis for specific investigations. If this idea were to be widely adopted, a series of coordinated research projects could be planned with sufficient flexibility to adapt the scheme to the material examined, and yet with enough common ground to make the results comparable.

The recent work of the International Center for the Terminology of the Social Sciences is discussed in Appendix D12.

Relationship to Citation Indexing Method.

The arguments and examples in this Appendix are based on minor modifications to extracts from a paper by Eugene Garfield (*). The reservations are due to the focus on documents as opposed to the focus of this project on concepts. The traditional philosophy of classification system design implies that individual entities (usually documents) can be treated as though they were independent of one another. This basic fallacy not only results in the loss of important informational links, but it is basically inefficient. Little or no effort is made to establish a possible relationship between the entity being classified and the entities already classified. There are exceptions to this rule, but generally the building-block development of human knowledge is not perceptibly reflected in traditional classification systems. In conventional word indexing systems, the indexers cannot afford the time to establish linkages between concepts.

Each addition to the body of knowledge is treated as one of a series of independent events, like molecules of a gas. But the literature is not an "ideal gas"--the molecules interact. Similarly, the body of knowledge, partly embodied in the literature, is composed of highly interrelated elements. It is a heavily cross-linked network. The clearly-visible linkages are those ordinarily provided by authors in the form of explicit citations. Less clearly seen are implicit references as in eponyms and neologisms. Almost invisible linkages exist in the natural language expressions which obscure the relationships, expecially to an unskilled observer. Conventional bibliography is essentially a simple listing or inventory of publications which disregards most of the interrelationships between the items in the inventory. In contrast, citation indexing integrates this necessary and useful listing in a huge graph or network. In this graph, each entity (in this case documents) is a node or vertex in a huge multi-dimensional network. By analogy, this model of the literature (which Garfield considers to be equivalent to man's knowledge) is like a large road map in which the cities and towns share varying degrees of connectivity. Even the smallest hamlets are nodes on the citation map of science.

Garfield refers to previous work of his on this type of historical map.(**) Since each document is an "event" and bears a date, a graphical history may be displayed, but with the important advantage of being able to show the interrelationships among events (see Figure D1). This is a legitimate starting point for the

(*) Eugene Garfield. "Primordial concepts, citation indexing and historio-bibliography." Journal of Library History, nº 2(3), 235-249 (1967).

see also: Eugene Garfield. "Science Citation Index; a new dimension in indexing." Science, 144, 649-654, (1964)

(**) E. Garfield. Citation indexing: a natural science literature retrieval system for the social sciences. American Behavioral Scientist, 7 (10) 58-61 (1&64).

historian. It should be noted that this technique has been in full operation for a number of years, in the form of the Science Citation Index produced commercially by the Institute for Science Information (U.S.A.).

Comment.

There is clearly no technical obstacle to handling conceptual entities in the same manner as documents. This would clearly be of value to both the historical and educational model types.

Garfield himself refers to the possibility of having such graphs displayed directly onto a computer-controlled TV screen (see Appendix B3), or plotted onto graph paper by a plotting device. Computers currently plot such graphs on standard line printers as output from the commonly-used PERT programs.

Garfield is only concerned with the time or historical dimension as a means of sequencing entities, and only with the citation relationship between such entities. There is no reason, however, why other dimensions and relationships should not be used: geographical, educational, logical, etc., corresponding in fact to more of the model-types listed in Appendix A3.

Relationship to the U.D.C./Dewey Classification Schemes (*)

Introduction

The Universal Decimal Classification (UDC), originally derived from the Dewey Decimal Classification (DC), has been, and is still being, extended for use in "classifying articles in periodicals monographs and documents of all kinds,"under the auspices of the Féderation Internationale de Documentation, which authorizes publication of various international editions of the UDC in different languages.

The general pattern of the D.C. is as follows: The whole field of knowledge is divided into 9 main classes, numbered 100/900 (U.D.C. 1/9), general works constituting a tenth class, 000. Each main class is divided into 9 subclasses, e.g., class 300 (Social Sciences) into 310/390 (U.D.C. 31/39). General works on the Social Sciences constitute a tenth subclass, divided into 301/309 according to the form in which the general subject is presented, e.g. 304 Essays. Each subclass is divided into 9 further subclasses, e.g. 320 (Political Science) into 321/329, general works constituting a tenth subclass, divided into 320.1/320.9. This division into ten may be continued indefinitely -- hence "decimal classification."

The use of the decimal notation has been one of the most successful features of the D.C., offering infinite hospitality to new subjects whilst using the best known and simplest of symbols -- Arabic numerals. It was these two factors, and the fact that the D.C. was a classification of subjects, independent of language or race, that chiefly determined the adoption of the D.C. by the Institut International de Bibliographie (I.I.B.) in 1895. Hence, the order of subjects in the D.C. and U.D.C. schedules is substantially the same.

But the D.C. introduced, in conjunction with its decimal notation, two other structural features which have been immensely significant in the development of the U.D.C.:

1. The synthetic principle, whereby recurrent series of concepts are arranged in consistent orders and allocated a consistent notation. For example, in class 400 (Philology) the order of subclasses within each language is the same -- Orthography, Etymology, etc., and the notation representing them is also the same, e.g. 425 English grammer, 435 German grammar, etc.

(*) This Appendix takes the form of extracts from the introductory pages to the Universal Decimal Classification; abridged English edition. London, British Standards Institution, 1961, (3rd edition).

2. <u>The use of a distinctive symbol to introduce a particular principle of division.</u> For example, the appearance of a zero in a D.C. number usually means that the subject represented by the preceeding digits is now divided by the principle "Form of presentation"; so 677.05 means the subject Textile Manufacture 677 presented in the form of a periodical.

It should be clearly understood that "decimal classification" refers essentially to the structure of the notation, since it is theoretically absurd, as Dewey stated, to divide each class into just nine subclasses. This means that the "expressiveness" of the notation in reflecting co-ordination and subordination is often sacrificed. If more than 9 subclasses have to be accommodated, some must "share" a number, e.g. 941 (Scotland) and 941.5 (Ireland); if less than 9 subclasses are required, they are spread over the 9 numbers, thereby shortening the length of the class number - e.g. 592/595 Invertebrates.

Two distinctive modifications of the D.C. pattern by U.D.C. may be noted. Firstly, as an international scheme, U.D.C. removes the occasional American emphasis found in the D.C. nitation -e.g. at 329 (Political Parties). Division of any subject by place finds all countries equally provided for. Secondly, whilst U.D.C., like D.C., is a general classification, not a composite of special classifications, the U.D.C. does provide, via its Points-of-view numbers, a mechanism whereby the structure of the general scheme may be adjusted to the needs of a special classification.

The Universal Decimal Classification (U.D.C.) is a scheme for classifying the whole field of knowledge. It can be applied both to the literature which records knowledge, and to the catalogues, indexes, etc., which refer to the literature. It enables these to be arranged in such a way that all references to information on a particular subject can be brought together and the information located with the minimum of searching.

Availability.

The preparation of the 2nd international edition of the U.D.C. involved some forty specialists, under the general editorship of Otlet and La Fontaine (co-founders of the Union of International Associations), who were concerned chiefly with the Humanities, and Mr. F. Donker Duyvis (later General Secretary of the F.I.D.) who was responsible for most of the sections on Science and Technology. Containing some 70,000 subdivisions, this 2nd edition in French was published during the years 1927-33 under the title of"Classification Décimale Universelle" and has since served as the authoritative basis for all subsequent schedules, full and abridged.

Of the three full international editions based on the 2nd edition, the only one completed since the end of the Second World War is the 3rd (German) edition, comprising 7 volumes of tables and a 3-volume alphabetical index. The 4th (English) edition is still in preparation, and the only sections so far published are 0, 5 and parts of 62 and 67, and 669. Rather more has appeared of the 5th edition (again in French): namely 0,2,3, 61 and 62 and 65, but difficulties are being encountered and progress is slow. Other editions in Japanese and in Spanish have been begun, and one in Portuguese is contemplated.

Abridged editions have been been published in Czech, Dutch, English, Finnish, German, Japanese, Polish, Spanish, Swedish, and other languages, while Arabic and Portuguese editions are in preparation. Only a very brief outline schedule exists in French, but a standard abridged text is included in the important new 3-language edition issued jointly by the Deutscher Normenausschuss (D.N.A.), and the British Standards Institution (B.S.I.), with the Association Belge de Documentation (A.B.D.) and the Union Française des Organismes de Documentation (U.F.O.D.).

Development

The development and maintenance of a system as comprehensive and widely-used as the U.D.C. could hardly have been achieved without some form of effective supervision and control. The ultimate authority on general U.D.C. policy and development is the International Committee on Universal Classification, on which all national member committees of the F.I.D. are entitled to be represented, but its membership is too diffuse to maintain continuity between F.I.D. Conferences. Day-to-day control and supervision are vested in the more compact Central Classification Committee (C.C.C.), which consists of the F.I.D. General Secretariat at The Hague, where an up-to-date master copy of the complete U.D.C. is maintained, based on proposals submitted by U.D.C. revision committees or individuals in many parts of the world.

Principles

Three basic principles are evident in the Universal Decimal Classification:-

1. It is a <u>classification</u> in the strictest sense, depending on the analysis of idea content, so that related concepts and groups of concepts are brought together, and the arbitrary and often haphazard systematization of alphabetical and other arrangements is avoided.

2. It is a <u>universal classification</u> in that an attempt is made to include in it every field of knowledge, not as a patchwork of isolated, self-sufficient specialist groupings, but as an

integrated pattern of correlated subjects. This universality at the conceptual level is supported by notational devices, which permit the linking of simple main numbers (for simple ideas) either with other main numbers or with auxiliaries denoting Place, Time and similar commonly recurring categories - in each case forming combined or compound numbers. If separate index entries are made under each single main number forming part of a combined number, the complex subject can be located from each point of view, regardless of any subjective approach on the part of the classifier.

3. It is a universal decimal classification, constructed on the principle of proceeding from the general to the more particular by the (arbitrary) division of the whole of human knowledge into ten main branches, each further subdivided decimally to the required degree. This principle can be applied to any notation, but Arabic numerals are internationally familiar in a recognized order, whereas letters and other symbols are not.

Structure

A. Main structure

The whole field of human knowledge, regarded as unity, is divided into ten main branches denoted by decimal fractions as follows:-

- .0 Generalities:methodology, documentation,scripts; recording, collection and dissemination of information
- .1 Philosophy, metaphysics, logic, ethics. Psychology
- .2 Religion. Theology
- .3 Social sciences: including statistics, law, education
- .4 Philology. Languages
- .5 Pure science, mathematical and natural
- .6 Applied science: medicine and technology
- .7 The Arts, including architecture, photography, entertainment and sport
- .8 Literature
- .9 Geography. Biography. History.

On this foundation, the notation is built up by continuous extension of the decimal fractions, on the principle of proceeding from the general to the particular. Thus, every concept within the domain of pure science is represented by a decimal fraction greater than .5 and less than .6, the subdivision being carried to any required degree, as shown by the following example:

- .5 Pure science, mathematical and natural
- .51 Mathematics
- .52 Astronomy, geodesy
- .53 Physics, with mechanics
- .531 Mechanics (of solid bodies)
- .531 7 Measurement of geometrical and mechanical magnitudes
- .531 71 Measurement of length, linear dimensions

- .532 Fluid mechanics
  etc.,etc.

B. Auxiliaries

Auxiliary numbers and signs are a means of eliminating repetition by grouping recurring subordinate concepts such as language, form, place, time and point of view

- (i) addition sign +, used to link commonly associated concepts, e.g. 622 + 669 for mining and metallurgy.
- (ii) extension sign,/, used to denote a range of concepts which collectively form a branch of knowledge, e.g. 624/628 for civil engineering.
- (iii) relation sign, : , used to link related concepts of equal value, e.g. 31:63 for statistics as applied to agriculture.
- (iv) language sign, = , used to give the language of the document, e.g. 22.05=30 for the Bible in German
- (v) form sign, (0...), used to give the nature of the document (periodical, book, etc.), e.g. 58(021) for a comprehensive botany handbook or manual
- (vi) place sign, (1) to (9), used to give the geographical range of the subject denoted, e.g. 385(43) for the German railway system.
- (vii) nationality or race sign(=...), used to indicate the racial aspects of the subject denoted, e.g. 291.33 (=947.5) for witchcraft among the Eskimos.
- (viii) time sign,"...", used to allocate dates or other time aspects to the subject denoted, e.g. 341 "1898.12.11" for international law as at 11th December 1898.
- (ix) alphabetical subdivision used to cover particular features, e.g. 820 (Shakespeare) for the works of Shakespeare
- (x) point of view sign, 00...., used to indicate the broader aspects of a subject from a particular viewpoint, e.g. 622.009 for the social and ethical aspects of mining
- (xi) note on application sign, -..., used to indicate the manner in which the subject denoted is used, e.g. 331.64-055.2 for labour services for women
- (xii) synthetic sign, '..., used to build up compound numbers e.g. 547.29'26 for carboxy-acid esters

Comment.

1. Some examples of typical numbers generated by the UDC are
   --159.9+331.826+37.048.2

   --338.984.4: 622.33+669.1(4)

   --621.039.004.14:327.3

   --572+930.6+41(98/99)

This sort of variable length/multiple sign number is definitely not oriented toward standard data processing methods. (One computer system is however in operation in the USSR which handles these codes).

2. There is no attempt to take care of semantic differences associated with multiple use of the same term, each to denote a subtle difference of meaning which can only be classified by a definition in the form of a series of phrases. (see Appendix A2) This can be handled, but only by increasing the depth of the indexing to create rather unwieldy code numbers.

3. Despite references to its applicability to the "whole field of knowledge" the scheme is primarily intended and used for the classification of documents. Most of the people and organizations involved in its use are documentalists with a commitment to document handling. For example, the UDC coordinating body is the International Federation for Documentation. Little value is seen in divorcing knowledge from the documents in which an attempt is made to record it. This divorce is however now possible with new techniques of handling information.

4. The development of the system leads to unbalanced depth of coding. The best example is the field of pure science which is confined to 5. The tremendous development in this field since the scheme was conceived means that a code must be made quite lengthy before it is significant, as opposed to the case in a less developed field such as philology, 4.

5. Revision and reconceptualization is very difficult and slow as the following extracts indicate. The process of reconceptualization is not seen to be an essential feature of the advance of knowledge, but rather as annoyingly inconvenient to continued development within traditional knowledge hierarchies.

In revision, conflicting tendencies are inevitable: the older established users, having built up extensive U.D.C. catalogues over the years, tend to resist changes, whilst the newer and prospective ones, with little or no existing U.D.C.-classified material, are often anxious for drastic rearrangement, sometimes in the interests of logic alone. However, the majority of those active in U.D.C. revision have accepted a policy of compromise: they endeavour to make the classification better for established users, and more attractive to non-users; to reduce confusions and shortcomings, and to introduce new concepts without disturbing the existing tables too much.

The governing rule in all U.D.C. revision work is that the significance of a particular number may be extended or restricted, but may not be completely altered. If a U.D.C. number (with any subdivisions) is obsolete, it may be "cancelled", which means that its use is no longer authorized because a better or more up-to-date arrangement has been developed under another number, often a "free" (unused) number. Eventually, the cancelled number becomes "free" (by disuse) and may then be authorized with a completely different significance, but only after a period of 10 years, which is considered the minimum necessary to enable

current users to readjust their files and to avoid widespread confusion of the older and newer meanings. This policy undoubtedly slows down revision, but probably creates least dissatisfaction, especially when applied with discretion and some degree of flexibility.

6. The UNISIST Study Report on the feasibility of a World Science Information System has the following to say about the UDC and other systems:

> "Librarians and information specialists would generally agree that a world-wide scheme of subject categorization is needed to facilitate document and information exchanges...Opinions differ, however, when it comes to deciding which scheme best suits the purpose. Several encyclopedic classifications are in competition -- the Dewey Decimal Classification, the List of Subject Headings used in the Library of Congress, the Colon Classification, the Universal Decimal Classification, etc. -- and although the last named has benefitted from extensive international support through FID, it is by no means the unique candidate for world-wide recognition as the standard subject category list. Its advantages and shortcomings were examined by the UNISIST Working Group on Research Needs in Documentation, who came to a twofold conclusion: (a) organizational and technical measures could be taken to obviate the managerial drawbacks of UDC, e.g. slow revision procedure, infrequent re-editions, etc.; (b) on the other hand, no clear answer could be given to the more controversial question of overall or local inadequacy, as regards the content and structures of UDC divisions...further studies and experiments are required to assess the potential value of UDC in its present state, as the unique world list of subject headings for broad categorization, or "shallow" indexing of documents." (p.95)

7. As the UNISIST extract above acknowledges, UDC is one amongst many classification schemes which are in competition. The tendency for different classifying groups to favour different category breakdowns should be contained and facilitated within an information system and not left to deteriorate into sordid squabbles which do not recognize the value to knowledge advance of alternative views, and a continuing effort at reconceptualization, restructuring and redefinition of knowledge.

Relationship to the UN/OECD Aligned List of Description (*)

Introduction.

An important international effort to coordinate subject indexing approaches has been made by libraries of international agencies concerned with "economic and social development"in its governmental sense. Those initially involved were:

--International Labour Office (ILO)
--International Committee for Social Sciences Documentation
--Deutsche Stiftung für Entwicklungsländer
--Food and Agriculture Organization of the UN
--Organization for Economic Cooperation and Development (OECD)

The following have already (in 1969) expressed their intention to cooperate in the undertaking:

--GATT for international economic relations
--ECMT for transport economics
--UNIDO for industrial development
--UNESCO (Social Sciences Department) for sociology
--ICSSD for economics

Other organizations such as

--World Health Organization (WHO)
--International Atomic Energy Agency (IAEA)

have declared their readiness to determine, in the light of their particular activities, the most suitable description for inclusion in a lexicon of economic and social development. In addition this work is followed with close interest by the United Nations Inter-Agency Working Party on Indexing and Documentation in which many of the above bodies participate. (The UN agencies are under pressure from ECOSOC to establish "a central index of the major documents of the United Nations' system as a measure that should enhance the usefulness of existing documentation", although some agency representatives are quite sceptical about the feasibility and even the desirability of this effort.)

The project for an international lexicon of economic and social development originated in a suggestion made by the Secretariat of the United Nations to the OECD Development Centre in 1964. Word lists were drawn up which resulted in 1966 in the publication of a draft Aligned Description List which was a sort of compromise between a number of agency descriptor lists and indexes prepared by reducing their essential contents to simple

(*) The material in this Appendix is based primarily on extracts from the introduction to the Aligned List of Descriptions, Paris, OECD, 1969.

elements, or "uniterms", which lent themselves to multiple combinations. This was tested by 14 national and international organizations.

Work on a new Aligned List continued to completion in 1968. The aim was not to establish a list of keywords taking account of other descriptor lists at the level of an arbitrary chosen common denominator, but rather to combine all the vocabularies in use into a single whole, leaving each responsible organization to make the adjustments deemed necessary by common agreement. In 1968 it was agreed that all UN agencies should be urged to cooperate in this effort in Engligh and French and German, and that extension to Russian and Spanish should be envisaged.

Descriptors.

The descriptors used are single words or compound expressions. The single words are meaningful terms with a retrieval value, the expressions are formed, in order to achieve a certain specific character from the outset (e.g., community development, rural development, etc.) or to avoid the ambiguities inherent in the exclusive use of uniterms. As a general rule the descriptors are substantives and given in the singular.

Certain descriptors having more than one meaning are accompanied by an appropriate "scope note" delimiting their use.

Structure.

Since one of the underlying reasons for the Aligned List was to ensure consistency of the differend descriptor lists, it was necessary to group all the descriptors into semantic fields. It would obviously have been impossible to bring any inconsistencies to light from a purely alphabetical list of several thousand terms.

There is nothing systematic about the proposed structure, which is not based upon any preconceived ideas or theories and differs from all the models offered by hierarchical classifications; it simply emerged gradually, once it was started to arrange the descriptors by affinities. It seemed convenient for the aim in view and nothing more. It is fully recognized that there may be a score of other possible arrangements, all equally practical; the essential thing was to decide on one.

It is at present arranged in ten major blocks numbered from I to X. Within each of these blocks a certain number of relatively extensive semantic fields have become evident. Each of these fields can be identified by one or more descriptors. These are simple "recognition signals"; they do not really cover the field, since they are members of it; they are not imposed on it, as a generic term is imposed on a more specific terms, they merely serve to call it to the attention of the analysts and ensure its inclusion in the block in strictly alphabetical order. In most cases, these extended semantic fields are broken down into more restricted units numbered 1, 2,3,4, etc. The series is generally headed by a unit 0 relating more directly to the descriptor (or group of descriptors) which best identifies the semantic field. For greater convenience, and without necessarily giving them a particular label, a certain number of descriptors which belong to these fields and seem significant to their content have been grouped between brackets at the head of the different restricted units.

This way of structuring the vocabulary may seem complicated, but is really simplicity itself, and has the advantage of always being easy to revise, since the restricted units can be shifted from one field to another, or rearranged in new fields according to the requirements or logic of more specialized documentary undertakings. Above all, we have sought to avoid the inflexibility of hierarchical classifications which freeze knowledge at a given moment of time and give their terms no meaning except in the light of the available whole.

In the process of preparing these semantic groupings, it has proved possible to pinpoint and eliminate a certain number of inconsistencies between the descriptor lists. In addition, this grouping has two other advantages: it should guide analysts in the search for pertinent descriptors; in the absence of scope notes for all terms used, it also makes it possible to avoid troublesome ambiguities, by prescribing the correct usage of certain polysemantic descriptiors by the mere fact of situating them in proximity to other non-ambiguous descriptors.

Unquestionably, this grouping is imperfect; certain of its sections could be further developed. In any event, it should be regarded as a simple working instrument, and not as a mandatory structure."

Use and Development.

"In its present form, the Aligned List does not constitute a genuine thesaurus, but it is substantially more than a simple list of keywords for documentary use. We have refrained from carrying it any further for two main reasons; it was important, in the context of effective international cooperation, to show the exact state of the work undertaken by each of the participating institutions, and it was just as important to think of the future, by leaving access to the List open to other organizations who might wish to develop those sections appropriate to their special interests."

Comment.

1. The Aligned List is a success for those bodies who most closely collaborated in its elaboration. The organizations have a similar perspective and a common frame of reference, namely "development".

   It is claimed that the List is open to other organizations, but this is clearly only to the extent that such bodies subscribe to the structuring or the semantic fields already established -- or else are so specialized that their descriptions constitute an isolated sub-set which interacts with none of the existing fields.

   The degree of agreement already achieved is difficult to establish. There are rumours of "problems of revision" and that the system is now "coming apart at the seams" and these are difficult to dispel because the official reports all seem to have a public relations component.

   One major difficulty is that to the extent that a participating organization is less committed to the doctrine of "development" as a mission-orientation, the more questionable the "agreed" semantic fields become. The scheme is the result of governmental decisions defining the semantics of economic and social development. "Development" has been a significant term since the late 1950's. There is no guarantee that it will not be replaced by intergovernmental decisions on "environment" or some other mission-oriented term which will shift all the semantic fields. Furthermore, if at any time several such programmes have equal status, multiple over-lapping fields would be required and cannot be supplied.

   The list of descriptions is therefore an excellent model for the purpose for which it was conceived, but cannot be considered of permanent value as a means of handling evolving, alternative patterns of knowledge.

2. The List is term-oriented, by definition. The semantic field is of secondary importance. Computer programs have been developed to handle the terms in different languages. These programs are organized to handle the alphabetic terms and obtain other language equivalents. Because it is language oriented, rather than code number-oriented, interlanguage problems arise before a term has been agreed in the other languages. There is a "favoured language" problem for different parts of the list at different times, as the introduction to it explains.

   > "In the first place English dominates the major part of the List which is arranged in semantic fields; secondly, French, and, to a lesser extent, German, are generally used as simple languages of transla-

tion. Even where the descriptors proper to the different organizations have been translated (from English, or from German to French or English) by these organizations themselves, or under their control, the list which they constitute differs quite appreciably, by the mere fact of being a translated list, from a list which might have been prepared in the same language. There is therefore nothing surprising about the somewhat subordinate character of the list in French, or the inadequacies of the English list in the parts translated from German, or conversely, a certain lack of substance in the German list translating the English. This is a purely transitional situation. The prolonged use of the Aligned List along its three parallel lines will gradually allow these deficiencies to be made up; either by the replacement of ambiguous terms by more apt descriptors, or above all, by the introduction in French and Encligh of synonyms which form a more reliable guide to lead the analysts towards the most useful word or expression."

From the perspective of this proposal, the following operations have been blurred together to contribute to this "transitional" situation:

-- entities are labelled by terms
-- terms have to be classified into semantic fields to be incorporated
-- terms have to be translated and agreed as terms to avoid language dependence

When terms are dropped, the reverse procedure affecting the structure of the list must be followed.

Each of these steps involves operational and intellectual difficulties which tend to slow down and resist modification. Much more flexibility would have been obtained by a number orientation.

3. The List makes great efforts to be flexible by being term-oriented. To do this it has had to avoid hierarchical classification of any depth. This choice is not in the interests of those users who need a "deep" classification structure.

Relationship to the UNISIST Proposals for a World Science Information System.

The United Nations Educational, Scientific, and Cultural Organization (UNESCO) and the International Council of Scientific Unions (ICSU) created a joint committee in 1967 to study and report on the feasibility of a world science information system. The report was published in 1971 (*) and its recommendations were the subject of a major intergovernmental conference at Unesco in October 1971.

It is useful to look at this report, because there are suggestions in it that its scope should be extended to cover the "behavioural disciplines" despite a primary emphasis on the "basic sciences " (p.136).

"The problem the Committee was called on to review is a complex one. In its popularly recognized form, it has been unfortunately termed the "information explosion". It is frequently alleged that scientific and technical articles and reports are increasing at a rate which makes it extremely difficult for scientists to keep up with the work of their colleagues. Faulty distribution practices and understocked and understaffed libraries make access to these reports difficult; once access is obtained linguistic barriers interpose comprehension difficulties.

There are the more familiar characteristics of the problem. Less obvious, but more radical, are the changing needs of the world scientific community for information. The interdisciplinary approach to problems of the environment, for example, requires information drawn from a variety of sciences: chemistry, biology, sociology, to name only a few. The emerging needs of applied science, technology, and engineering add further complexities. The classic information services, the scientific journals, abstracting and indexing services, libraries, have all demonstrated a cultural lag in accommodating to these new requirements. The achievement of new and flexible forms of information services to meet these new needs is the fundamental problem..."

The report notes that:

"...so many institutions are already engaged in sponsoring or conducting projects subservient to UNISIST's objectives... that it would be both unreasonable and impracticable to conceptualize at this stage a transfer or subordination of their duties to a new international organization, henceforward acknowledged as the unique authority in this area. For one thing, few if any of the existing agencies or services would be ready to accept the authority..."(p.127)

(*) UNESCO and ICSU. UNISIST; study report on the feasibility of a world science information system. Paris,UNESCO,1971.

It therefore recommends the establishment of an ongoing project to move towards a world science information system conceived as follows:

"A 'World Science Information System' in this case, is any complex set of rules and media that may be devised with the purpose of actualizing this concept of world-wide information sharing in the transfer of scientific and technical information from scattered producers to dispersed users in all regions of the earth. This preliminary definition has several implications. (a) A major one is that there is no room in this "system" for a centralization of document processing in a single world institute, as proposed by some...the world system is to be thought of as a "system of systems", or better a network of systems, whose components are the operating information systems of the world whatever their scope...(b) However, there is a sense in which the proposed scheme may qualify as a system in its own right. The common "rules and media" that provide the basis for the interconnection of the components are often designated as systems, or even networks...What is meant then is that some systematization of current practices is needed..."(p.80-1)

Goal of UNISIST

The report is deceptive on a rather fundamental point. The UNISIST project, it is stated, stands for:

> "The unimpeded exchange of published or publishable scientific information and data among scientists in all countries" (p1)
>
> "The world-wide availability of scientific documents or data should be acknowledged as one of the ultimate goals of UNISIST..." (p.115)

It would appear from this that the UNISIST criteria is the maximization of diffusion of documents and information. The relationship to the diffusion and availability of knowledge is not explored. In fact, the transfer and availability of knowledge is identified with the transfer and availability of information.

> "Knowledge (scientific and technical): the subject and findings of research (facts, theories, hypotheses, etc.) as embodied in scientific "information" and "data" (Glossary, p.148)
>
> "Information (scientific and technical): the symbolic elements used for communicating scientific and technical "knowledge", irrespective of their nature (numerical, textual, iconic, etc.), material carriers, form of presentation, etc.
>
> The word "information" in this report, is not differentiated from "documentation"; it refers both to the substance, or content of scientific documents, and to their physical existence." (Glossary, p.148)

The blurring goes so far that one suspects distinctions were not clearly established in the minds of the Committee or significant to it. As an illustration, the Committee had a "Working Group on the Evaluation, Compression, and Organization of Scientific Information" (list of groups, p. 152), which is referred to as the Working Group on "Evaluation, Compression and Organization of Scientific Knowledge" (p.103) and whose work is referred to as the "evaluation and compression of scientific documents or data" (p.138), and as "improving the quality of scientific documents through evaluation, compression, etc." (p.20). These remarks would be trivial were it not that knowledge, information, and documents each demand a different approach to maximize transfer and availability.

Briefly, documents pose a physical handling, transfer and filing problem (which may be eased by reproduction at a distance). Information, consists of signs which can be read, transferred, manipulated and filed electronically. They function as symbols of units of human knowledge, but only during the short-duration process of being read for meaning. Knowledge transfer depends on the ability of the momentary psychological system "sign and reader" to generate an unambiguous, coherent and consistent meaning in the mind of the reader, and conversely to convert a distinct meaning or concept into a suitable sign which can be interpreted with equal ease by another reader. Information, in the form of signs, can be read without resulting in the transfer of knowledge and particularly of the knowledge intended (e.g. undecipherable hieroglyphic writing can be "read" without knowledge transfer).

Availability of knowledge

Information can be made available on the location and content of documents containing information (e.g.as abstracts). The documents may even be made physically accessible (e.g. to the point of being in a pile on one's desk). These accomplishments do not constitute "availability of knowledge."

The Study does not recognize that the period covered by the proposed system is one in which increasingly, it is almost impossible for the decision-maker or researcher to determine what information from which discipline is "relevant" (*). If he attempts to order all the relevant documents (or even subscribes to the appropriate abstracting service), the purchase or transport costs will be prohibitive (except to a small elite); if he waits for all the relevant information, it will be too late for him to make a useful decision; if he gets all the relevant information in the form it currently takes, he will have neither

(*) "...how is a practitioner of any one discipline to know in a particular case if another discipline is better equipped to handle the problem than is he? Is would be rare indeed if a representative of any one of these disciplines did not feel that his approach to a particular organizational problem would be very fruitful, if not the most fruitful..." (R.L. Ackoff, Systems, organizations, and interdisciplinary research.)

the time, the training, nor the inclination to read it all; and if he reads and comprehends it all, he will not have the time or the ability to convey his understanding to those whose support he must obtain to carry a vote on the matter or, ultimately, to the man in the street.

> "Among the responses to such pressure is greater specialization. Yet...this expedient is not always satisfactory, for the degree to which one specialty impinges on another also is increasing, and with it the amount of information with relevance to any one field of endeavour." (SATCOM, .178-9)

Any attempt to divide up the task merely poses once more all the problems of adequate coordination and integration of programmes and the need for a clear overall perspective. A multidisciplinary synthesis cannot be effectively conveyed in a report. The shorter the report, the less depth and detail it can contain, and the less credible it becomes, particularly if the validity of the argument depends on many successive steps. The longer the report, the less likely it is that it will be read and understood.

> "Consider this dilemma: while our technological abilities to generate and disseminate potentially useful data have increased manyfold in the past few years, man's physical capacity to register and to process potentially informative data has probably increased very little, if indeed at all. The sheer volume of data that crosses the typical executive's desk today should serve to spotlight the inadequacies of the education and development of our acquisition strategies and practices. But no gain in ability could offset the widening gap between the exponentially-increasing quantity of data available for consumption and man's very limited capacity for acquiring and processing useful information." (*)

Even if the signs are in the reader's native language and in the jargon of his discipline, knowledge transfer has not necessarily been significantly facilitated. For when and if he has the time to digest the symbolic value of the signs contained in the document, they may not constitute an unambiguous, coherent communication. The signs may have a symbolic or conceptual value for the reader which differs from that of the author (*).

(*) Lee Thayer. Communication and communication systems; in organization, management, and interpersonal relations. Homewood, Irvin, 1968,P.202.

(**) Colin Cherry.(World Communication:Threat or Promise? N.Y., Wiley, 1971) notes with respect to a BBC study that: "within a country's own borders it is only recently, after 45 years of broadcasting, that the importance of word whoice is really becoming appreciated....It is now being realized that many words commonly used by broadcasters to people of their own countries are unknown in meaning to many listeners....If this is so with internal, national broadcasting, what about overseas broadcasting in foreign languages? What misunderstandings have been innocently created? (p.16)

Knowledge via documentation

It is questionable, in view of present trends, whether knowledge transfer can continue to be effectively accomplished primarily via document transfer. The United Nations is potentially the most significant institution in existence and is at a vital nexus of multidisciplinary, international knowledge transfer -- which it currently accomplishes via documents (*). And yet it has a documentation problem (which in a sense is equivalent to that of many, if not most, other large organizations).

> "This issue has been repeatedly recognized by the General Assembly, the Economic and Social Council, the Joint Inspection Unit and nearly a dozen of other UN bodies as one which directly affects the functioning of the UN. Suffice it here to note that in 1970, the UN, both in New York and Geneva, produced nearly a million page documentation in all languages. The massive volume of documentation produced by the UN prompted a former President of the General Assembly, Mr. Lester B. Pearson of Canada, to remark that "the United Nations is drowning in its own words and suffocating in its own documentation." The Joint Inspection Unit stated recently in its report submitted to the present General Assembly session that "the inspectors do not hesitate to say that the point of saturation has now been reached and indeed overstepped." (**)

The last quote in fact continues with the significant phrase "and that the law of diminishing returns is taking over... Beyond strictly financial considerations, therefore....the future usefulness of the Organization may well hinge on its ability and determination to set once and for all, and strictly enforce a reasonable but drastically reduced ceiling to the volume of documentation its various bodies call for and its services produce." (***)

Stemming the generation of new knowledge in developed countries, is however, not as feasible as lowering the birth rate in developing countries (****). To severely reduce one means of storing and disseminating such knowledge, without seeking a more appropriate complementary medium, could only be counter-productive and unsatisfactory.

The limitation of the documentation system approach can also be usefully studied if the problems of physical accessibility and

(*) UNITAR/EUR 3/2, 1971, p.2. "Only recently the Secretary-General of the United Nations affirmed that the Organization's most important working tools were documents. Thus the main medium for conveying information consists of documents."

(**) UNITAR. The Interest of the United Nations Institute for Training and Research in the question of United Nations documentation. Geneva, UNITAR/Eur 3/1, 1971,p.1.

(***) UN Document A/8319, 2 June, 1971 (or JIU/REP/71/4)

(****) UN Document A/7576, 25 July 1969, para.2, shows that document production by New York HQ increased by 50%, from 1964 to 1967, to 600 million page-units. This does not include production of the regional or Geneva offices or specialized agencies.

indexing are considered eliminated. What problems would persist if one had a computer terminal beside one's desk giving zero-cost access to all the world's literature as published (*), with an immediate view and copy of any page desired, and translated from any language? This is an ideal documentation system -- but no synthesis or overview of knowledge in a field emerges. "Reviews of the literature" by persons with the special slant of particular schools of thought are the only substitute. It is questionable whether one's mobility in "semantic space", outside one's own narrow territory (in which one knows what one wants to know and who knows it) would be significantly improved.

One is not exposed to alternative hierarchies of conceptual nexuses linked directly or indirectly to more distant nexuses from which relevant knowledge may be obtained. (There are no "heights" in documentation systems -- the general is filed with the particular cf. the treatment of documents with an interdisciplinary emphasis.) The potential value of a knowledge-oriented information system as an active stimulus for creative social change and problem-solving may even be directly proportional to its ability to draw attention to the existence of established relationships of low probability (i.e. low entropy) between concept nexuses. This is not a criteria of document information systems where the emphasis is -- for cost reasons -- on facilitating access to those documents which are most probably relevant in terms of demand frequency.

Natural science vs. social science.

Treatment of documentation as synonymous with information and both as symbols of knowledge (or the presence of knowledge) creates the illusion that the world science information system will in some way make knowledge more accessible.

This may in fact be true in the case of the scientific and technical information covered by the UNISIST project. For there, invariants in the objective world are represented by signs which can in most cases be directly and unambiguously attached to the object in question, to the satisfaction of the natural science community. The sign for the object and the conceptualization of it are intimately and unambiguously related. Another sign in another language may be used but the rules of transformation are clear (the natural language verbiage is another matter, but is less significant). It is a case of "one sign, one concept, one object". It is therefore possible to infer that knowledge transfer tends to accompany information transfer.

(*) One could think in terms of a personal library of 30 million books. A recent UNITAR document (UNITAR/Eur/ 3/2). notes that there will probably be one million journals in 30 years time. Currently it is estimated that about 2000 books (i.e. 1 million pages) appear every minute throughout each day.

(This inference may however be very dangerous in the case of non-Indo-European language users, for whom the "objective" nature of the world may appear less significant. (See Appendix E2) But any extension of the world science information system, as it is conceived, to the social sciences would only be of superficial significance if the above distinctions were not made clear. This is because in the social sciences, most of the debate concerns the relation between perceptual invariants detected (by the consensus of a group), signs (selected by the group) and the associated conceptual meaning -- as has been recently pointed out by Jean Piaget (*):

> "All the social and human sciences are more or less closely concerned, in their diachronic aspects, with the development of knowledge (as a subject)... The foregoing considerations show that the human sciences, in so far as they necessarily include in their field of study the subject of knowledge -- the source of the logical and mathematical structures on which they depend-- do not merely maintain a set of interdisciplinary relations between one another ... but are part of an extensive circuit or network that really covers all the sciences... It was essential to recall this so as to be able to shape our conclusions in such a way that they might succeed in revealing the true significance of interdisciplinary relations.
>
> "For their significance far exceeds that of a mere tool for facilitating work, which is all they would amount to if used solely in a common exploration of the boundaries of knowledge. This way of viewing collaboration between specialists in different branches of knowledge would be the only possible one if we admitted a thesis to which far too many research workers still unwittingly cling --that the frontiers of each branch of knowledge are fixed once and for all, and that they will inevitably remain so in the future. But the main object of a work such as this ... is to push back the frontiers horizontally and to challenge them transversally. The true object of interdisciplinary research, therefore, is to reshape or reorganize the fields of knowledge, by means of exchanges which are in fact constructive recombinations." (p.521-524, emphasis added)

The natural sciences are therefore primarily interested in the debate on the, usually tangible, content of categories (which are considered to be relatively permanent), and the dynamic lies in subdividing the categories and discovering relationships between their content. Whereas the social

(*) Jean Piaget. General problems of interdisciplinary research and common mechanismes. In: Unesco. Main trends of research in the social and human sciences. Paris, Unesco, vol.1, 1970, pp. 467-528.

sciences, unable to latch onto an unambiguous content, are primarily interested in the categories themselves and their interrelationships, and the dynamic lies in reformulating, reshaping, and regrouping the system of categories in an effort to get closer to the content (*). It is clear that the natural sciences could easily adjust to an arbitrary permanent category hierarchy, whereas the social sciences would be straight-jacketed and ill-served by any such system.

Shuffling documents and signs might facilitate the transfer of meaning and knowledge between those who could identify the representative of the group for whom a particular set of meanings could be consistently and unambiguously attached to the signs. But even within that group, advances in knowledge and reconceptualization have to be carefully related to the original set of meanings. However, making the documents and signs of that group available to other "outside" groups would only introduce "noise" and confusion. A knowledge-oriented information system would be needed to avoid such confusion and facilitate fruitful interaction between different schools of thought within the social sciences.

Perhaps the clearest example of the need for a concept- or knowledge-oriented approach in the case of the social sciences is given by the confusion of meanings associated with the concept "democracy". Few people know that Unesco arranged an expert meeting to clarify its meaning. The meeting concluded that at least thirty distinct meanings were required and in use. (*) The report was withdrawn for circulation for political reasons -- it is political dynamite. It means that in most international debates (in which the word is a vital element of the consensus of interest and common goal on which the discussion is founded) participants are simply talking past one another, and resolutions containing the word are of questionable significance. In fact, the multiplicity of interpretations implicit in term-oriented discussions and report production may be considered a direct stimulus to the production of further reports giving clarifying or alternative interpretations -- thus further clogging document systems.

(*) Both natural and social science have conceptual parsimony as a criterion, whereas the "sciences humaines" are interested in multiplying the number of possible concepts and increasing their variety.

(**) Mentioned by F.A. Casadio, Director, Societa Italiana per l'Organizzione Internationale.

The Shepard's Citator coding technique.

D.A. Strickland has suggested (*) that the Shepard's Citation coding system for American case and statutory law might provide suggestions as to how to handle the concept coding problems on computer.

The beauty of the Citator system is that it too is designed to convey accurate information about ambiguous concepts and about ideas the meaning and validity of which are themselves in dispute. The basic logic of it is this: The units of analysis (judicial decisions and statutes) are ordered according to real time. Since these elements are already referred to by number (e.g. 351 U.S. 147,1956, meaning a certain volume of the <u>U.S. Reports</u> at a certain page, for the given year), the task of the citator is somehow to characterize the relation of the elements across time. Note that at this level the <u>content</u> of a unit need not be described or characterized one way or the other; what <u>is</u> represented is the relationship between two units.

The simplest way of connecting the units is to designate which of them explicitly cites which others. Hence, a later case would not be cited at all, under the entry for an anterior one, unless the former were somehow dependent on the latter. For example:

<u>63 Mich. 709 (1961)</u>

j 64 Mch 328
d 86 Mch 96
f 94 Mch 18
115 Mch 667
(etc.)

The citation without a lower-case prefix is merely one which relies on the case-in-chief in <u>other than</u> one of the ways designated by the standard prefixes. The prefixes stand for judgements about the nature of the connections,e.g. that the later case dissents from (d) the earlier one, or reverses (r) it. (The need for precision in this area stems of course from the American adherence to <u>stare decisis</u>.)

This approach is possible with the coding system in zone b.1 (Appendix A6).

Strickland then suggests two possible refinements:

(*) Internal note to COCTA members, 9 October 1971. The text of this Appendix is entirely based on the contents of the note.

(i) use of codes to indicate topological relationships

- B is included in A
- B includes A
- B overlaps A
- B is obverse of A
- B contradicts A (i.e. "has nothing to do zith A")
- B is included in A and includes C
- etc.

These features are covered by the coding system in Appendix

(ii) indication of the degree of consensus on the characterization of a concept "inasmuch as reasonable men will differ not only about the exact meaning(s) of a political concept, but also therefore about the simplest relations between such concepts." Such coding is possible in zone b.2 (see Appendix A6).

Use of the International Standard Book Numbering Technique

There is a strong temptation to adopt a technique similar to that of the International Standard Book Numbering (ISBN) system now used (on the reverse of all recent book title pages) to give a unique code to each book. This number consists of 10 digits made up of the following parts:

-- group identifiers (i.e., national, geographical, language or other convenient group). An "agency" coordinates the allocation of numbers within each group e.g., one for Anglo-American publications ("0"), one for UN system publications, etc.

  The group identifier is allocated by an international standard book numbering agency (in formation *).

  (This could be considered as a concept filing centre identifier allocated by some loose coordinating body.)

-- book publisher identifiers. The publisher identifier is allocated internally within the group by the group agency. (This could be considered as an accredited concept filing source identifier allocated with respect to the filing centre for which it locates new conceptual entities)

-- book title identifiers. A block of sequence numbers is reserved for each publisher to permit him to select the next available for the next book. (This could be considered as a block of sequence numbers for concepts, so that each accredited source can select the next number as each new concept is identified.)

-- check digit. This ensures that the code has been correctly transcribed and input to the computer. A computer pre-generated list of "available" sequence numbers incorporates this digit (which is calculated on a modular 11 with weights 10-2, using X in lieu of 10 where 10 would occur as a check digit).

The total length is 10 digits, but the three identifiers only total 9 digits. In order to avoid wastage of numbers or lack of sufficient numbers, publishers with a large book output (of which there are few) have a two or three digit identifier so that the title identifiers can use six or five digits. A

(*) See proceedings of International Organization for Standardization. Technical Committee 46, Working Group 1 on ISBN/ ISSN (ISO/TC 46/ WG1). Lisbon, April-May, 1971. For accessible description, see: Suzanne Honoré. La numerotation normalisée internationale du livre. Bulletin des Bibliothéques de France, 14, 28, August 1968, p. 321-333, bibl.

small publisher (of which there are many) has a five or six digit identifier so that the title identifier can use two or three digits. The publisher identifier is therefore selected on the basis of his output using from two to six digits as required. Hyphen separators are used.

The temptation to use this system should however be resisted. While the significance attached to the digits is only "administrative" and has no "theoretical" implications, problems of overflowing the allocated blocks are bound to occur. The system will "bulge" in unpredictable areas as the U.D.C. has done. It is also questionable whether so much significance should be placed on the source which, once the concept has been incorporated, will quickly become irrelevant within the network of other related concepts from other sources.

Sources for social science concepts

Guidance in limiting scope can be obtained by concentrating, in the light of the priorities in Appendix A7, on concepts mentioned in such publications as:

1. David L. Sills (Ed.). International Encyclopedia of the Social Sciences. Macmillan, 1968.

   Approx. 2500 main entries; 50,000 cross-reference entries, covering the concepts, theories, and methods of the following disciplines:

   Anthropology -- includes cultural, economic, physical, political, social, and applied anthropology, as well as archeology, ethnography, ethnology, and linguistics.

   Economics -- includes econometrics, economic history, the history of economic thought, economic development, agricultural economics, industrial organization, international economics, labor economics, money and banking, public finance, and certain aspects of business management.

   Geography -- includes cultural, economic, political, and social geography, but not physical geography.

   History -- includes the traditional subject-matter fields of history and the scope and methods of historiography.

   Law -- includes jurisprudence, the major legal systems, legal theory, and the relationship of law to the other social sciences.

   Political science -- includes public administration, public law, international relations, comparative politics, political theory, and the study of policy making and political behaviour.

   Psychiatry -- includes theories and descriptions of the principal mental disorders and methods of diagnosis and treatment.

   Psychology -- includes clinical, counseling, educational, experimental, personality, physiological, social, and applied psychology.

   Sociology -- includes economic, organizational, political, rural, and urban sociology; the sociologies of knowledge, law, religion, and medicine; human ecology; the history of social thought, sociometry and other small-group research; survey research; and such special fields as criminology and demography.

   Statistics -- includes theoretical statistics, the design of experiments, non-sampling errors, sample surveys, government statistics, and the use of statistical methods in social science research.

2. Julius Gould and W.L. Kolb (Ed.) A Dictionary of the Social Sciences. (Compiled under the auspices of UNESCO). New York, Free Press of Glencoe, 1964.

This volume is the result of international meetings and national pilot projects (1952-1956) under the auspices of UNESCO to define the key concepts most widely employed in various social science disciplines. The experts recommended that the everyday usage of the terms defined should be given as well as the most widely accepted scientific usages, which should be illustrated by short quotations from the literature. The aim was to find synthetic scientific definitions that would constitute a common denominator in the different usages.

It was recognized that the selection of 1000 "general" concepts and terms used in the social sciences would be a difficult and, in part, arbitrary undertaking. Tentative selections were made from a study of the literature in the fields of political science, social anthropology, economics, social psychology, and sociology -- the aim being to select terms that were general and/or in some way basic to the discipline concerned.

For the most part terms were omitted that were unduly technical or appeared to be used only in the analysis of minor or local phenomena. As attempt was made to exclude those terms about whose meaning there was little dispute or where little could be added to a standard dictionary definition.

In deciding upon the number of terms to be drawn from each discipline, the editorswere guided by the desire to achieve a rough balance between the disciplines. Despite this, concepts from political science and sociology were in the majority. On the other hand, it was noted that many of the concepts were "general" in a special sense, in that they were used in two or more social science disciplines.

Each entry was divided into a number of sections:

A. giving the core meaning or meanings of the term as used in one or more of the social sciences

B. giving a historical background of these meanings and/or more detailed discussion

C/D/E etc. giving more historical background plus details of the controversies and divergencies of meaning. The aim was to clarify the extent and sources of divergence and to describe the many convergences that could be noted.

The entries were prepared by individual scholars, in some cases with "second opinions".

3. UNESCO. Main Trends of Research in the Social and Human Sciences. Paris, Unesco, (Part one: social sciences, 1970 819 p; Part two: human sciences, 1972?).Also in French edition.

These volumes are potentially useful as a guide to selecting concepts and sources. Part one is confined to the "law-

seeking" social sciences and covers sociology (Lazarsfeld), political science (Mackenzie), psychology (Piaget), economics (several authors), demography (Bourgeois-Pichat), linguistics (Jakobson), interdisciplinary problems (Piaget), mathematical models (Boudon), problem-focused research (de Bie), comparative research (Rokkan), and research policy (Trist).

4. International Committee for Social Sciences Documentation. International bibliography of the social sciences. London, Tavistock, 4 annual volumes (sociology, political science, economics, social and cultural anthropology).

These volumes do not, of course, attempt any synthesis or overview from which key concepts could be extracted. They do, however, have good indexes which could assist in the location of frequently-occurring concepts. The bibliographical entries are also classified into subject groupings in a helpful manner.

5. Key textbooks in each discipline.

In each discipline there are a few key textbooks which are recognized as offering the clearest insight into the concepts of the discipline.

6. Other sources.

Consideration could be given to using the major abstracts for each discipline to isolate the key concepts used with a certain frequency (e.g. Sociological Abstracts, Psychological Abstracts, etc.). The U.S. Dissertation Abstracts might also be useful.

7. Specialized multi-lingual dictionaries.

For example:

(1) Günter Haenich. Dictionary of international relations and politics; systematic and alphabetical in four languages (German/English/French/Spanish). Elsevier, 1965.

This dictionary has 5778 terms with equivalents in the four languages.

(2) I. Paenson. English/French/Spanish/Russian Systematic Glossary of Select Economic and Social Terms. Oxford, Pergammon, 1964.

Attempts to present a system of inter-related concepts which reflect a vertical hierarchy and are presented within a continuous text in a systematic exposition of a given subject (see Appendix D12).

International Organizations Possibly Interested

International Federation for Documentation (Classification Group)
International Sociological Association
International Political Science Association
Society for General Systems Research
European Centre for Coordination of Research and Documentation in the Social Sciences
Institute of International Law
International Association for Analytical Psychology
International Association of Legal Science
International Association of Applied Psychology
International Association of Individual Psychology
International Institute of Administrative Science
International Union of Psychological Science
International Commission for a History of the Scientific and Cultural Development of Mankind
International Committee for Historical Sciences
International Council for Philosophy and Humanistic Studies
International Institute of Sociology
International Law Association
International Institute of Philosophy
International Social Science Council
International Federation of Societies of Philosophy
International Society for General Semantics
International Union of Anthropological and Ethnological Sciences
Pugwash Conference on Science and World Affairs
International Union of Orientalists
World Academy of Art and Science
International Federation for Modern Languages and Literature
International Association of Universities
European Society of Culture
Interamerican Society of Psychology
International Bureau of Differential Anthropology

Relationship to the SATCOM Recommendations

In the U.S.A. the National Academy of Sciences and the National Academy of Engineering appointed a committee on Scientific and Technical Communication (SATCOM) which undertook a three-year survey (1966-68) funded by the National Science Foundation.

This resulted in a report (*) and recommendations.

Of major concern was the increase in the volume of such information, the emergence of new disciplines, and of new links between existing ones, and the increasing diversity of user groups and user needs. The Committee made recommendations with respect to improvements "to the structuring, flow and transfer of scientific and technical information and insight."

On the question of further research, the Committee reported that "More exciting than retrieval of information from a static store is evolutionary indexing, in which user's additions, modifications, restructuring, and critical commentaries steadily improve the initial indexing..." NSF funding of investigation into this approach was recommended.

The report laid great stress on the interdisciplinary information problem. "With the expansion of the body of recorded information, the likelihood that all the information which could be of use in a given operation will have its origin in the geographic, temporal or disciplinary neighborhood of (its) potential point of applica-tion decreases." It concludes that an area requiring further action "related to the slowly knitting, massive, mission-oriented programs of recent years which deal with major social concerns ...Economic, demographic and sociological information will have to be readily available and used in complete integration with engineering, geographic, and other relevant kinds of information."

(*) Scientific and Technical Communication; a pressing national problem and recommendations for its solution. Washington, National Academy of Sciences, 1969.

International Center for the Terminology of the Social Sciences (INTERCENTRE)

This centre, under the leadership of Dr. Isaac Paenson, has been active in the production of glossaries as an aid to translators and specialists in the economic and social disciplines. Dr. Paneson specifically makes the point that the glossaries are not "word lists with sets of definitions," but on the contrary are attempts to present a system of inter-related concepts which reflect a vertical hierarchy and are presented within a continuous text. He recognizes that the direct translation of a word or phrase from a foreign language is only part of the way to comprehension, and has aimed at conveying not merely the right word to use in a particular context, but also the idea the word expresses. Every effort is evidently made to ensure that the texts are authoritative. The glossaries produced so far have been supported financially or otherwise by contacts with Unesco and the International Social Science Council. An effort has been made to maintain contact with national and regional centres interested in these approaches.

Systematic Glossary of Selected Economic and Social Terms (English/French/Spanish/Russian). Pergamon Press, 1964.

Section headings: Demand, Production, Business Economics, Labour and Social Security Questions, Financial Questions, Economic Theories, International Trade, Alphabetic Index of Terms in all four languages.

The Glossary is almost entirely concerned with economic terms, the only social terms included are those with respect to social security. The general approach is to take, topic by topic, each aspect of economics and for each concept within each aspect to give a sentence or phrase and its equivalents in each language. There is not a great deal of emphasis on differences of meaning in specific terms. There are some exceptions to this; for example, different economic theories are given in some detail. "National income" is given with various alternative definitions. But in general, each term is treated as having a single meaning.

No attempt is made to show the interconnections between terms in any systematic way. The structure is not a very deep one, and is based on his breakdown of economics. Concepts are not broken down into sub-units, in the way that is planned in this project.

Systematic Glossary of the Terminology of Statistical Methods (English/French/Spanish/Russian). Pergamon Press, 1970.

This is prepared on the same basis as the previous volume. Again the emphasis is on providing a clear definition of each term within a systematic exposition of the subject. The presentation is given in the attached extract.

Glossary of Selected Terms in Public Law.

INTERCENTRE is now in the process of completing terms in this subject area, and will shortly be publishing the result.

Clearly, the work of this Centre could represent an important aid to this project, and it is also very probable that the output of the project would be of great use to the Centre in preparing further glossaries. There is, however, a very clear distinction between the concerns of this project, namely to clarify differences in meanings and to register different concepts, and the INTERCENTRE emphasis on defining terms within a specific framework.

C. PRELIMINARY TREATMENT OF DATA

(1) Before studying the various movements of **time-series**[1] it is essential to insure the comparability of the observations included by making necessary adjustments for, *inter alia*:

(a) **variations in the interval of time**[2]* /caused by fluctuations, resulting partly from the incidence of holidays, in the numbers of **calendar days**[3] or **working days**[4] per month, per quarter, etc./;

(b) **population changes**[5] /to eliminate their influence **time-series**[1] are often calculated on a **per capita basis**[6]/;

(c) **price changes**[7] /to eliminate their influence value data forming **time-series**[1] are divided /**deflated**/ by the **price** {**indexes**[8a] / **indices**[8b]} for the corresponding periods and thus **quantum indexes**[9] are established/.

C. TRAITEMENT PRÉLIMINAIRE DES DONNÉES

(1) Avant d'étudier les différents mouvements de **séries chronologiques**[1] il est essentiel de s'assurer de la comparabilité des observations effectuées, en procédant aux ajustements nécessaires pour tenir compte, entre autres, des :

(a) **variations de** {**l'intervalle de temps**[2a]* / **la période des observations**[2b]} /dues à des fluctuations qui résultent en partie de l'incidence des jours de fête sur le nombre de **jours civils**[3] ou de **jours ouvrables**[4] par mois, trimestre etc./ ;

(b) **variations de la population**[5] /pour éliminer leur influence, les **séries chronologiques**[1] sont souvent calculées **par tête**[6]/ ;

(c) **variations des prix**[7] /pour éliminer leur influence, les données de valeur constituant des **séries chronologiques**[1] sont divisées par les **indices de prix**[8] pour les périodes correspondantes, et les **indices de** {**quantités**[9a] / **volume**[9b] / **production**[9c]} sont ainsi établis/.

D. ANALYSIS OF THE (SECULAR) TREND

(1) The first step to be taken in the analysis of **time-series**[1] is to plot them on **ordinary** {**plotting**[2a] / **arithmetic**[2b]} **paper** or **semilogarithmic**[3] or **logarithmic**[4] **plotting paper** [s. Ch. IV, A.2., (2) and (3)]. The **secular trend**[5] may be {**fitted** / **drawn**} on such a chart—free-hand or with a transparent ruler—as a curve or as a straight line thus enabling us to study it as well as the deviations from it due to **cyclical**[6], **seasonal**[7] or **random factors**[8] [s. B., (1), above]. This method of **trend-fitting**[9] contains, however, a certain subjective element and is very approximate. If we want to make **forecasts**[10] based on the

* So-called 'calendar variations'; cf. A., (2), above.

D. L'ANALYSE DE LA TENDANCE (GÉNÉRALE)

(1) La première étape de l'analyse de **séries chronologiques**[1] consiste en leur représentation graphique sur du **papier** {**quadrillé**[2a] / **millimétré**[2b]} **ordinaire** ou du **papier quadrillé semi-logarithmique**[3] ou (doublement) **logarithmique**[4] [v. Ch. IV, A.2., (2) et (3)]. La **tendance générale**[5] peut être ajustée sur un tel graphique à main levée ou à l'aide d'une règle transparente suivant une courbe ou une droite, ce qui permet de l'étudier en même temps que les écarts par rapport à elle, dus aux **facteurs cycliques**[6], **saisonniers**[7] ou **accidentels**[8] [v. B., (1), ci-dessus]. Cette méthode d'**ajustement de la tendance**[9] contient, pourtant, un élément subjectif et est très approxi-

* Cf. A., (2), ci-dessus.

The concept of semantic fields (*)

The semantic field theory (*) was first put forward by Jost Tries and is based on the conception of fields as closely-knit sections of the vocabulary, in which a particular sphere is divided up, classified and organized in such a way that each element helps to delimit its neighbors and is delimited by them. Their contours fit into each other like pieces of different shapes in a mosaic. In each field, the raw material of experience is analysed and elaborated in a unique way, differing from one language to another and ofter from one period to another in the history of the same idiom. In this way, the structure of semantic fields embodies a specific philosophy and a scale of values.

The field theory has been strongly criticised from various quarters and some of the claims put forward by its champions are no doubt extravagant and unconvincing. The neatness with which words delimit each other and build up a kind of mosaic, without any gaps or overlaps, has been greatly exaggerated. This is true only of specialized and rigidly-defined systems; in ordinary language, vagueness,synonymy, ambiguity and similar factors will produce a much less tidy picture... Quite apart from overlaps between the various conceptual spheres, it is clear that many of these have no systematic organization of any kind.

These limitations must not, however, be allowed to obscure the outstanding importance of the field theory.

Ullmann cites: its introduction of a truly structural method into a branch of linguistics, the possibility of formulating problems which would otherwise pass unobserved; and a method of approach to the problem of the influence of language on thinking.

A semantic field does not merely reflect the ideas, values and outlook of contemporary society, but it crystallizes and perpetuates them: it hands down to the oncoming generation a ready-made analysis of experience through which the world will be viewed until the analysis becomes so palpably inadequate and out-of-date that the whole field has to be recast.

At this point, the field theory links up with another recent development in linguistics, the so-called Sapis-Whorf hypothe-

(*) This Appendix consists of extracts from: S. Ullman. Semantics; an introduction to the science of meaning. Oxford, Blackwell, p. 243-253.

(**) For recent work see: L. Weinberger, Von Weltbild der deutscher Spache. (Dusseldorf, 1953-54, 2nd edition). F. Hiorth. "On the relation between field research and lexicography." Studia Linguistica, 10 (1956), pp. 57-66.

thesis on the influence of language upon thought. (see Appendix E2) Modern philosophers, "linguistic analysts" and others, are deeply concerned about the possibility that some philosophical problems are pseudo-problems generated by the structure of our languages. Benjamin Lee Whorf approached the question in a novel and fruitful way: by comparing our own European languages --"Standard Average European", as he called them -- with the totally different structure of American Indian idioms. His researches convinced him that each language contains a "hidden metaphysics",that it embodies a unique way of viewing the world and imposes this outlook on its speakers. "The linguistic system of each language," he argued," is not merely a reproducing system for voicing ideas, but rather is itself the shaper of ideas, the program and guide for the individual's mental activity, for his analysis of impressions, for his systhesis of his mental stock-in-trade. We dissect nature along lines laid down by our native languages."

During the last few years, a new concept of semantic fields has been evolved by the French linguist Georges Matoré: c'est en partant de l'étude du vocabulaire que nous essaierons d'expliquer une société. Aussi pourrons-nous definir la lexicologie comme une discipline sociologique utilisant le materiel linguistique que sont les mots." His teaching and example have also stimulated a number of enquiries into specialized vocabularies, ranging from feudalism to railways and from fashion to medicine. These investigations have found a focal point in the recently-established centre of lexicological studies at Besançon, where large-scale research projects are being organized with the aid of modern mechanical devices. (*)

(*) The centre, headed by B. Quemada, publishes a series entitled,"Cahiers de Lexicologie,"and a "Bulletin d'Information du Laboratoire d'Analyse Lexicologique".

Use of several languages and translation problems.

A. Absence of a lingua franca.

It would be optimistic to expect wide acceptance of the system if it was based on one language only. The UNISIST Study notes (pp. 72-73) that:

(i) English now accounts for about 40% of the world literature, regularly yielding (as are French and German) to the rising group of "Eastern" languages e.g. Slavic, Chinese and Japanese.

(ii) No one can predict what the situation will be twenty or fifty years ahead, nor does anyone possess reliable data on the present use of foreign language materials in the scientific community (but see Appendix E2,section E).

(iii) The position of English as a lingua franca of science is contested by some governments either to consolidate a new country via a national language or in the belief that language can be artificially maintained as a vehicle of a culture.

(iv) The chances of securing international acceptance of English as the standard language of science are, in present circumstances, very poor.

B. Language preferences.

Apart from these aspects, there is the extremely serious problem that social scientists in one language group tend to either ignore foreign language material or find it "less relevant" to their particular concerns. This is particularly significant across the English, French, German divide. Concepts given in foreign languages may be difficult to comprehend if one is less than completely at home with the language in question. An unconscious hostility to concepts expressed in foreign languages may even build up (see section E).

C. Language group incompatabilities.

There is also the possibility that a concept may first be expressed or may only be expressible, in a given foreign language. It would be an advantage to be able to file it as such and worry about the translation afterwards. The author who has done much to emphasize the difficult-to-comprehend contrasts between meanings in the standard Indo-European languages and those in other language groups is Benjamin Lee Whorf. These contrasts are well-illustrated in the following extract from one of his papers (*):

"The growth of the Indo-European language-culture complex dates from ancient times. Much of its metaphorical reference to the nonspatial by the spatial

(*) B.L. Whorf. Language, Thought, and Reality. New York, Wiley, 1958, 278 p. ("The relation of habitual thought and behaviour to language") was already fixed in the ancient tongues, and more especially in Latin. It is indeed a marked trait of Latin. If we compare, say Hebrew, we find that, while Hebrew has some allusion to not-space as space, Latin has more. Latin terms for nonspatials, like educo, religio, principia, comprehendo, are usually metaphorized physical references: lead out, tying back, etc. This is not true of all languages -- it is quite untrue of Hopi. The fact that in Latin the direction of development happened to be from spatial to nonspatial (partly because of secondary stimulation to abstract thinking when the intellectually crude Romans encountered Greek culture) and that later tongues were strongly stimulated to mimic Latin, seems a likely reason for a belief, which still lingers on among linguists, that this is the natural direction of semantic change in all languages, and for the persistent notion in Western learned circles (in strong contrast to Eastern ones) that objective experience is prior to subjective. Philosophies make a strong case for the reverse, and certainly the direction of development is sometimes the reverse. Thus the Hopi word for "heart" can be shown to be a late formation within Hopi from a root meaning think or remember. Or consider what has happened to the word "radio" in such a sentence as "he bought a new radio" as compared to its prior meaning "science of wireless telephony."

"To sum up the matter, concepts of "time" and "matter" are not given in substantially the same form by experience to all men but depend upon the nature of the language or languages through the use of which they have been developed. They do not depend so much upon any one system (e.g. tense, or nouns) within the grammar as upon the ways of analyzing and reporting experience which have become fixed in the language as integrated "fashions of speaking" and which cut across the typical grammatical classifications, so that such a "fashion" may include lexical, morphological, syntactic, and otherwise systemically diverse means coordinated in a certain framework of consistency. Our own "time" differs markedly from Hopi "duration". It is conceived as like a space of strictly limited dimensions, or sometimes as like a motion upon such a space, and employed as an intellectual tool accordingly. Hopi "duration" seems to be inconceivable in terms of space or motion, being the mode in which life differs from form, and consciousness *in toto* from the spatial elements of consciousness. Certain ideas born of our own time-concept, such as that of absolute simultaneity, would be either very difficult to express or impossible and devoid of meaning under the Hopi conception, and would be replaced by operational concepts. Our "matter" is the physical subtype of "substance" or "stuff" which is conceived as the formless extensional item that must be joined with form before there can be real existence. In Hopi there

seems to be nothing corresponding to it; there are no formless extensional items; existence may or may not have form, but what it also has, with or without form, is intensity and duration, these being nonextensional and at bottom the same."

The differences are not restricted to high level abstractions such as "time" and "matter" but may permeate the whole perspective. The famous hypothesis associated with the work of von Humboldt, Sapir and formalized by Whorf suggests:

> "that the commonly held belief that the cognitive processes of all human beings possess a common logical structure which operated prior to and independently of communication through language, is erroneous. It is Whorf's view that the linguistic patterns themselves determine what the individual perceives in this world and how he thinks about it. Since these patterns vary widely, the modes of thinking and perceiving in groups utilizing different linguistic systems will result in basically different world views." (1)

> "We are thus introduced to a new principle of relativity which holds that all observers are not led by the same physical evidence to the same picture of the universe, unless their linguistic backgrounds are similar... We cut up and organize the spread and flow of events as we do largely because, through our mother tongue, we are parties of an agreement to do so, not because nature itself is segmented in exactly that way for all to see" (2)

Each language becomes a classification and organization of experience in its own right. As such each may be significantly different from the other and may structure the forms and categories by which the individual not only communicates but also analyzes nature, perceives or neglects particular phenomena or relationships, and constructs his model of the world (3).

A striking example of the possible differences is given by Marshall Walker in discussing the social factors which affect scientific models:

> "The language of the Wintu Indians of California seems to indicate a way of thinking quite different from our own. Imagine the surface of a table with a

---

1 F. Fearing. "An examination of the conceptions of Benjamin Whorf in the light of theories of perception and cognition". In H. Hoijer (Ed.) Language in Culture. American Anthropologist, 56, (1954), Memoir 79, 47.

2 B.L. Whorf. Collected Papers on Metalinguistics. Washington, Foreign Service Institute, Department of State, 1952.

3 See Wharf, *op.cit*.

book lying on it. The remainder of the surface is bare. In English one describes the situation by saying, "The book is on the table." In Wintu one says, "The table bumps". The English phrase has already committed the speaker to an entire analytical philosophy of the situation: (1) there are two objects; (2) there is a polarity such that one object is above the other; (3) there is an implication that the book is supported by the table. None of this analysis is present in the Wintu sentence, which is purely topological.... The scientist who wishes to be as objective as possible in his study of the external world will try to free himself from the possible constraints of his own language." (1)

Such languages may not have parts of speech or separate subject and predicate. In Indian Languages such as Nootke and Hopi events as a whole are signified. Instead of "a light flashed" or "it flashed", Hopi uses a single term, "flash" to signify that a happening has occurred. There is thus no distinction between tenses, for the Hopi has no general notion or intuition of time as a smooth flowing continuum in which everything in the universe proceeds at an equal rate, out of a future, through a present, into a past. The focus is rather on the totality accessible to the senses at a given moment with no distinction between present, past, or even the future or physically distant where the latter are accessible or represented in memory. Navaho is different again with little development of tenses by an emphasis on types of activity or aspects of action. The first concern of Indo-European languages can be defined as time; of Hopi, the validity a statement has (in terms of fact, memory, expectation, or custom); and of Navaho , the type of activity.(2)

Von Bertalanffy suggests that the Whorfian hypothesis may be extended. He argues that the categories of knowledge depend on biological and cultural factors. In particular, he argues that Aristotelean logic actually covers only the extremely small field of subject-predicate relations. The all-or-none concepts of traditional logic fall short of continuity concepts basic for mathematical analysis. He is with Whorf in hoping that other languages may permit basically different kinds of "science" which would represent other aspects of reality as well or even better than does the current scientific world picture.

The suggestion has been made, for example, that a language like Hopi might be better suited to verbalizing the concepts of modern physics than English. But some of the non-Indo-

[1] Marshall Walker. The Nature of Scientific Thought. Prentice-Hall, 1963, p.103

[2] These points are based on Ludwig von Bertalenffy's account of the Whorfian hypothesis in: "The relativity of categories" in: General Systems Theory, New York, Braziller, 1968.

European languages may also have important and hitherto unknown concepts concerning the functioning of social processes -- an area in which continuity is even more vital to understanding than is the natural sciences.

Some languages may in fact constitute rich sources of concepts which could prove useful to the understanding of organized social complexity. Little work seems to have been done on this possibility -- most of the examples refer to contrasts of interest to the natural science perspective. In fact the field of comparative linguistics seems to be made up of "one shot" studies with very little comparison. Where comparisons are made it is at the formal rather than the conceptual level (1), so that with the exception of a few startling examples which augur for a fascinating variety of thinking styles, little information is available. It may be that few linguists are competent to write on the concepts of more than one or two non-Indo-European languages, so that no wide-ranging study or classification is possible, and no "handbook" is available. The absence of such a study only helps to conceal the many differences from the Indo-European perspective-- the existence of such differences is certainly not widely recognized (2).

The whole argument raises the possibility that the computer record design envisaged (see Appendix A6) would not be sufficiently general and flexible to be able to "contain" the concepts of some other language groups. The either-or distinction between "entities" and "relationships" may only amount to a magnificient exercise in handling Aristotelian "substance" and "attribute" as represented in Indo-European nouns and predicate adjectives. Is the concept of distinct,

[1] One reason is that a major school of linguistics denies the need to consider "semantics" and "concepts", claiming that all understanding relevant to the discipline can be gained from analysis of syntax. A second reason may be, as Sapir has argued, that many linguists consider such languages "primitive" and therefore unlikely to constitute a source of concepts unknown to the Indo-European culture.

[2] Marshall Walker (The Nature of Scientific Thought) notes (p.103-4): "The student of science also has a vital need for comparative linguistics in order to acquire experience in the isolation of concepts from their language matrix. The usual language departments of a university are not much help for this type of study....There is need for a course for undergraduates (not language majors) which is designed to illustrate the expression of concepts by different language families. Pending the arrival of such courses the student of science will have to do it himself as best he can."

persisting "entities" common to all languages and can all concepts of "relationships" be adequately represented by graph-theoretic type areas?

David Bohm, a theoretical physicist interested in Piaget's and Gibson's work on the problems of perception, gives detailed arguments against permanence of "entities" and concludes (1):

> "it is clear that both in common experience and in scientific investigations, the objects, entities, substances, etc., that we actually experience, perceive, or observe have always (thus far) shown themselves to be only relatively invariant in their properties, this relative invariance having often been mistaken for absolute permanence" (p.14)

> "It is evident then that by considering entities and structures as relatively invariant, with an as-yet-unknown domain of invariance, we avoid making unnecessary and unprovable assumptions concerning their absolute invariance. Such a procedure has enormous advantages in research, because one of the main sources of difficulty in the development of new concepts -- not only in physics but also in the whole of science -- has been the tendency to hold onto old concepts beyond their domain of validity." (p.121-2)

Colin Cherry, a telecommunications engineer interested in the psychology of communication with developing countries, considers that relationships may not be meaningfully represented by graph-theoretic links and that other forms of representation might be preferable.

One response is in the work on linguistic universals. It is suggested that terms exist in all languages to designate objects which meet a condition of spatio-temporal contiguity. And, in general, that all languages are cut to the same pattern without there necessarily being any point by point correspondence between particular languages (2). It is recognized that work in this area is only at the early stages (3). A close look at the logical assumptions built into the computer record design seems to be necessary.

It would seem important to avoid losing the richness of alternative perspectives by confining this project to one or two languages in one language group -- particularly as the concepts inventoried are supposed to be in some way relevant to the cultures using such languages. That this is significant is indicated by the fact that --% of the world's population currently uses non-Indo-European languages (4). This includes the Chinese, who are unlikely to remain a

1 David Bohm. The Special Theory of Relativity. N.Y., Benjamin, 1965.

2 Noam Chomsky. Aspects of the Theory of Syntax. Cambridge, MIT, p. 29-30.

3 See:J.H.Greenberg (ed.)Universals of Language.Cambridge,MIT, 1963.

4

minor influence on world society. The argument that many learn an Indo-European "second" language is weak in that being present in classes at which such a second language is taught or used is no evidence that the language and its perspective "take" in the individual -- as most school leavers know. Even if they do take, it is questionable whether it is satisfactory to ignore the individual's problems of translating between the two conceptual systems.(1)

D. Problems of translation.

It may astonish many people to know that contemporary linguistics has concluded that translation between languages is theoretically impossible. Chomsky notes (p.202):

> "In fact, although there is much reason to believe that languages are to a significant extent cast in the same mold, there is little reason to suppose that reasonable procedures (not involving extralinguistic information) of translation are in general possible"

Georges Mounin, who notes the same conclusion, has summarized the theoretical difficulties prior to considering why, how, and within what limits the practical operation of translations is relatively possible (1).

Some of the difficulties he notes argue against any attempt to force this project into a unilingual mode.

- --certain languages have highly developed terminologies in areas where there are few Indo-European equivalents (e.g. the Pyalluo Indians and "salmon"; the Eskimos and "snow" (30 terms), some African languages and "palm trees", the Argentine gauchos and "horse colouring" (200)). There is little value in attempting a definitive translation when no exact equivalent exists.
- --the situation becomes more complex when dealing with socio-cultural terms, e.g. how can "brother" and sister be translated into Maya when that language only has terms for "younger brother" or "older brother" (2) Much closer to the concerns of this project is the simple problem of translating "people's capitalism" into French (3).

1 Georges Mounin. Les problèmes théoretiques de la traduction. Paris, Gallimard, 1963.

2 A special issue of the ETC (Institute of General Semantics),15,2, March 1958 is entirely devoted to interpretation and intercultural communication. It gives many examples of this sort of problem.

3. Georges Mounin, op.cit. p.67-68.

-- another excellent example, noted by Colin Cherry (XVV) is that whilst there is no difficulty in translating the colour "red" into and from Russian, the association in the two languages are very different. In English: blood red, red in tooth and claw, red with anger, red light district, etc. In Russian the translation of "red" is synonymous with "beautiful" and has associations equivalent to the English "golden" -- hence "Red Square" and the "Red Army" should be meaningfully translated as the "Golden Square" and the "Golden Army". How much has international tension been aggravated and reinforced by this simple error? Similarly, In Chinese, "red" is primarily associated with "joy", "prosperity", "luck", and "happiness" (*). Thus greeting cards, invitations, decorations, etc., are usually in red. (To what extent have the positive associations of the colour in the two cultures influences the marked success of socialism there, compared to that in Anglo-Saxon culture, where it has more negative association?)

E. Use of foreign language material by social scientists

A recent study of 1000 social science research information users in Great Britain has just been completed (**). It shows that 18% of the sample read English only, 75% read French, and 27% read German. Of those who said they were able to read a foreign language, only one-third regularly scan literature in that language. There is even a reluctance to follow up articles in another language.

It was also noted that 22% make no use of abstracts or indexes, 35% never use bibliographies, 22% do not use library catalogues, and 48% do not consult the librarian.

F. Administrative delays

If the attempt is made to translate every theoretical formulation into English, before filing, there will be a hold-up similar to that associated with the modelling activity. There is also bound to be disagreement as to the adequacy of translations. It may be preferable therefore to conceive of a Translation Phase in parallel with the filing, modelling, and term allocation phases, and to give priorities to the translation of given terms according to need.

(*) I am grateful to Mr. Thai Wo Tsan for this information.

(**) Maurice Line (Ed). Information Requirements of Researchers in the Social Sciences. Bath University, 1971, 2 vols.

A discipline's model as a "language"

For each entity or class of entities within a discipline we can attempt to indicate which other entities, relationships, attributes may be associated with it (i.e., in a political science parlance, what entities or classes of entities can the attribute "democratic" possibly be meaningfully applied to?). In a given discipline the number of such permissable relationships should be quite limited for a given concept -- and even if the number is large it can be reduced by redefining the entities in question as a class or classes.

This approach permits us to define the possible meaningful subunits of propositions which may be generated from the discipline vocabulary. In effect extremely stringent "grammatical" rules are thus set up for the creation of valid elements of sentences in the discipline jargon.

By this approach, extra conditions on sentence formulation are imposed, based on the knowledge obtained by use of the disciplines. In natural language such conditions cannot be imposed because many meaningless sentences in natural language are grammatically correct. Because of the richness and redundencies, size of vocabulary and indifference to truth values in natural language, any natural language project would lead to very large numbers of permissable relations which would be of little use, besides being impracticable. (Basically, in natural language, any adjective may be used to qualify any entity in the class of "nouns"; any adverb may be used with any entity in the class of "verbs", etc. -- and it is difficult to introduce restrictions at a more detailed level.)

Once the sub-propositional units have been determined (perhaps as "concept pairs") these in effect amount to new compound concepts (e.g. "voting procedure"). These may either be registered as new concepts or left as permissable "generateable" concepts (merely indicating the relationship between the component units, rather than showing the component units as components of a new compound unit). The choice would depend on the frequency of usage of the composed unit.

The procedure may then be repeated for the compound units, where this is considered useful, in order to build up the more complex permissable units commonly encountered. Clearly at a certain point in a particular domain:

-- it becomes difficult to determine whether higher level concepts are permissable because there meaningfulness is as yet untested, i.e. they are "new

-- it is decreasingly useful to create new units because of the quantity

In certain cases, however, the build-up can continue to the level of defining permissable propositions, i.e., a propositional inventory is built up from the units. Other sub-units are held in such a form that many probable propositions may be generated automatically for inspection and possible coding as requiring investigation, meaningless, false, etc. This procedure introduces further rules restricting the manner in which the units may be combined. Modification and additions may of course be made as new insights and data are obtained.

Once the concepts of a discipline are held in this structured form, some interesting investigations of levels of analysis and degree of equivalence may be made. Where a set of propositions exists employing a given entity (e.g. "nation") which is itself made up of sub-units (e.g."provinces"), or is itself a sub-unit of a larger entity (e.g.,"continental region") (*). "new" propositions may be systematically generated for the higher or lower level by treating the terms as equivalent. These propositions may then be inspected as before, to eliminate the obviously meaningless and inapplicable at the new level. The remaining propositions may be added to the inventory if required.

A similar approach may be adopted <u>between</u> disciplines. In some cases new insights may be suggested by treating key entities in different disciplines as equivalent and substituting the entity from the second discipline into the propositions of the first containing the proposed counterpart. (e.g.,"individual" in psychology may be substituted for "nation" in political science or vice versa; "cell"from biology for "organization" in organization theory(**)). In the case of a given set of propositions containing a limited number of concepts, equivalents for many of the concepts may be selected from the second discipline, so that only the formal structure of the first discipline proposition is retained. This amounts to a general system investigation of propositional invariance or isomorphy across discipline boundaries -- without the need to define any questionable "meta language" in which the isomorphy is established (***).

> "In fact, similar concepts, models and laws have often appeared in widely different fields, independently and based upon totally different facts. There are many instances where identical principles were discovered several times because the workers in one field were unaware that the theoretical structure required was already well developed in some other field." (*, p.33-34)
>
> "Thus, there exist models, principles, and laws that apply to generalized systems or their subclasses, irrespective of their particular kind, the nature of their component elements, and the relations or "forces" between them ... we can ask for principles applying to systems in general, irrespective of whether they are of physical, biological or sociological nature.... To take a simple example, an exponential law of growth applies to certain bacterial cells, to populations of bacteria, of animals or humans, and to the progress of scientific research measured by the number of publications in genetics or science in general. The entities in question, such as bacteria, animals men, books, etc., are completely different, and so are the causal mechanisms involved. Nevertheless, the mathematical law is the same." (*, p. 32-33)

Clearly, investigation of propositions generated by this technique would facilitate the elimination of meaningless and false propositions, and the identification of isomorphisms as distinct from superficial analogies.(**)

(*) On "levels of analysis" with respect to international studies, see Henry Teune,"Conceptual dimensions of linking international and comparative research" (Paper presented to the International Conference on the Relationship of Comparative and International Studies, Bellagio, 1971)

(**) See M.Haire. Biological models and empirical histories of the growth of organizations,in:M.Haire (ed),<u>Modern Organization Theory</u>. New York, Wiley, 1959, pp.272-306.

(***) Some very interesting math for the detection of such isomorphy is given in:P.Jaffard et.G. Poitou, "Introduction aux catégories et aux problémes universels." Paris, Ediscience, 1971.

(*) L. von Bertalanffy. General System Theory. New York, Braziller, 1968.

(**) L. von Bertalanffy, op.cit., p.84-5.

Future prospects: an ideal knowledge-representation system

The ideal "information" system in a given academic field has been sketched out as follows by the U.S. National Academy of Science Committee on Information in the Behavioural Sciences under the chairmanship of David Easton. The ideal is here portrayed (*) as a "computer analogue of the available, intelligent, and informed colleague."

> "Such an ideal colleague would read widely, have total recall, evaluate what he read; he would be able to reorganize materials, recognize fruitful analogies, and synthesize new ideas. In addition the ideal colleague would always be accessible and available to all, either in person or by phone. Finally, such a colleague would be sensitive to each research worker's needs. He would be aware of the general interests and current problems of each scientist, and he could adapt both the context and style of his communication to each researcher's knowledge, skills, and habits." (**)

There have been many reports on the improvement and integration of information systems and it would be futile and inappropriate to comment on them here. There seems, however, to have been little mention of what might be termed a "knowledge-representation" system (***). The ideal colleague above would be the key component in a knowledge-representation system -- he would, it is suggested, have no place in an "information" or "documentation" system as they are currently conceived. This Appendix attempts to clarify the distinction between the knowledge-oriented and document-oriented approaches to system design by comparing the functioning of a hypothetical knowledge-oriented system, now technically feasible, with the current approach. The intention is not to imply that the former should replace the latter but rather that the former offers various means of avoiding some of the key problems faced by the latter -- the two are however complementary. The distinction is basically between a synthesis or atomisation in

(*) Cited in the preface to: L. Larry Leonard (Chairman). Report and recommendations toward an international studies integrated information system. International Studies Association, Committee on Bibliographical and Documentation Services, 1969.

(**) David Easton (Chairman). Communication System and Resources in the Behavioural Sciences; by the Committee on Information in the Behavioural Sciences, Division of Behavioural Sciences of the National Research Council, Washington, DC, National Academy of Sciences, (Pub.1575), 1967, p.46.

(***) "Knowledge-representation could be considered to mean "information", but there are so many other interpretations of the latter that the new term seems appropriate here.

the handling of information as noted by J.M. Ziman:

> "I cannot emphasize too strongly the importance of this activity of intellectual synthesis... Any notion that we may have about the nature of science includes the belief that something like an overall pattern is to be discovered and described. What we need is scientific knowledge -- not more and more miscellaneous and unrelated information. The starting point for a search should not have to be an abstract journal or a computerized retrieval system -- it should be an encyclopaedic treatise or textbook where the information has been transformed into an intelligible pattern of thought...from which can be deduced the characterization of the particular datum, specimen or phenomenon that we are studying." (*)

The comparison is done in parallel column for ease of understanding.

(*) J.M. Ziman. "Information, communication, knowledge." Nature, 224, 25 Oct 1969, p.323

Information	Knowledge-representation
1. Index tends to be based on simple hierarchy or alphabetic listing of subject, author and title, which can be handled on catalogue cards.	1. "Index" constitutes a complex network giving a representation of entities and relationships and the dynamics of any points under debate which can only be handled by multi-dimensional computer programming techniques.
2. Users want documents; the index is a temporary inconveninece to gain access to the document.	2. Users want access to the "network index" which represents the items of knowledge and their relationships which they need; documents are a temporary inconvenience if it is necessary to re-examine data and detailed arguments justifying the entities and relationships incorporated. Document access is a secondary problem for which a documentation system may be used.
3. Author has "published" when document is in circulation and "available"; index entries of little significance to author.	3. Author has "published" when the appropriate knowledge structure in the "index" has been modified; incorporation in "index" (through a terminal) is a high priority for the author.
4. Research is conducted primarily using documents as a stimulus to creativity.	4. Research is conducted primarily using the knowledge-representation structure as a stimulus to creativity, i.e., on the graphical representation.
5. Access to knowledge via documents, means multiple reproduction and transfer of documents to a variety of libraries where they may or may not be used.	5. Access to knowledge is direct and does not require reproduction and transfer of documents. (Only one copy of the document justifying the amendment need exist on microfiche so that copies need only be prepared when the data and arguments must be re-examined in detail.)
6. Documentation system is embarrassed when faced with obtaining "ephemeral" or "phantom" material which has not been made commercially available through the few standard channels.	6. See 5.
7. Out-of-date, rejected, low quality, false, old documents are retained in the system and index with no index indication to that effect.	7. Out-of-date, rejected, false etc. entities or relationships are eliminated from the system by listing them on paper (or other "documents") with the bibliographical source from which they were obtained (i.e. they are available if required but do not clog the system).

8. Only the knowledge held in the documents physically available is accessible. The index only notes the documents held in the documentation centre in question.	8. All knowledge is on-line, although the supporting documents may not be physically accessible
9. Thinking momentum is constantly interrupted when access to new documents is required. (Long delays, 2-3 months, are normal; 50 months or more from initiation of research to appearance in abstracts)	9. Thinking momentum is maintained since the essence of any new domains of knowledge is always accessible -- all the links and entities are there (Delays are measured in seconds).
10. Authors status, pride and interest associated with visible document on some library shelves.	10. Author's status, pride and interest are associated with the visible entities and links in the graph representation accessible to all.
11. Author's domain of interest and home "territory" are unclearly defined.	11. Author's domain of interest and home "territory" are visibly defined.
12. The key figures in a discipline and the relationship between their spheres of influence are unclear.	12. The "luminaries" in a particular discipline are all visible together with the relationship between their spheres of influence.
13. Alternative concepts or contradictory evidence can be conveniently ignored in a document or textbook without too much risk -- particularly where the counter argument comes from another discipline (or a school of thought publishing in a different language).	13. Alternative concepts, relationships or contradicting evidence is immediately forced on one's attention -- even in the case of relationships linking to other disciplines.
14. Interdisciplinary links are ignored if the author has no interest in them.	14. Interdisciplinary links are already held in position whether the author wants to ignore them or not.
15. Documents carry a lot of text which is verbal packaging for the main points, or didactic in intent, used to honour the elders of the profession, or provided in order to define the frame of discourse. Much of it is repeated in other documents on the same point.	15. Non-essential material is unnecessary because the points are in many cases already embedded in the knowledge-representation system. Arguments can be directed specifically to the use and relationships between particular entities. Such compacted arguments might also be directly accessible on call -- but only as a clarifying presentation.

16. Any panoramic summary of knowledge in a discipline -- the standard textbook-- must remasticate all the extant views which are visibly significant from the author's perspective. The author must "redo" the whole discipline environment to provide the framework for any new contributions of his own. There is no guarantee that the rephrasing (necessary for status and copyright reasons) of other people's arguments will make them any clearer. One result is to add a large wad of duplicate material to the documentation system, often of doubtful literary quality.

17. The documentation system does not permit of any permanent representation of knowledge in a particular domain. Each verbal summary extant at a particular moment is under criticism and subject to reserve from different schools of thought within the discipline. In this important respect a document arising from a single group of authors can never contain the totality of views in a domain of knowledge. It is only the non-concretized interaction between a succession of documents which approximates to it. These invisible qualifiers on any document are a feature of the "collective mentality" of the members of the discipline. The knowledge of the discipline at any moment is very much in (and between) the hands of its members rather than on paper or in a row of books.

16. The panoramic view of the entities and relationships in the discipline is always available and up to date. The author's extra contribution is all that needs to be added - he does not need to recap the whole environment.
Since the academic's status is bound up with his specific modifications to the knowledge structure and not the verbalizations held in a document, the problem of adequate verbalization may be handled separately; Hopefully a limited number of skilled verbal presentations, from a minimum number of different perspectives and literary styles, could be constantly updated by professional writers using the best verbal arguments by any appropriate academics where appropriate.

17. Each entity, link, and qualification is indicated in the knowledge representation system. In effect one "layer" of the "collective mentality" of a discipline is rendered visible. Each modification to knowledge in the domain is entered on an hour-by-hour basis.

18. Different styles of documents are produced on the same topic for research, education, public information, program management, policy making, etc. purposes. The same material is repeated, with some extensions and some omissions, for each audience. But because it generally requires a person with a different style of thought to present each type of document, lags in the incorporation of the latest arguments or vital new evidence tend to be evident, so that there may be marked differences between the entities and relationships incorporated into each. This leads to a "spastic" or "aphasic" response to new situations, by different portions of society.

No attempt is made in a document established for one purpose to relate the elements of knowledge to those of other purposes.

19. The documentation problem is aggravated by the "publish or perish" code which governs much of academic life. Unless an academic publishes, he is "invisible" --he loses status in the eyes of his superiors. A curriculum vitae is judged as much on the number of artilces, books, etc., as on the quality.

20. Disciplines are psycho-social groups in which professional status and advances in knowledge are intimately related. At present intra-disciplinary communication is via documents for the knowledge advances, but the status and credibility of particular documents, and their authors are governed by ongoing informal word of mouth communication centered upon elders who set the fashions and designate ap-

18. The entities and relationships entered on the basis of research insights are also used for other purposes. Instead of producing different documents and reprocessing the insights, different "filters" are used in presenting or displaying the entities and relationships to different audiences. In this way, each new research insight is immediately incorporated into each other form of knowledge-representation -- each portion of society works from the same data base. (Problems registered by non-research bodies are immediately evident as a challenge to research.)

In this way if an element of knowledge represented cannot be understood, the user merely calls for a new method of representation (of the same knowledge), possibly using isomorphs (or even analogies) from a domain with which he is familiar. At any point he can move into a programmed learning mode and work from simple representations.

19. By switching emphasis to the specific entities and relationships which the academic has formulated, successfully, confirmed or criticized -- his status is determined by the bonds and entities with which he is associated. Each of his contributions is "visible" until it is superseded and is not subject to the vagaries of the documentation system.

20. In the knowledge representation system, it is quite evident which issues are currently under debate and the manner in which the demise of a set of entities and relationships will entrain the fall of a whole set of dependent elements. It is also evident who are the key proponents or opponents -- directly or indirectly -- of particular knowledge elements. Ideally the knowledge representation system would also act as a continually updated voting board for each entity and relationship. Each addition to the structure

20.) proved new fashions (and thus provide a needed element of stability). The procedure may be fairly democratic in that on each topic there are invisible collages of proponent and opposition "parties" in a "lower house", each with an eloquent voting constituency. The approval of the "upper house" of elders is required. It is by this on-going formal-informal debating mechanism that the disciplines stance at any one time is determined. But the channels by which members of a discipline are exposed to new views and indicate or withdraw their support, are controlled, sometimes rather undemocratically, by well-placed elders. There is a tendency for new and contrary views to have difficulty in obtaining a hearing. This may slow the development of the discipline and make it somewhat dependent upon a form of intellectual nepotism and "smoke-filled club room" democracy.

(20.) of knowledge would cause some individuals in the profession to indicate a modification in their pattern of allegiance. At any one time it is then evident how much support a particular knowledge element can muster and exactly where the weak links in the chain of support are. The vague auras of influence which are symptomatic of the document-oriented system are replaced by a precise picture of the state of the game. Contrary views are represented on an exactly equal basis and are not subject to the vagaries of the journal referee system.

21. Many academics subscribe to the building block approach to the advance of knowledge - particularly in the natural sciences (e.g.chemistry can be considered to be a sky-scraper under construction, with 30 floors completed and in use, the 31st and 32nd under construction. Partitions in the lower floors are modified as required by new insights. In the more human sciences, the view might be that each academic constructs his own mansion inspired by the elements of the style of his neighbors and predecessors.) This is only a useful metaphor, however, since there are no rec recognized "building blocks"

21. The "building blocks" are the entities and relationships added to the network which constitutes the "building". Any part of the network can be displayed as a visible representation of the "building" on which academics are working. This has the advantage of being a multidimensional dynamic structure in which any element can be questioned and modified without endangering the whole. The emphasis is on a community adding entities and links to a single existing visible whole, however many levels and domains it may be subdivided into -- individual initiative, whatever its domain, is related to that of the whole.

(21.) and no concretized "building" -- it also suffers from the severe disadvantage of constituting a "frozen pyramid" concept of knowledge categories. Each individual does his own thing with no attempt to relate it to the whole.

22. The forum of academic debate is concretized as a scattering of journals and other documents. There is little interaction between the journals but the debate is somewhat summarized in a scattering of abstracts in which the contents index gives some indication of the interventions on related topics.

22. The knowledge representation system constitutes a thinking forum in which the juxtaposition of relevant ideas from all sources is maximized. The researcher is exposed to a pattern of theoretical formulations in the process of being continually improved, and to which he can contribute. A dozen or more specialists in a particular field (the "invisible college" for that topic) cannot contribute simultaneously to ideas being written on one memo pad. They can do so via electronic dialogue support systems which help them to respond to each other's ideas (even if they are a continent apart) with a rapidity that allows each of them to maintain thinking momentum. Even in such a rapid debate the paternity of each emerging formulation is identified and registered. This mode of operation should be compared with some discussions between academics interested in the same topic in which progress is frustrated because if someone thinks of a good idea he wants to "publish" it (to gain credit) before contributing to the thinking momentum of his colleagues - this may take months.