EMPIRE INFORMATION SYSTEM
PHASE III
FINAL TECHNICAL REPORT

February 1967

Electromagnetic Compatibility Analysis Branch
Naval Weapons Laboratory
Dahlgren, Virginia

Contract N178-8883

HOWARD RESEARCH DIVISION
CONTROL DATA CORPORATION
7735 OLD GEORGETOWN ROAD
BETHESDA, MARYLAND 20014

EMPIRE INFORMATION SYSTEM

PHASE III

FINAL TECHNICAL REPORT

February 1967

Electromagnetic Compatibility Analysis Branch
Naval Weapons Laboratory
Dahlgren, Virginia

Contract N178-8883

HOWARD RESEARCH DIVISION
CONTROL DATA CORPORATION
7735 Old Georgetown Road
Bethesda, Maryland

ABSTRACT

The primary objectives of Phase III were to evaluate the system as a test data base for future research and development and to broaden the system base for testing its operational capabilities to support EMC information requirements. This report describes the continuing activities of the EMPIRE program, including innovations or modifications, mechanisms and rationale applicable to acquisitions, indexing/abstracting, thesaurus revision and retrieval implementation. Test and evaluation for intra- and inter-indexing consistency, indexing comparison with other systems, and the utilization of synthetic queries are discussed. Finally, recommendations for continued research and development are presented.

OUTLINE

SECTION I

1.0 BACKGROUND

1.1 Inception

EMPIRE had its inception in June of 1963 when the Naval Weapons Laboratory granted a contract to Howard Research to develop a technical information support system for the Weapons Interference Reduction Effort (WIRE). This system, known as EMPIRE, an acronym for Electro-Magnetic Pheromena Interference Repository, went through a development phase which provided the basic EMC Thesaurus, abstracted and indexed 250 documents, and outlined an implementation schedule.

1.2 Phases I and II

Phase I, February through July 1964, accomplished the following objectives:

- Indexed and abstracted an additional 850 documents.
- Further refined the EMC Thesaurus.
- Redefined system parameters and procedures.
- Produced catalog cards, technical abstract listings, and an inverted file.
- Sketched a search scheme for use with an optical coincidence retrieval system.
- Stressed the advantages and specified the use of microfiche for documental support of the retrieval system.

In Phase II, December through January 1966, the following tasks were accomplished:

- Indexing and abstracting of additional documents.
- Major revision of the EMPIRE Thesaurus to include the structuring of generic, specific relationships and related concepts.
- Produced catalog cards, technical abstract listings, and a file inversion.
- Initiation of a pilot microfiching program.
- An acquisition program for direct receipt of DDC materials for processing.
- Specification and implementation of the PAB (peek-a-boo) storage and retrieval system.

1.3 Initiation of Phase III

The present phase, commencing February 1966, was directed toward continuing the tasks which were established under Phase II of this program. Specifically, these tasks included the following:

- Expanding the acquisitions program to include the selection of pertinent articles appearing in the open literature.
- The establishment of a production schedule to include accessioning, analysis, production and delivery of items on a timely basis.
- Initiation and exploration of a current awareness program.
- Updating, correction and further refinement of the EMPIRE Thesaurus.
- Indexing test and evaluation, including inter-indexer consistency, self-consistency, comparison

with indexing in other systems, synthetic queries; and user questionnaire development.

- Document entry and further implementation of the PAB retrieval system.
- Development of literature search procedures and on-site demonstrations and training for the PAB retrieval system.
- Continuation of the bibliographic, indexing and abstracting effort, including the production of catalog cards, coded indexing worksheets, and technical abstract listings.

Further, contacts were to be made with various government agencies and facilities containing operating information storage and retrieval systems in order to exploit all available resources and to explore compatibility with the EMPIRE System.

SECTION II

2.0 SYSTEM OPERATION

2.1 Introduction

The primary objective of Phase III was to evaluate the system as a test data base for future research and development and, at the same time, to broaden the system base for testing its operational capabilities to support engineering information requirements. The former was accomplished, the results appearing in Section V of this report in the form of recommendations for continued research and development. For the latter, the following sub-sections describe the continuing activities of the EMPIRE program including innovations or modifications, mechanisms and rationale. Procedural details will be found in Appendix D.

2.2 Acquisitioning

The basic sources of EMC, EMC-related, and weapons systems literature acquired during the overall development of the EMPIRE system have included: 1) selected internally and externally generated documents from the NWL collection; 2) direct procurement of DDC documents by HRD in accordance with lists of selected TAB items prepared by the technical director (initiated during Phase II), and 3) an extensive file of applications, submitted to the tri-service Joint Frequency Panel (known as JF/12 documents), provided to the contractor by the technical director.

Phase III has seen the continuation of acquiring and processing literature drawn from the files of the Naval

Weapons Laboratory and the Defense Document Center. In addition, a program for extending the EMPIRE collection to include significant articles appearing in the open literature has been initiated. The following table summarizes the document mixture of the EMPIRE information base:

	PHASE				
	I	II	III	TOTAL	%
AD Reports, Digests and Misc. Tech. Reports	1104	735	527	2366	50.0
JF/12		871		871	18.4
ECM Reports			121	121	2.6
Open Literature			1371	1371	29.0
Totals	1104	1606	2019	4729	100.0

With regard to AD reports, the task of procurement went quite smoothly, once the FOIR was established. All items were requested, with exception of reports on limited distribution, by HRD on DDC's Form 1 -- hard copy and film copy -- in accordance with lists of selected TAB items prepared by the technical director. The film version of the documents was batched and then sent on to NWL; hard copy was remitted after processing.

On the basis of a suggested list of periodicals believed to be productive of non-trivial information relevant to EMPIRE, a budgeted item for acquisition of the open literature was included in Phase III. These consisted of professional technical and trade journals for the most part, supplemented by reprints, preprints, symposia proceedings,

standards and specifications, house organs, application notes for hardware components, instrumentation, and an occasional monograph of particular pertinence, e.g., "Radio Spectrum Utilization", a report of the Joint Technical Advisory Committee of the IEEE and EIA. Closely associated with the acquisitions process -- in fact, ancillary to it -- was attendance at special meetings or lecture series of local chapters of national professional societies where state-of-the art talks or review papers were delivered.

To live within budgeting limits and yet get wider coverage, resort was had to photoreproductions of selected articles from the more specialized journals where the yield of accessions per issue did not warrant purchase. Additionally, citations of current articles and conference meeting programs with abstracts affording a preview of addresses or papers submitted were utilized to procure complimentary copies or informal publications from the authors, if available.

It developed, in time, that the major emphasis of Phase III was placed on the open literature. Since there was no lack of material on subjects peripheral to RFI, which will ultimately assume greater importance, integrated circuits or developments in semiconductor devices, such as MOS field effect transistors, are examples -- results of a tentative nature regarding experimental devices or processes which were not easily subject to verification or duplication were generally excluded from the EMPIRE corpus. Similarly, analytical or theoretical papers, whose

simplifying assumptions or qualifications indicated limited value for design, troubleshooting or problem solving, were passed over in favor of treatments of demonstrated or more general utility.

Source or authorship was yet another criterion: it is fairly obvious that in developmental work on devices, a prospective vendor may make exaggerated claims. These normally would be excluded from journals whose contents are refereed. They have a way, however, of getting included in the variegated population of symposia or conference proceedings. Hence, this type of publication -- often the preliminary version of a formal journal article or a digest of a report -- was subjected to strict scrutiny for purposes of selection. On the other hand, items appearing in house organs such as "General Radio Experimenter" or the "Hewlett-Packard Journal", which describe improved measurement techniques and not just refinements in hardware were quite acceptable.

A list of journals for which subscriptions were placed is contained in Appendix A.

2.3 Bibliographic Description, Indexing and Abstracting

2.3.1 Cataloging

The practice of bibliographic description and subsequent catalog card preparation conforms generally with the standardization procedures devised during Phase II. Hence, descriptive cataloging returned as a clerical function, with final editing by an experienced cataloger to review the proper forms of entry for corporate author,

personal author, title and other elements of the bibliographic entry. Basic sets of seven catalog cards were issued for each document in the report literature: 1) corporate author(s), 2) title, 3) personal author(s), 4) AD number (or report number), 5) contract or task number, 6) NWL security number, and 7) accession number sequence. Where the NWL security number was non-existent, as in the case of direct acquisitions from DDC, a set was provided, filed by accession number, for eventual assignation of a NWL number.

Bibliographic description of the open literature did not present a major problem, although the style of entry changed somewhat. For the open literature, the title was the first entry, followed, in enclosed parentheses, by the journal or periodical nomenclature, volume number, periodical number, article page numbers (inclusive) and date. The remainder of the bibliographic description proved routine. Basic sets of 5 catalog cards were issued for each article in the open literature: 1) title, 2) periodical or journal, 3) personal author(s), 4) and 5) accession number sequence.

A special series of classified documents, treating electronic countermeasures and counter-countermeasures, was processed during the year. In order to retain the unclassified nature of EMPIRE catalog cards, titles were suppressed when classified, and replaced by the designation "Classified Title" in parentheses. However, the bibliographic information in the technical abstract listing included all proper titles due to the overall classified designation of the abstract listing.

2.3.2 Indexing and Abstracting

The guidelines for indexing and abstracting remain unchanged from Phases I and II, with continued intensive analysis of selected materials for conceptual information and parametric data. These are characterized by an informative abstract, i.e., specifics rather than generalized statements concerning a documents content, and indexing terminology which identifies the concepts conveying the significance of each document. However, while indexing and abstracting methodology remains relatively unchanged, the diversified nature and format of the several types of literature examined has had an overall effect on the indexing and abstracting effort.

Reports, generally, are of a uniform format, enabling the abstractor to seek in specific portions of the document for specific items of information in preparation of an abstract. Articles from the open literature, however, vary considerably in format and style, are by nature concise, and possess a high incidence of theoretical, analytical and mathematical treatments accompanying the recording of basic research. These factors require more effort by the abstractor to scan and evaluate the document and they generally demand greater technical ability and comprehension.

The abstracts accompanying the professional journals -- IEEE Transactions, Proceedings of the IEE, Journal of Applied Physics are examples -- are usually adaptable to EMPIRE abstract requirements; trade journals

do not possess this feature, necessitating complete preparation of an abstract.

The actual indexing of an article from the published literature does not differ from the indexing of a report.

Consideration was given to the inclusion of personal author affiliations, where cited in the open literature. A file of this nature would be useful, e.g., in determining where a particular endeavor is being undertaken. To provide for this provision, the corporate affiliation was noted on the indexing worksheets.

Following the indexing task, the PAB codes are entered on the indexing worksheet, utilizing the EMPIRE Thesaurus. During this Phase, this function was accomplished by clerical personnel, preliminary to an edit by indexer/abstractors who review coding, examine synonymy, and consider new terminology.

Indexing worksheets and technical abstract listings were prepared for documents 1839 through 2000, 5001 through 5365, 7000 through 8370 and 9001 through 9121 in separate classified and unclassified issues for greater accessibility. In addition, the first 1257 documents, processed in Phase I, were re-indexed to conform with present terminology and subsequently coded for document entry. Since all documents had been delivered to NWL, resort was had to original index worksheets and the abstract listing. Further, the first 250 abstracts were originally produced as single-page entries; these have been re-typed and reproduced conforming with current practice.

2.3.3 Thesaurus Revision

An examination of Volumes I, II, and III of the revised EMPIRE Thesaurus indicated several minor errors which were corrected during the current Phase. These included duplicate PAB codes, misfiled terms, and additional synonyms. Furthermore, new terms and codes were generated from coding of the first 1257 documents, and the indexing of new accessions. These revisions and additions were reflected in the monthly addenda.

Concentration in the open literature, which assumes a larger proportion of basic, theoretical, and analytical content than the report literature, has affected the nature of new terminology. Typical of this emphasis are the named effects, analytics and techniques applicable to developments and discoveries in the state-of-the art, e.g., solid-state physics and associated phenomena (see Appendix B). These developments are evidenced by totally new terms or concepts. Standard reference tools do not include them due to their novelty or an application which is not commonplace. They may be new or highly sophisticated techniques. As such, they require intensive documentation with respect to usage and technical application. With continued assignment and increasing occurrence of newer terminology, references to the occurrences can be collected and definitions formulated. This will have a continued impact on the indexing task and thesaurus construction efforts, and prove valuable as an information source within itself.

2.4 Current Awareness

An announcement bulletin is an effective device for

bridging the gap between issues of the abstract listing, providing technical personnel with awareness of new accessions containing significant information. Several bulletins were prepared during the year, differing in format and method of production. In the preliminary issue, produced as a by-product of catalog card publication, spacing was provided for distribution and routing information.

Bibliographic information was organized as an alphabetic listing under DDC subject categories. A subsequent issue was divided into broad subject categories pertinent to EMC interest. Manual production procedures largely destroy the economies of efficiently producing the bulletin on a regular basis. Future issues, based on the capture of data for EDP would be produced automatically.

2.5 PAB Implementation

The final technical report, Phase II, described the structure, input, search philosophy and output of the EMPIRE PAB card retrieval system. However, successful operation and evaluation of the optical coincidence scheme of storage and retrieval awaited complete entry into the decks of all indexed items. Therefore, the initial task in Phase III was to encode and drill these documents, including approximately 1257 technical reports and 871 JF/12 applications.

The task of converting the index terms to PAB codes for these documents required a review of indexed terminology. A particular problem existed for the first 250 documents due to incompatibility of the index terms to the revised Thesaurus. In some cases, the documents had to be re-indexed by utilizing

the abstract. Following encoding, the drilling function was assumed by utilizing clerical personnel. The procedure envolved pulling the cards corresponding to the PAB codes on the index worksheet, setting the machine to the proper accession number and drilling, returning the cards to the deck in random order, and continuing sequentially. The task required speed to meet an imposed deadline, yet a high degree of accuracy. Spot checks were made to monitor the effort.

In order to further determine the accuracy of the drilling effort an analysis was made of the corporate author dedicated decades by using the catalog cards as reference. It was determined that a number of corporate authors had not been entered, due to the lack of quality control and to the fact they were new and had not been coded. It was also determined that there existed a discrepancy between the Primary and Secondary cards. These items have now been entered into the system or rectified.

In summary, document accession numbers 1 through 2000, 9001 through 9121 and the 871 JF/12 applications, from the 2001 through 5000 series, are now encoded and drilled. Training in the actual use of the retrieval system, the formulation of search prescriptions and search techniques are discussed in subsections 2.7 and 2.7.1 of this report.

2.6 System Compatibility

During Phase II, a program to investigate other governmental and commercial indexing and abstracting services, which conceivably could supplement EMPIRE resources, was initiated. This study included both field trips and the

critical evaluation of published reports, exploring compatibility in terms of coverage, analytical apparatus, publication and distribution policies and procedures, and hardware and software programs. This effort was continued during the current phase; for example, the IR literature collection now numbers well over 550 select reports on information systems methodology and practices which have been reviewed and evaluated. Other efforts are discussed in the following subsections.

2.6.1 Field Trips

Field trips were conducted at the following facilities: Applied Physics Laboratory, NASA - Documentation, Inc. Facility, Electromagnetic Compatibility Analysis Center, David Taylor Model Basin (NARDIS), Army Library (ASDIRS), and the Naval Ordnance Laboratory. Results of the visits were chronicled in the monthly status reports. Indexing services for DDC were assessed for compatibility with EMPIRE; results are shown in Section IV of this report.

2.6.2 EMPIRE/ECAC Compatibility

As part of a continuing effort to broaden the data base of the EMPIRE system, the possibility of establishing mutually beneficial efforts between the Electromagnetic Compatibility Analysis Center (ECAC) and EMPIRE systems was proposed. The goal of this effort was to: 1) coordinate and integrate future ECAC and EMPIRE processing materials, thereby improving the services available from each, and 2) incorporate past ECAC indexed material that was determined to be of historical value into the EMPIRE date base. To

accomplish this goal, three tasks were presented, the first of which was an evaluation of the feasibility of combining work from the two systems, explored during the current year.

ECAC indexing/abstracting worksheets corresponding to a list of fifty AD documents which were processed into the EMPIRE system were requested. These were necessary to compare and evaluate the indexing depth and quality and the information content of the abstracts for the two systems. Further, a copy of the ECAC KWIC printout of all indexed documents with corresponding abstract listings was requested to determine the historical and technical value of material entered into the ECAC system. A partial receipt of the indexing/abstracting worksheets were compared; the test and the results are discussed in Section IV.

A number of on-site demonstrations were held at ECAC and the Naval Weapons Laboratory. Particular interest was expressed by ECAC personnel in the EMPIRE JF/12 file, specifically the treatment of parametric data and the retrieval capabilities of the PAB system. During the demonstrations, compatibility between the two systems was discussed. In summary, the main area of concern is thesauri terminology differences. A program to analyze and compare the two thesauri resolving differences and ambiguities, must precede any further effort in exploring overall system compatibility.

2.6.3 Project LEX

The Naval Weapons Laboratory requested HRD attendance at Session V of Project LEX, for the pur-

pose of assisting in the preparation of a DOD Technical Thesaurus. Session V of Project LEX was concerned with terminology in the subject areas of navigation, communications, detection and countermeasures, areas of primary interest to the EMPIRE program. Following briefings by Heston Heald, Director of Project LEX, and Eugene Wall, the project manager, participants were assigned to the panel responsible for developing portions of the projected DOD thesaurus concerned with navigation, communication, detection, and countermeasures.

The modus operandi was to critique a computer printout on cards representing main terms, which displayed all relationships including synonyms, scope notes if any, narrower and broader terms (hierarchical or class relationships) and related terms. Panel members were free to accept, reject, alter or supplement, scope, etc.

To assist the panel in the task, there were additional computer printouts showing all terms and their stratifications, frequency counts, and the sources, i.e., thesauri (about 250) from which derived; a good selection of standard technical dictionaries, in addition to the Webster unabridged; reference works, such as the multi-volume Thewlis Physics; and any monographs or texbooks brought by panel members. There were panel leaders permanently assigned to Project LEX who were efficient recorders, being thoroughly familiar with the ground rules. The services of Eugene Wall and Terry Gillum of DDC were at the panel's disposal for advice and guidance.

The scheme for qualifying and associating terms was quite orderly and generally consistent: analyze and judge usage; resolve ambiguities and homographs with scope notes or definitions; determine synonymy; arrange terms in hierarchical or family relationships and decide whether they had a truly reciprocal generic-specific relationship; then relate them among themselves on the basis of having some bearing on the term with which they were being linked.

2.7 Training

Demonstrations were held periodically during the year at NWL for the benefit of operating personnel. Basic principles of coordinate retrieval were reviewed, search techniques discussed and thesaurus format and look-up described, and the application of the abstract listing was explained. Synthetic questions, based on related abstracts, were used to illustrate handling of a query, formulation of search prescription and strategy (including narrow-to-broad and broad-to-narrow patterns), and actual manipulation of term PAB cards and readout equipment for successful retrieval. Document entry into the PAB system was explained, and actual input was accomplished by utilizing NWL clerical personnel in collaboration with HRD personnel.

2.7.1 Search Procedures

To assist the user in utilizing the EMPIRE optical coincidence retrieval system and simplify search routines, a literature search request form was devised (See Appendix C). The form specifies step-by-step procedures for conducting various types of searches, including specific or general

search patterns and the use of logical sums or negations. Further, definitive instructions for search preparations were included.

2.8 Procedures

Procedures relating to acquisitioning, bibliographic description, indexing and abstracting were revised slightly with respect to Phase II. Appendix D contains instructions and guidelines relating to these operations.

SECTION III

3.0 STATISTICAL SUMMARY

3.1 Introduction

The assessment of time required to complete the various acquisition and processing functions during Phase III has permitted the projection of a valid production schedule. In return, this schedule was utilized to measure actual performance. The following subsection discusses the statistical assessment of total document time, followed by a statistical summary of cumulative document totals to date for the various activities during Phases I, II, and III.

3.2 Document Processing

A valid breakdown of the time required to totally integrate a document into the system was made. Time assessed in this study included all functions necessary for document processing, from initial observation of the document (scanning for relevancy) to total incorporation. Results from this study, averaged over the corpus tested, are summarized in the following:

Event	Average Time Per Document in Minutes
Document Selection	5
Bibliographic Description	10
Bibliographic Editing & Correction	2
Indexing/Abstracting	45
Abstract Editing & Proofing	10
Abstract Typing & Correction	10

Index Worksheet Coding & Editing	8
Total Processing Time	90

These figures are not total production figures; thesaurus changes, terminology checks, and synonym problems continue to present difficulties. Further, abstract strip and paste-up, catalog card filing, document control, and overall reproduction figures are not included.

Average time per document includes both professional and clerical operations. Professional time is equal to approximately two-thirds of the total time.

STATISTICS

Acquisitions	Phase I & II	Phase III
1. Requisition	2982	1901
2. Dupe check	2982	1901
3. Logging	2934	1795
4. Labeling	2934	1795
5. Delivery	2710	2019
Bibliographic Description - Cataloging		
1. Master & index EAM worksheet	2710	2019
2. Production of cards for document	2710	2019
3. Assembly into 7 sets	2710	2019
4. Alphabetic/numeric sorting	2710	2019
5. Delivery	2710	2019
Indexing		
1. Analysis	2710	3275 *
2. Coding	1454	3275
3. Duplication	2710	2019
4. Delivery	2710	2019
Abstracting		
1. Composition	2710	2019
2. Editing	2710	2269 **
3. Typing	2710	2269
4. Proofing & correction	2710	2269
5. Strip & paste-up	2710	2269
6. Reproduction (collating) and binding	2710	2269

	Phase I & II	Phase III
7. Delivery ----------	2710	2269
PAB Entry		
1. Produce term cards ----------	3000	677
2. Check code against term card -	3000	677
3. Drill ----------	1454	1528
4. Delivery ----------		4000
Thesaurus		
1. From indexing sheets develop new descriptors & codes -----	3414	691
2. Keypunch & list ----------	3414	691
3. Edit for cumulative or monthly addenda ----------	3414	691

* Includes re-analysis of D 0001 - D 1256

** Includes D 0001 - D 0250

SECTION IV

4.0 TEST AND EVALUATION

4.1 Introduction

Research was conducted in four related areas, all using descriptors that had been assigned to documents and journal articles entered into the EMPIRE System. The first test was an investigation into the consistency of indexing by individual indexers. The second test determined the consistency of indexing performed for documents, each of which had been indexed by two indexers. The third test was an evaluation of the types of indexing accomplished by the EMPIRE staff in comparison with the indexing of ECAC and DDC. The fourth test involved retrieval of documents by manipulation of the PAB cards at the NWL.

4.2 Indexer Self-consistency

Twenty-five articles in the open literature previously indexed by EMPIRE indexers were reindexed by the same indexers to see how their second indexing correlated with the first. Beginning with D 7500, every tenth article was selected, for a total of twenty-five items.

An assignment of a descriptor in the first indexing was indicated on a tally sheet, and search was made in the second indexing for a match, for the same term or one approaching it in meaning, indication being made of the second occurrence or lack of it. Next, a record was made of the number of terms introduced in the second indexing that had not been included in the first. "Housekeeping"

descriptors, e.g., article, charts, graphs, etc., were excluded from these tallies for several reasons, mainly because they are not significant for indexing evaluation.

The actual percentage mean for the sample used in this indexer self-consistency (intra-indexing consistency) study was calculated from the matches. The hypothetical number of matching terms predicted (expected mean) for each article was developed by application of the sample mean. A graph was drawn showing the percentage variation from the expected mean for each individual article included in the test (See Figure 4-1). A plus percentage correlation and a minus percentage correlation showed that more and fewer terms respectively, were matched than had been anticipated on the basis of the percentage mean. The agreement sought was based on a comparison of the number of matches per article with the average of the number of descriptors assigned in the original and the test indexing. The range from -15.8 to -33.3% for the self-consistency test can be justified in part by the appearance of marginally valid terms selected from the article in question. The test involved more terms than the original for almost every case. Examination revealed that more terms had been drawn from the body of the article, which were generally less indicative of significant context. The standard deviation of the intra-indexing test was 15.8% around the population mean value (%) of 72.4%. It was demonstrable that the indexing of articles whose subject matter was peripheral, depended more on the background of the indexer than was the case for

INTER-INDEXER CONSISTENCY

VARIATION FROM A HYPOTHETICAL MEAN FOR DESCRIPTOR MATCHING

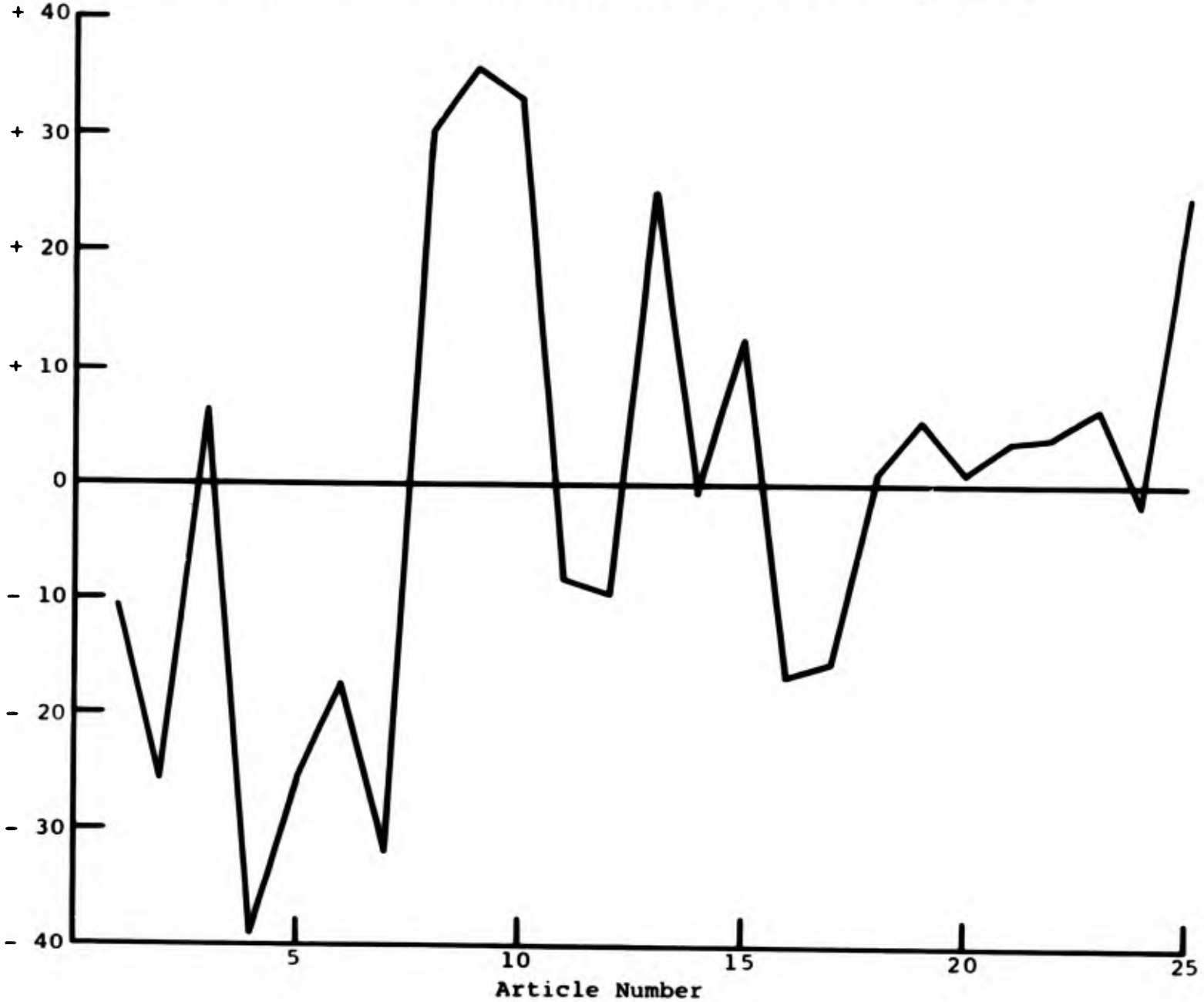


Figure 4-2

central topics, where the indexer relied on the authority of the Thesaurus for term assignment. Critical descriptors consistently were commonly assigned and the matches were preponderantly found among the first terms indexed. In the case of the peripheral concepts, the descriptors were found to be chosen chiefly from the body of the article subsequent to the introductory material and tended to be more specific, less representative of the article's scope, and, of necessity, less frequently selected in both tests.

4.3 Inter-indexer Consistency

Twenty-five articles were again selected on the basis of every tenth article indexed, beginning with D 7500, to ascertain the consistency of indexing done by a second indexer with that done by an earlier indexer. Again, the "housekeeping" descriptors were eliminated from the test. The tallying technique used for the self-consistency test was used again. A match percentage mean for the test sample used in the between-indexer consistency (inter-indexing consistency) study was again obtained from the matches found between the original and repeat descriptor assignments. The hypothetical number of matching terms predicted for each article was again determined by application of sample mean. A graph was made showing the percentage variation from the sample mean. (See Figure 4.2) A plus percentage agreement and a minus percentage agreement showed a positive and negative correlation, respectively, based on the mean percentage (expected value). The agreement sought was in

the comparison of the number of matches per article with the average of the number of descriptors assigned in the original and the test. The range from +19.75 to +40.00% for the inter-indexing test most likely reflected the appearance of articles, whose contents were equally familiar to both indexers and contained key terms thoroughly covered in the Thesaurus. The second indexing of the article again consistently brought in more terms than the original indexing. The conclusion inferred again is that more terms were selected from the body of the article, were often less generally relevant, and, consequently, did not contribute noticeably to the correlation sought in the number of matching terms. The standard deviation of the inter-indexing test was 19.75% around the sample mean value (%) of 58.8%. The lower correlation of the inter-indexing test is explicable by considering the factors of differing backgrounds of the indexers, and the consequent difference in emphasis in the assignment of descriptors. The most critical descriptors, however, again were commonly assigned and the most frequent matches were found among the first terms indexed. The unmatched terms were evidently selected primarily from the more diffuse body of the article. As a result, the terms again tended to be more specific than warranted.

The results of the inter-indexing consistency test are not as favorable as one would wish. One explanation is that one indexer may have confined his term assignments to the descriptors listed in the EMPIRE Thesaurus, constraining

them to fit those descriptors. The other indexer was more free in term assignments. The degree of consistency is regarded as acceptable, the hope being that, with more experience and more explicit scope notes in the Thesaurus, the indexers would attain more consistency.

4.4 Comparison of Indexing of Other Systems

A test was performed to determine the validity of the EMPIRE depth indexing effort as compared with the efforts of other systems. The test consisted of a term-by-term comparison of indexed descriptors of randomly chosen documents from EMPIRE, DDC, and ECAC. The population of the test was approximately one percent of the EMPIRE data base at the time of testing. The test was divided into three general areas: (1) total average number of descriptors per document for each of the three systems; (2) common assignment of terms between the three systems; and (3) valid descriptors that one system had which the other systems did not include (The choice as to which descriptors were valid or not, was a subjective judgment dependent on the tester's technical ability in EMC and his experience with information and retrieval techniques). The general criterion used by the tester was the specific technical applicability of a descriptor to a document, with an abstract being used to represent the document.

In the first area under test, ECAC assigned 17.0 descriptors per document, EMPIRE assigned 32.9, and DDC assigned 15.0 (The DDC identifier file was not available

for inclusion in the testing). The results indicate that a user of the EMPIRE System can ask specific questions and narrow the search pattern in order to attain retrieval of more pertinent documents.

In the second area under test, determination was made of the assignment of descriptors by one system that were common to, or duplicates of, the descriptors indexed by each of the other systems. DDC and ECAC had 5.7 terms that had been asigned in common. EMPIRE had common assignments of 7.2 with ECAC and 7.1 with DDC.

The third area investigated the pertinent descriptors missed by one system in comparison with the assignments of descriptors by the other two systems. The data obtained are shown in Figure 4-3.

FIGURE 4-3

PERTINENT DESCRIPTORS MISSED BY	COMPARED TO		
	EMPIRE	ECAC	DDC
EMPIRE	-	1.8	1.5
ECAC	4.0	-	2.7
DDC	3.0	1.6	-

The DDC and EMPIRE systems had the least number of relevant terms missing per document. Their indexing terms are generally more representative of the document which they describe than ECAC descriptors are. For this particular aspect of the testing, the greater specificity of the EMPIRE System was purposely neglected by ignoring the narrow terms.

It can be concluded from these tests that the EMPIRE descriptors are of excellent utility in an EMC information retrieval environment.

4.5 Synthetic Queries

A test was devised by which the probability for retrieval of a known document, from a set of randomly chosen documents, could be determined. The test population from DDC documents and J/F 12's was approximately one percent of the entire data base. The test consisted of a series of technical questions which were formulated from the abstracts of the documents and were searched on the EMPIRE retrieval system.

The evaluation of the results give evidence of 1) the effect of inconsistencies introduced into a retrieval system and 2) the probability of successful retrieval. There were, on the average, 5.2 descriptor terms for every question asked, or 245 descriptors in all. Of the total number of descriptors employed, 72, or 29.4%, were found to be unretrievable for the document from which the terms were originated. This result seemed high, so a further analysis was conducted. Exactly half of the terms missed (36) were trivial housekeeping descriptors which did not hamper retrieval or imperil conciseness. This left thirty-six terms, 14.7%, unaccounted for, which could be useful in identifying a particular document.

There are at least three operational areas in which relevant terms will not be retrievable. These areas are: 1) the indexing of the document, 2) the coding of the

worksheet, and 3) the drilling of the PAB cards. Examination of the original worksheets indicated that of the 36 terms missed, 14.5 missed due to lack of indexing, 9.5 were missed due to coding inconsistency, and 12 were missed due to incompleteness in drilling. To correlate these results with the retrievability of a specific document is, to a certain degree, a subjective judgement based on the user.

The manner in which the search and retrieval is implemented has an effect on the retrievability of a document. To use the system best, the PAB cards must be placed in order of importance, i.e., the least important PAB cards being placed on the reader last so that they can be removed first when a retrieval is not indicated. The determination of the importance of the order for the cards is a subjective judgement which is peculiar to each individual. The amount of personal fluctuation was not directly determined because only one person implemented the actual retrieval operation, but a mean deviation was computed that reasonably accounts for such fluctuations. The probability that any one document could be identified by use of all the descriptors obtained from the test questions was 28.3%. By removing the least important cards first, the document retrieval probability rapidly rose from 28.3% to 82.5% as shown by Figure 4-4. In region A of the Figure, the slope of the curve decreases markedly. This area is the transition between the irrelevant and relevant descriptors missed in the assignment of descriptors for the search operation. The deflection and the subsequent curve indicates the

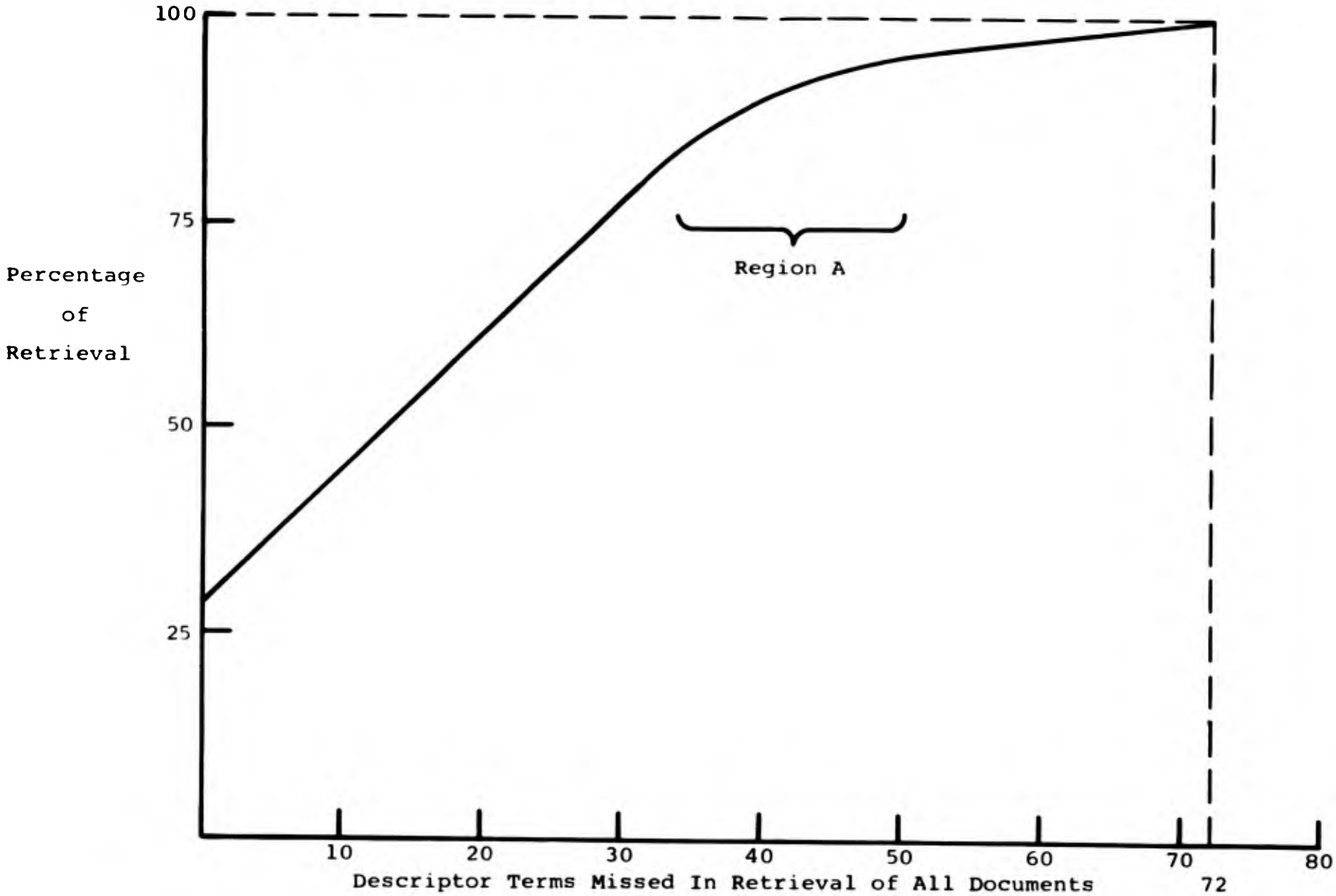

100
75
50
25
Percentage
of
Retrieval
Region A
10
20
30
40
50
60
70
80
72
Descriptor Terms Missed In Retrieval of All Documents

Figure 4-4

removal from the search of relevant terms decreases the rate of probability for retrieval. The area of transition takes place between 81 and 91% and has a mean of 86%, which agrees to within 0.7% of the 85.3% implied earlier in this section (subsection 4.5, second paragraph). A mean deviation of $\pm$ 5% can be attributed to human fluctuations in the ordering of PAB cards and the generalization of the search. It is to be noted that the search is not broadened between the 28.3% and the 86%. The search must be assumed to be broadened between 86% and 100%, at which point retrieval of a valid document is guaranteed.

The data and analysis indicate that except for drilling inaccuracy, the retrieval probability is within acceptable bounds.

4.6 User Questionnaires

Questionnaires have been developed to elicit user response as to the operation of the EMPIRE System. The data derived from this survey will be evaluated to determine how the system is operating, thereby permitting rectification of deficiencies and improvement of overall services. The questionnaires, developed for the retrieval system and the abstract listing, are shown in Appendix E.

SECTION V

5.0 RECOMMENDATIONS

5.1 Introduction

The EMPIRE data base now incorporates a fairly representative cross-section of the literature on electromagnetic compatibility and related topics. It includes report literature, classified and unclassified; the open literature of journal articles and symposia; and data-oriented files, such as field reports and J/F applications. It would appear to be appropriate at this stage to consider (1) how its coverage might be extended to make it more comprehensive within budgetary limitations; and (2) how it can be put to use as an informational resource for technical staff and management. The options suggested below are not exhaustive but suggestive of directions for development.

5.2 Input/Output Developments

To augment the retrieval capabilities of the PAB file, capture of the bibliographic and indexing data on magnetic tape should be undertaken at least on a pilot scale. The input would be manipulated by the Control Data INFOL program which requires a minimum of formatting for efficient output of a variety of useful products. Among these are: 1) an enhanced retrieval capability due to the additional logic formulations the program accommodates; 2) permuted arrangements of bibliographic data, such as KWIC; 3) announcement bulletins; 4) special bibliographies,

manipulation of the data for testing, and printout of data for review.

5.2.1 Acquisitions

To achieve more comprehensive coverage of the literature of electromagnetic compatibility in the open literature, as its scope is presently defined, is hardly feasible unless the budget for acquisitions is liberalized. On the other hand, exhaustive coverage within a narrower specialty would permit stretching of allotted funds to include a broader range of materials. If, for example, the scope of collection were limited to antennas, it would be possible to include significant monographs - even textbooks - and conference papers or pre-prints available to attendees, not to mention the abundant commercial literature. It is suggested that in the upcoming effort a fairly narrow technical topic be selected for exhaustive open literature coverage. It is desirable that the subject matter, despite its restricted scope, should be of central importance to the mission of NWL.

5.2.2 Automatic Indexing

If the text or bibliographic descriptions and the abstracts or significant extracts of an article or report are read into the computer, a small experiment based on statistical measures and/or associations might be attempted to indicate the feasibility of automatic indexing. It would not be an operation on natural language text pure and simple, as these have been demon-

strated to be far too costly and impractical. The text would be pre-edited for essential data elements capture. Normalization through matching against thesaurus entries would also be involved in the process.

5.2.3 Selective Dissemination of Information

A pilot program, employing the INFOL program, could be initiated to match current accessions against profiles to be compiled for a handful of key managerial and technical personnel. The "hits" or matches would be listed and distributed to the participants, who, on the basis of the abstracts, could signify their interest in receiving the complete text. It is suggested that this program be limited to the unclassified literature to avoid unnecessary complications.

5.3 Management Information

A technical information system does not operate in a vacuum. It is supposed to respond to needs tied to specific programs. These programs are determined by management, based on the criterion of whether they promote the organization's objectives in fulfillment of its assigned missions. But missions are adjusted to meet changing situations, and their objectives have to be redefined. To make these decisions, management must be aware of technological developments. Some of this information is derived from official reports and interaction with technical staff. Intramural or departmental intelligence must be supplemented by related work in

progress in outside circles.

5.3.1 Project Reports, 1498's

Recognition of the foregoing has led to reporting of RDT & E efforts on DOD-wide scale. The Defense Documentation Center has been assigned the task of collecting DD Form 1498's, which report the substance of research and development projects, and acting as a referral center for relevant details. It should be one of the functions of EMPIRE to seek information contained in pertinent DD 1498's for the twin purposes of uncovering new sources of information and publications for its collection, as well as calling to the attention of management similar work - perhaps duplicative - conducted by other organizations and facilities. On the basis of the information obtained by follow-up, it is often possible to decide that current projects may have been rendered technologically obsolete by other developments, or to supplement results to date with findings disclosed by related projects elsewhere.

5.3.2 Data Package Production

It is true that a single treatment or presentation of subject matter cannot satisfy a varied clientele. This is the reason why information services frequently employ intermediaries or mechanisms which attempt to give individual services. In NWL, there are three or four major groupings of personnel who depend upon EMPIRE to some degree: 1) technical management; 2) scientists and engineers; 3) supporting technical groups; and 4) admin-

istrative services. If the services were custom-tailored to the constituent groupings of the clientele, greater acceptance, hence wider utilization, of EMPIRE's resources would result. However, this could prove quite expensive from the point of view of cost-effectiveness, and would demand a closer liaison between EMPIRE and its clientele. It might be feasible, at the expense of some extra effort by EMPIRE, to modularize its products such that those items appropriate to management would be detached for management; those portions included for senior technical personnel would be delivered to them; and so on. For example, it might be advantageous to report to management a breakthrough in integrated circuits instrumentation or technology which enables the various laboratories to deal with these components or circuits economically and reliably. Technical staff would be interested both in the theory and details of operation. Technicians would be interested in procedures and laboratory processes. Administrative personnel would be concerned with evaluation and specification and standardization for monitoring and procurement purposes.

5.4 Significant Information Identification

Characteristic of the pace of technology is the rapid appreciation and assimilation of new developments. Walter Brattain states that the significance of the transistor effect, announced by Bardeen and himself, was immediately picked up by the technical community and put

to work. There is very little delay between the announcement of a discovery and the development of literature about it.

5.4.1 Terminology Utilization

One means of identifying the new developments is through the language used in describing them. These may be wholly new terms or unfamiliar combinations of terms whose precise import has to be inferred from the context and the citations. As the terms are considered for inclusion in the thesaurus, novelty may be due either to expanded coverage or to a new development in the state-of-the-art. In the former case, the usual reference tools will define the term adequately, and inclusion in the thesaurus is determined without much difficulty. In the latter case, one must await confirmation by the author's peers, which usually is not long delayed. In either case, the term is documented, whether by accepted authorities or from the context and supporting literature. These should be brought to the attention of the clientele in the form of concise explanations compiled into a specialized glossary.

5.5 Training Tools

While the principal purpose of an information system is current and retrospective awareness, regard for the more personal needs of the clientele should not be overlooked. Specialization often shuts out a technologist from knowledge of developments on the periphery of his

field, or from mastery of disciplines which have a bearing on his design problems. To bring him up-to-date on these developments, at least two useful services can be made available: Selective Dissemination of Information (SDI) and special bibliographies.

5.5.1 SDI as a Training Tool

A pilot scheme of SDI is suggested for application to key technical personnel at various levels of responsibility. After ascertaining their information requirements within the constraints of EMPIRE, a profile of interests would be formulated in terms of EMPIRE descriptors. With each delivery of abstract bulletins, individual notices would go out to the participants, directing their attention to those specific items conforming to their interest profile. A reading of the biblio-abstracts should enable them to select and request those items most likely to be of profit to them, either in connection with work in progress or to bolster their knowledge of techniques and processes related to their responsibilities. Both profiles and matching schemes would benefit greatly from appropriate feedback.

5.5.2 Special Bibliographies

When lasers were introduced, a great many applications were predicted for them. However, it wasn't until certain power levels, whether pulsed or CW, became possible and modulation of the beams was feasible, that they became of interest to microwave propagation and reception schemes. Similarly, holography at present appears to have a great deal of potential for use in photogrammetry and optical in-

strumentation for the study of high-speed phenomena. The compilation of special bibliographies on topics such as these would appear to be desirable for the purpose of alerting management and technical personnel to developments likely to have great impact on their activities --- if not immediately, then in a not very remote future. It goes without saying that standard procedures borrowed from other disciplines could also be the basis of such special bibliographies, not to mention demand bibliographies reflecting the state-of-the-art of new or ongoing projects.

5.6 System Improvements

5.6.1 Thesaurus Automation

Depends on machine construction.

When a thesaurus exceeds a certain size, manual update or revision becomes a cumbersome operation. The present thesaurus could profit from computerization in several respects. Indexing terms could be compared against the present list, both as they occur and in permuted form to judge whether they are new, in which case they would be printed out for decision on how to treat them. Generic and specific terms would be automatically cross-referenced. It would also be possible to determine whether closeness of association between related terms afforded a basis for the construction of useful descriptor hierarchies.

5.6.2 ECAC and Other Compatible Sources

It is highly desirable to be aware of resources which could supplement the data base of the EMPIRE System. It is particularly advantageous to have a degree of compati-

bility which would enable easy transfer of information between the available systems operating in the same or closely related fields and disciplines, perhaps even missions. An effort should be made to seek out the groups known to fit this description for the purpose of making cooperative arrangements resulting in an augmentation of local resources and a reduction of duplicative activities. ECAC is an obvious candidate.

5.6.3 Compatibility with National Systems

There are national systems such as DDC, NASA and special technical information centers with whose products and services it would be useful to have a reasonable degree of compatibility. The joint EJC-DOD program under COSATI auspices entitled Project LEX was an effort to achieve compatibility at least on a macrothesaurus level. When the new thesaurus is available, a study should be undertaken to determine whether a common basic vocabulary couldn't be achieved between it and EMPIRE.

5.7 Theoretical Improvements

5.7.1 Man and Information Problem

In evaluating the effectiveness of an information system, some factors may be present which are difficult to isolate. One of these is the capacity of man to absorb information. It is the experience of most scientists and engineers that the only manner in which they can stay on top of all the information in a specialty is to keep narrowing its scope. Otherwise they must content themselves with

a good working knowledge of the subject supported by an acquaintance with sources of information to be tapped on an ad hoc basis. Many solutions have been proposed for the situation where the problems confronting the technologist are interdisciplinary, but they founder on the notion that knowledge is ultimately a personal affair. In the following sections some tentative approaches are suggested.

5.7.2 Information Formatting

In Section 5.3.2 there was a suggestion of producing information packages in modular fashion to take account of the variations in need and technical sophistication of the clientele. Here it is only necessary to add that close working relationship is a prerequisite to formatting information for ready assimilation by the clientele. For example, if it is ascertained that a technician lacks the necessary analytical apparatus for a design or measurement problem, two courses can be followed: (1) providing the necessary study materials for the appropriate theoretical concept, e.g., Fourier waveforms analysis; or (2) presenting the pertinent material at a less sophisticated level, e.g., approximate formulas, nomographs or tables.

5.7.3 Value of Information

Objective measures of the value of information have, except in rare instances, eluded the students of information systems. It can be demonstrated, for example, that a patent search prevented an unwise investment in a research and development program. "Re-inventing the wheel"

is not easily demonstrable in other instances of R&D. Since the information is frequently useful only under certain circumstances and at a specific time, it may be that a "duplication" of effort is as, or more, effective than a search whose results are in doubt. A further question is what happens with information stored up against the day that it might prove useful. This latter question really goes to the heart of most information systems which attempt to anticipate requirements for information. As suggested previously, the only way to establish the utility of information is through feedback from the user which is obtained through a close working relationship between information system service personnel and the clientele.

5.7.4 Time Effect on Information

Two things are fairly well-established about information in the sense of well documented, sound contributions: (1) information is constantly in flux; (2) the rate of obsolescence is not quite as great as it is popularly presumed to be. The problem, then, is keep a close watch on developments in order to determine: (1) significant contributions; (2) current status. A citation index can accomplish the first type of recognition. A review of state-of-the-art report can evaluate the currency or relevance of previous accessions to present requirements. Both these avenues should be pursued.

APPENDIX A

Title	No. of Articles	Percent
AIAA Student Journal	1	.1
American Documentation	3	.2
Aviation Week	1	.1
Avionics	1	.1
The Bell System Technical Journal	42	3.0
Computer Design	5	.4
Control Engineering	16	1.2
Cryogenic Engineering News	7	.5
Datamation	2	.2
Data Processing Magazine	1	.1
Design News	3	.2
The Electronic Engineer	1	.1
Electronic Engineering Design	18	1.3
EEE Circuit Design Engineering	12	.9
Electromagnetic Design	2	.1
Electromechanical Design	16	1.2
Electronic Design	38	2.8
Electronic Industries	18	1.3
Electronic Packaging and Production	9	.7
Electronic Products	6	.4
Electronics	43	3.1
Electro Technology	36	2.6
Evaluation Engineering	2	.1
Frequency	6	.4
The General Radio Experimenter	9	.7
Geo Marine Technology	9	.7
Hewlett-Packard Journal	4	.3
Hydraulics & Pneumatics	1	.1
IBM Journal	18	1.3
IEEE Journal of Quantum Electronics	9	.7
IEEE Journal of Solid-State Circuits	5	.4
IEEE Spectrum	24	1.8
IEEE Transactions on Aerospace and Electronic Systems	48	3.5
IEEE Transactions on Antennas and Propagation	74	5.4
IEEE Transactions on Audio and Electroacoustics	7	.5
IEEE Transactions on Automatic Control	8	.6
IEEE Transactions on Broadcast and Television Receivers	6	.4
IEEE Transactions on Broadcasting	4	.3
IEEE Transactions on Circuit Theory	6	.4
IEEE Transactions on Communication Technology	38	2.8
IEEE Transactions on Education	5	.4
IEEE Transactions on Electromagnetic Compatibility	25	1.8

Title	No. of Articles	Percent
IEEE Transactions on Electron Devices	25	1.8
IEEE Transactions on Engineering Management	5	.4
IEEE Transactions on Engineering Writing and Speech	3	.2
IEEE Transactions on Geoscience Electronics	3	.2
IEEE Transactions on Human Factors in Electronics	2	.1
IEEE Transactions on Industrial Electronics and Control Instrumentation	3	.2
IEEE Transactions on Industry and General Applications	7	.5
IEEE Transactions on Instrumentation and Measurement	16	1.1
IEEE Transactions on Magnetics	19	1.4
IEEE Transactions on Microwave Theory and Techniques	23	1.7
IEEE Transactions on Power Apparatus and Systems	12	.9
IEEE Transactions on Reliability	10	.7
IEEE Transactions on Sonics and Ultrasonics	10	.7
IEEE Transactions on Systems Science and Cybernetics	5	.4
Information Storage & Retrieval	2	.1
Instruments and Control Systems	3	.2
International Science & Technology	1	.1
Journal of the Acoustical Society of America	46	3.4
Journal of Air Traffic Control	5	.4
Journal of Applied Physics	49	3.6
Journal of the Association for Computing Machinery	12	.9
Journal of the Audio Engineering Society	4	.3
Journal of the Franklin Institute	26	1.9
Journal of Geophysical Research	1	.1
Journal of the Optical Society of America	15	1.1
Journal of Research of the National Bureau of Standards	1	.1
Journal of the S.M.P.T.E.	5	.4
Machine Design	6	.4
Microwave Journal	44	3.2
Microwaves	19	1.4
Missiles and Rockets	1	.1
Ordnance	1	.1
Phillips Technical Review	7	.5

Title		No. of Articles	Percent
Photogrammetric Engineering		9	.6
Physical Review		2	.1
Physics of Fluids		5	.4
Proceedings of the IEE		91	6.6
Proceedings of the IEEE		115	8.4
Proceedings of the National Electronics Conference		58	4.2
Product Engineering		3	.2
RCA Review		5	.4
The Review of Scientific Instruments		39	2.8
Science		3	.2
SIAM Journal on Applied Mathematics		16	1.1
Space/Aeronautics		17	1.2
Space Technology		2	.1
Technology Week		4	.3
Technometrics		11	.8
Telemetry		4	.3
Test Engineering		2	.1
UnderSea Technology		5	.4
Total	93	1,371	100.0%

APPENDIX B

FUNC	DESCRIPTOR/DEFINITION	PAB
	017 SUPPORT EQUIPMENT	
017	AN/USQ-17 COMPUTER SET	4S91
	054 TECHNICAL JOURNALS + TRADE PUBLICATIONS	
054	AMERICAN DOCUMENTATION (JOURNAL)	0533
054	IBM JOURNAL OF RESEARCH + DEVELOPMENT (JOURNAL)	0537
054	JOURNAL OF GEOPHYSICAL RESEARCH (JOURNAL)	0532
054	JOURNAL OF THE ASSOCIATION FOR COMPUTING MACHINERY	0534
054	PHOTOGRAMMETRIC ENGINEERING (JOURNAL)	0535
054	TECHNOMETRICS (JOURNAL)	0536
	060 ENGINEERING AND EMC TERMS	
060	AMMONIA	4S88
060	ATOM, ATOMIC	4S87
060	BOLTZMANN TRANSPORT EQUATIONS	4S84
060	COVARIANCE	4S8J
060	DELTA-FUNCTION, DIRAC DELTA-FUNCTION	4S82
060	DENSITOMETRY, DENSITOMETER	4S81
060	DIFFERENTIATION (MATHEMATICS)	4S9J
060	DIRAC DELTA-FUNCTION, DELTA FUNCTION	4S82
060	EIGENFUNCTION, EIGENVALUE	4S7J
060	ELECTROSTRICTION	4S8J
060	ENTHALPY	4S9G
060	ERGODICS	4G64
060	FERMI LEVEL	4S9J
060	FREQUENCY STANDARD	4S8G
060	FUEL CELL	4S8S
060	GEOLOGICAL STRUCTURE, GEOLOGY	4N9J
060	GEOMAGNETISM, TERRESTRIAL MAGNETISM	4G6G
060	GREEN'S FUNCTION	4GH7
060	HILBERT TRANSFORMS	4P81
060	MAGNETOMETER	3W8J
060	MAGNETO-OPTICAL EFFECT	3W7S
060	NUMERICAL ANALYSIS	4S8J
060	TERRESTRIAL MAGNETISM, GEOMAGNETISM	4G6J

APPENDIX C

LITERATURE SEARCH REQUEST

(Instructions on Reverse)

1. LSR No.		3. Name		5. Department	
2. Date		4. Title		6. Tel. No.	

7. Information Needed for (Project, Task, Work Area, etc.)

8. STATE Question in Narrative form

9. Analyze Question into Subjects

Concepts	Rank	Search Terms	Pab Code(s)

10. Limitations

a. Coverage (Time frame)
   All _____
   Past Year _____
   Past 3 Years _____
   Other _____

b. Security
   All _____
   Exclude:
   Secret _____
   Confid _____
   Open Literature _____

c. Response Time
   ____ 1. ASAP
   ____ 2. 24 hours
   ____ 3. Complete search

11. Output Needs
   1. List of Documents _____
   2. List of Abstracts _____
   3. Actual Documents _____
      1. Hard copy _____
      2. Microfilm _____

SEARCH PROCEDURE

1. Where the request is for a specific document, known by AD and/or auxiliary number, corporate author, personal author, exact title, or project number, the catalog card file is the preferred searching medium.
2. The Termatrex optical coincidence system is most effective for logical products, i.e., the co-occurrence of two or more concepts within a document. Only those documents containing <u>all</u> of the selected terms will be retrieved.
3. For the most efficient search, select the term cards for the most specific concept first. Then proceed to select the next most specific concept, and so forth. This frequently will correspond to the ranking order specified in Item 8 on reverse. Merely viewing a single term card for a highly specific concept may be sufficient to answer the search request without the superimposition of other term cards.
4. Having selected the term cards in a particular order, place them on the viewer. With the aid of the slide ruler, ascertain and note the document numbers.
5. If no documents are retrieved, it may be necessary to broaden the search. Removing cards successively will generally yield more documents, some of which may be responsive to the question.
6. If a poor response is recorded, despite the broadening of the question, consult the Thesaurus for broad, narrow, and related terms. Employ these terms either in place of or combination with the original terms.
7. For logical sums (all or any of the selected terms) place each term card successively on the viewer, recording the individual document numbers. Match these sets of numbers, with one another to eliminate duplicate document numbers.
8. For logical differences or negations, proceed as you would for logical products, recording all documents disclosed by the viewer. Then place the excluded terms on the viewer and eliminate those documents matching the retrieved.
9. For cancelled, withdrawn or obsolete documents, place the green transparency over the selected term cards. The documents shown by the clear light are no longer available.
10. Having selected the documents, check the Bibliographic Abstract Listing to determine whether the selected documents are relevant or pertinent to the request.

INSTRUCTION FOR PREPARATION

Item 1: Serial number assigned in order of receipt.

Item 2-6: Self-explanatory

Item 7: If no specific project, task, work unit, etc., state purpose for which desired.

Item 8: a) Formulate question as clearly as possible. Provide context, e.g., objectives of program or study for which desired. Indicate whether selective or broad search is desired.

b) <u>Selective Search</u> - Polarization of camouflage target cross-sections limited to stationary targets.

<u>Broad Search</u> - Target cross-sections.

Item 9: a) <u>Concepts</u> - List words, terms or phrases which define the content of the question (e.g., antennas, polarization, camouflage, absorbers reflectivity, target cross-section, signal processing, aspect angle, detection, profiles, clutter, directivity and ground return). If your request is limited to major vessels, weapon systems, project names, sub-systems, AN equipments, etc., consult function codes 1-53 for specific names.

b) <u>Rank</u> - Indicate in descending order (one is highest) the weight or relative importance of the individual topics.

c) <u>Search Terms</u> - These are terms found in the Thesaurus. Using the Thesaurus, match as closely as possible the concepts in 8a - adding terms pertinent to the request.

Item 10: Designate the time coverage, security limitations, and desired response time.

Item 11: Self-explanatory

APPENDIX D

EMPIRE SYSTEMS AND PROCEDURES

1.0 Introduction

The EMPIRE System continues to expand coverage through an intensified effort to select and maintain sources of EMC and EMC-related documentation. Currently, the primary source of EMC material is the open literature, including pertinent professional and technical journals; house organs and trade publications; proceedings of conferences, meetings, symposia, and workshops of professional groups, societies, and associations; and reviews, annuals, monographs, and state-of-the-art summaries. Additional techniques utilized in obtaining pertinent open literature information include the following:

1) scanning of announcements from publishers, professional and trade associations, and government agencies.

2) scanning of abstract reviews, bulletins, and announcements.

3) scanning of reviews appearing in journals and other publications.

Additional sources of EMC material are the files of the Naval Weapons Laboratory, including technical reports, manuals, newsletters, Joint Frequency Allocations, progress or status reports, and technical memoranda, and direct acquisition of technical reports and publications from the Defense Documentation Center. A list of selected items is provided by NWL after a review of the Technical Abstract Bulletin. These items are then requisitioned directly from

DDC. The following paragraphs present detailed procedures for document acquisition and processing.

2.0 EMPIRE Procedures

2.1 Procedures for Acquisition and Control of Report Literature

a) Establish FOIR (Field-of-Interest Register).

b) Transcribe list(s) from NWL of selected items to DDC Form 1.

c) Submit requisition for hard copy documents and microfilm to DDC.

d) A tickler file is maintained for all documents ordered. As the documents are received, they are so noted in the file. The file is checked at four week intervals for documents not yet received.

e) The available microfilm is forwarded to NWL as soon as a sufficient quantity has accumulated. In no case, however, are they retained by HRD beyond two weeks after receipt.

f) Hard copy documents are checked against the catalog card file for duplication.

g) The documents are divided into classified and unclassified groupings.

h) The documents are analyzed for relevancy and format. Symposia and periodicals are assigned sequential numbers for individual papers beginning with an overall document number.

i) The remaining documents are assigned sequential accession numbers, after arrangement in AD number sequence.

j) The documents are labeled with the assigned accession numbers and entered on the EMPIRE Document Control Log Sheet.

Documents received directly from NWL are sent via NWL courier in batches from 100 to 250 documents. A transmittal listing provides the document identification numbers and titles. Upon receipt, the documents are checked against the transmittal list and a receipted copy is returned to NWL. The respective documents are then processed as described above.

2.2 Procedures for Acquisition and Control of Open Literature

a) purchase subscriptions to pertinent journals, symposia, and abstract reviews

b) scan all journals and additional materials for references to other significant information; acquire by loan, reproduction or purchase.

c) analyze journals and symposia for relevant articles and notate.

d) check selected articles against catalog card file for duplication.

e) assign sequential accession numbers, label and enter on the EMPIRE Document Control Log Sheet.

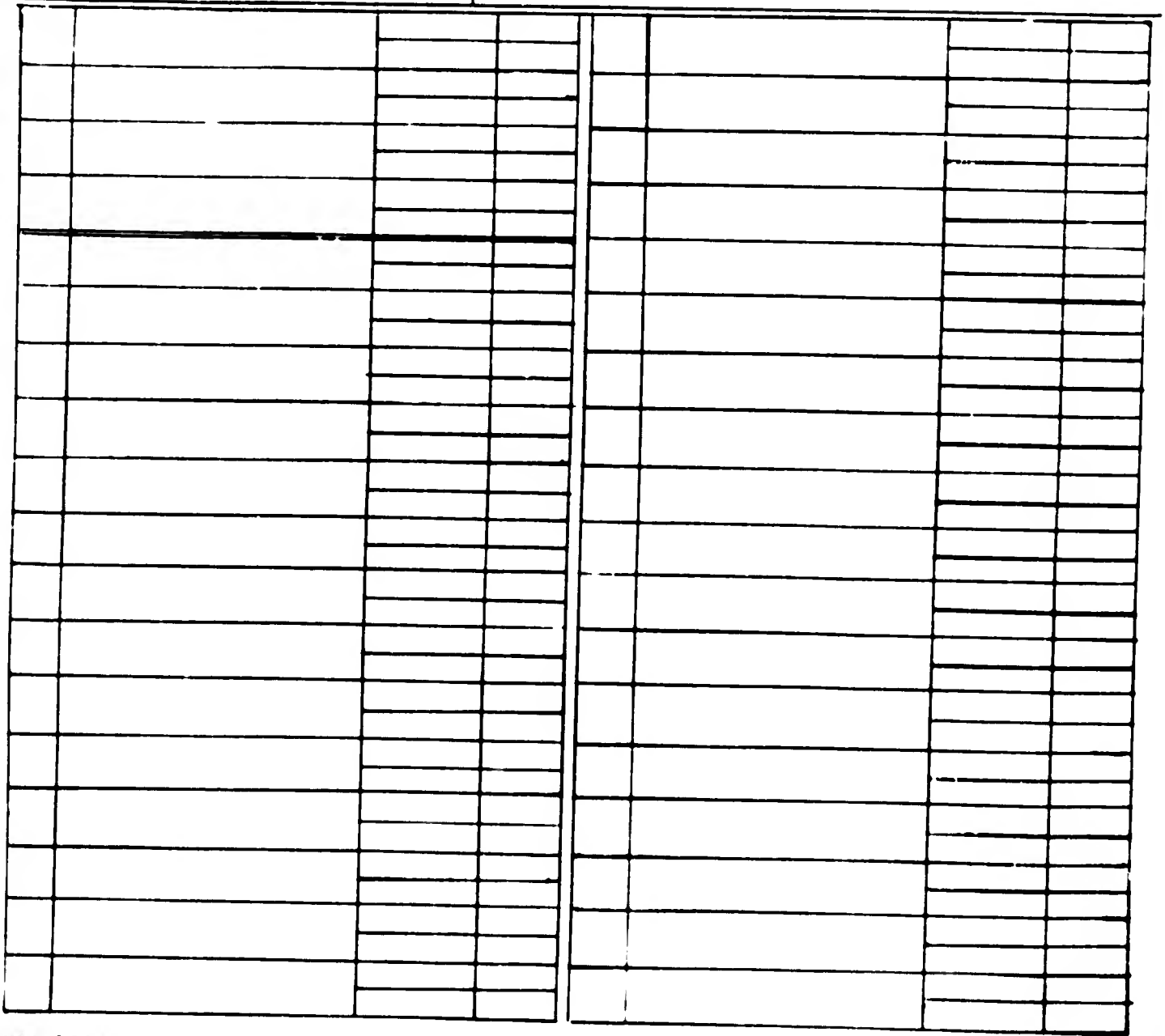
2.3 Bibliographic Description - Cataloging Procedures

During Phase I of the EMPIRE System, a new format was designed to provide single entry capability for producing catalog cards, index worksheets, and the bibliographic-abstract listing. The same basic format was utilized during Phase II and Phase III. The form is shown in Figure E-1.

(Card Classification)
1-3 Empire No. (NWL Acc. No.)

1-45 Source
2
3-9 TITLE (Pages)
4 (Descriptive Notes)
5
6-9 Author 6-39 Date
6-45 Class. 6-51 No. 6-63 NWL No.
7-9 Aux. Nos.
7-45 Contract No.

Codes: Cards 8
9,15,21,27,33,39,45,51,57,63 69,75



The descriptive cataloging information is entered in a three-by-five-inch square at the upper right-hand side of the form. EAM key-punching instructions indicating the card and column position of the respective information is preprinted in the upper left-hand corner of the form. Below this information, space is provided for entering two columns of descriptive terms and their respective PAB card entry codes.

The document work set consists of a plain bond original sheet and an EAM indexing worksheet. The bibliographic information is entered on this work set. The original sheet is used to produce the catalog cards; later, the document abstract is typed directly below the bibliographic information and the sheet is then utilized to produce the bibliographic abstract listing.

The bibliographic information is typed directly from the source document. Each document is examined to provide the following information:

Report Literature	Open Literature
EMPIRE Accession Number	EMPIRE Accession Number
Corporate Author(s)	Article Title
Document Title	Source Publication
Personal Author(s)	Volume Number
Date	Periodical Number
Paging	Article Page Numbers
Contract Numbers	Date
NWL Accession Number	Personal Author(s)
NWL Security Number	
AD Number or Report Number	
Document Security Classification	

After editing, catalog cards are produced from the three-by-five typed area on the plain bond sheet. Four sheets are arranged on a 914 Xerox photocopy unit to produce an offset master, from which catalog cards are run off. Sufficient copies of each card are produced to provide seven basic sets for NWL, and another set is retained by HRD for duplicate checking.

The cards for NWL are sorted into the following card sets:

Report Literature	Open Literature
Title	Title
Corporate Author(s)	Source Publication
Personal Author(s)	Personal Author(s)
AD Number	Accession Number
Contract Number	
NWL Accession Number	
NWL Security Number	

The basic set may be augmented by additional cards for joint corporate or personal authors. The set may consist of fewer cards for a variety of reasons, including the nature of the document such as Joint Frequency Applications which utilize only three catalog cards, or where particular information such as contract number or personal author are not pertinent to the document.

2.4 Indexing Procedures

Indexing procedures generally conform to those presented in the Appendix of the EMPIRE System Final Report

published in January of 1964. General housekeeping concepts, such as document types, corporate authors, primary or secondary EMC classification, and principal subject continue to be a part of document analysis. Special presentations such as graphs and tables continue to be noted.

Indexing seeks to identify the concepts which convey the significance of each document. Initially the title, abstract, and table of contents are examined for relevant items of information. Then, the report narrative is scanned to locate all additional significant items, including vehicles, equipments, materials, and components discussed in the document, along with methods and definitions presented, and pertinent parametric EMC data. The stated results of the document, including problems, conclusions and recommendations require particular scrutiny.

When miscellaneous minor components and subsystems are mentioned in the document, but are not central to its purpose, they are indexed by generic descriptions and coded up to the first level of AN designations. If the components and subsystems are important to the document's intent, they are indexed and coded at their specific level.

The EMPIRE Thesaurus standardizes all index terms and concepts. However, the indexing terminology in not limited to the existing Thesaurus. Descriptive terms which do not appear in the Thesaurus are noted on 3" x 5" index cards along with their usage. These new terms are reviewed and analyzed; if approved, new PAB codes are generated, and coded on the index worksheet.

Following the indexing task, the PAB codes for each indexed term are entered on the index worksheet. The EMPIRE Thesaurus is utilized for this function. Presently, the coding function is accomplished by clerical personnel followed by an edit by indexer/abstractors who validate coding and examine new terminology.

2.5 Abstracting Procedures

Abstracting is a concomitant function of indexing. The author's abstract may be adapted, if competent. If the author's abstract is inadequate, a new abstract is prepared.

There are two major types of abstracts, each intended to perform a certain function. The first type, an "informative" abstract, is a condensation of the original document. The second type, an "indicative" abstract, describes what the document is about. The informative abstract is currently being used for the EMPIRE System.

The function of the informative abstract is to obviate as much as possible the reading of the original document. Ideally, this type of abstract provides all the important ideas that the document contains, and presents quantitative and qualitative data. Excess wordage and detailed explanations are omitted. As a means of conveying information, it is often an adequate substitute for the original document.

Abstracts are written in conjunction with indexing because the indexing worksheet provides a compact outline of the document's subject matter. After a number of abstracts have been drafted, they are reviewed for accuracy

of technical content and edited for style and syntax. The abstracts are then typed on the plain bond original directly below the 3" x 5" inch area containing the bibliographic information. The abstract is proofread and corrected. The biblio-abstracts are then dummied up, two columns to an oversized matrix, which is photo-reduced to produce 8½" x 11" offset masters for the abstract bulletin.

The bibliographic abstract listing is in ascending accession number sequence. The classified abstracts are printed and bound separately from the unclassified abstracts and marked accordingly.

APPENDIX E

INTRODUCTION TO THE INFORMATION RETRIEVAL AND ABSTRACT LISTING QUESTIONNAIRES

Questionnaires have been devised to oversee the operation of the EMPIRE Information Retrieval System and the Abstract Listing Bulletin. By determining and obviating the weaknesses of these services, better services can be provided in the future. The ultimate goal is to provide you with the best possible services at the lowest cost. Your cooperation in completing the questionnaires is needed to aid us in achieving this goal.

The Information Retrieval questionnaire should be completed at various time periods. Question numbers 1-5 should be completed at the time the search is conducted; question number 6 should be completed after you have read the abstracts. The remaining questions should be completed after you have read the documents. (If you did not request any documents, complete the remaining questions after you have read the abstracts).

The Abstract Listing questionnaire should be completed at two time periods. Question numbers 1-3 should be completed after you have read the abstracts; the remaining questions should be completed after the documents are read. (If you do not request any documents, complete all questions after you have read the abstracts.)

INFORMATION RETRIEVAL QUESTIONNAIRE

1. What is your job title? _____

2. Please state the question asked to the information retrieval system?

3. Would you list the terms used in the search?

4. Who conducted the search?

   _____ Personally conducted the search

   _____ Librarian

   _____ Secretary

   _____ Clerk

   _____ Technical assistant

   _____ Other (please state job title)

   _____

5. How many document accession numbers were identified from the results of the search? _____

6. Did you refer to the abstract listing before requesting the documents?

   _____ Yes _____ No

7. How many documents were requested from the abstract listing? _____

8. How many identified documents were relevant to the problem? _____

9. Was the information derived from the retrieved documents beneficial to you in any of these ways (you may check more than one way)?

_____ Completely answered the question

_____ Partially answered the question

_____ Background information

_____ Other (please explain) _____

_____

10. a) Are you satisfied with the performance of the information retrieval system?

_____ Yes _____ No

b) If No, please check the reason for your dissatisfaction (you may check more than one reason)

_____ System did not identify any documents

_____ Too few documents were identified

_____ Question was not answered

_____ Too many documents did not pertain to the problem

_____ Other (please explain) _____

_____

ABSTRACT LISTING QUESTIONNAIRE

1. What is your job title? _____

2. How many abstracts from the abstract listing were of interest to you? _____

3. How many documents were requested from the abstract listing? _____

4. In general, are the abstracts adequate as a substitute for the original documents?

   _____ Yes _____ No

5. Which abstract do you feel is the most important to you in your work (only indicate the abstract number)?

   D _____

6. Was the information contained in the abstract beneficial to you in any of these ways (you may check more than one way)?

   _____ Complete solution of a problem related to your work

   _____ Partial solution of a problem related to your work

   _____ Background information for present or future work

   _____ Stimulated you to think of new ideas

   _____ Verified a past solution to a problem

   _____ Other (please explain) _____

   _____

7. Would you have obtained the information contained in the abstract without the aid of the abstract listing?

   _____ Yes _____ No

BLANK PAGE

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*
Howard Research Division
Control Data Corporation
7735 Old Georgetown Road, Bethesda, Md. 20014

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE
EMPIRE Information System, Phase III
Final Technical Report

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

5. AUTHOR(S) *(First name, middle initial, last name)*

6. REPORT DATE
28 February 1967

7a. TOTAL NO. OF PAGES
70

7b. NO. OF REFS

8a. CONTRACT OR GRANT NO.
N-178-8883

b. PROJECT NO.

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
U.S. Naval Weapons Lab.
Dahlgren, Virginia

10. DISTRIBUTION STATEMENT

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT

The primary objectives of Phase III were to evaluate the system as a test data base for future research and development and to broaden the system base for testing its operational capabilities to support EMC information requirements. This report describes the continuing activities of the EMPIRE program, including innovations or modifications, mechanisms and rationale applicable to acquisitions, indexing/abstracting, thesaurus revision and retrieval implementation. Test and evaluation for intra- and inter-indexing consistency, indexing comparison with other systems, and the utilization of synthetic queries are discussed. Finally, recommendations for continued research and development are presented.

DD FORM 1 NOV 65 1473

UNCLASSIFIED

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
EMPIRE						
Information System						
Information Storage and Retrieval						
Electromagnetic Compatibility						
Optical Coincidence						
Thesaurus						
Indexing						
Abstracting						
System Evaluation						
Documentation						