RL-TR-91-169
In-House Report
June 1991

# INTELLIGENCE ANALYST ASSOCIATE (IAA): A PRELIMINARY ANALYSIS & SPECIFICATION

Carrie G. Pine



*APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.*

Rome Laboratory
Air Force Systems Command
Griffiss Air Force Base, NY 13441-5700

This report has been reviewed by the Rome Laboratory Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RL-TR-91-169 has been reviewed and is approved for publication.

APPROVED:

THADEUS J. DOMURAT, Chief
Signal Intelligence Division
Directorate of Intelligence and Reconnaissance

APPROVED:

WALTER J. SENUS
Technical Director
Directorate of Intelligence and Reconnaissance

FOR THE COMMANDER:

IGOR G. PLONISCH
Directorate of Plans and Programs

If your address has changed or if you wish to be removed from the Rome Laboratory mailing list, or if the addressee is no longer employed by your organization, please notify RL(IRAE) Griffiss AFB NY 13441-5700. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

# REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188



| 1. AGENCY USE ONLY (Leave Blank) | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | June 1991 | In-House Jun 90 - Feb 91 |

**4. TITLE AND SUBTITLE**
INTELLIGENCE ANALYST ASSOCIATE (IAA):
A PRELIMINARY ANALYSIS & SPECIFICATION

**5. FUNDING NUMBERS**
PE - 62702F
PR - 4594
TA - PR
WU - OJ

**6. AUTHOR(S)**
Carrie G. Pine

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Rome Laboratory (IRAE)
Griffiss AFB NY 13441-5700

**8. PERFORMING ORGANIZATION REPORT NUMBER**
RL-TR-91-169

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
Rome Laboratory (IRAE)
Griffiss AFB NY 13441-5700

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
N/A

**11. SUPPLEMENTARY NOTES**
Rome Laboratory Project Engineer: Carrie G. Pine/IRAE/(315) 330-3038

**12a. DISTRIBUTION/AVAILABILITY STATEMENT**
Approved for public release; distribution unlimited.

**12b. DISTRIBUTION CODE**

**13. ABSTRACT (Maximum 200 words)**

Intelligence analysts who process and analyze data from unformatted text are currently faced with a number of serious problems. These include an overwhelming volume of input text, and a lack of automated tools to help process and analyze the text. The Intelligence Analyst Associate (IAA) is the first effort in a long-term approach to solving the unformatted text problem. This paper presents a preliminary definition of the operational problem, and a preliminary specification of the requirements for solving it.

**14. SUBJECT TERMS**
Automatic Data Base Generation  Message Processing
Natural Language Understanding

**15. NUMBER OF PAGES**
52

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| UNCLASSIFIED | UNCLASSIFIED | UNCLASSIFIED | UL |

TABLE OF CONTENTS

FIGURES



# IAA DEFINTIONS, ACRONYMS, AND ABBREVIATONS

## ACRONYMS, ABBREVIATIONS

| | |
|---|---|
| ADBG | Automatic Data Base Generation |
| DB | Data Base |
| FTD | Foreign Technology Division |
| IAA | Intelligence Analyst Associate |
| info | information |
| intel | intelligence |
| NLU | Natural Language Understanding |
| req. | requirement |
| RL | Rome Laboratories (formerly RADC) |
| WPAFB | Wright-Patterson Air Force Base |

## DEFINITIONS

AUTOMATIC DATA BASE GENERATION (ADBG) - The function of creating data base records from unformatted text input (implementation-independent).

ANALYSIS (of intelligence data) - The process of reasoning about intelligence data, in order to determine its potential significance, and to gain useful information.

CONCEPT - A "concept" refers to a subject of interest, but at a higher conceptual level than a topic. A concept may incorporate several lower-level topics, and may imply certain relationships. E.g., the TOPIC "bombing" is only one subject which may fall under the CONCEPT "terrorist attack".

DATA - The essential raw information contained within a piece of text. E.g., who, what where, when, how many, what with, etc.

ESSENTIAL DATA - The basic, necessary data contained within a piece of text. E.g., who, what where, when, how many, what with, etc.

FREE-TEXT - Unformatted text.

INFORMATION - Meaningful/useful knowledge.

PRECISION - The ability of the system to find ONLY the information on the topic requested.

PROCESSING - The process of finding and extracting data on topics of interest from free-text, for analysis.

RECALL - The ability of a system to find ALL of the information on a requested topic

TOPIC - A subject of interest; a person, place, thing, or time.

# 1 OVERVIEW

## 1.1 THE OPERATIONAL PROBLEM & ITS SIGNIFICANCE

Intelligence Analysts who process and analyze data from unformatted text are currently faced with a number of serious problems. These problems include an overwhelming volume of input text, a lack of automated tools to help process and analyze the text, and the negative consequences which result.

Analysts are overburdened by the large volume of input text they should look through to find data of interest for analysis. They simply do not have enough time to read through all the text which might contain data relevant to their topic of interest. As a result, there is an increased likelihood that the analyst will miss potentially critical intelligence data. This is even more true during times of crisis, when the volume of incoming message traffic drastically increases.

Ironically, the time analysts do spend looking through text for important data reduces the time they actually have left to analyze the data. Less time for analyzing the data means less time to reason about its implications, and less time to formulate solid conclusions about what the data implies - the whole point behind finding the data in the first place.

Once data on a topic of interest is found, it is generally not put into any kind of a structured, computer-usable format. As a result, application software (SW) which could be developed to facilitate Analysis cannot be utilized. This includes graphic display utilities, which could improve analysts' ability to conceptualize the meaning and significance of the data they are dealing with. It also includes utilities which could be developed to support analytical modeling of the data.

Missed data, insufficient time to analyze data once it is found, and a lack of automated utilities to help process and analyze data from text, all negatively impact the analysts' ability to produce complete and accurate analyses. But the quality of the intelligence analyses are not the only things affected; so is the quality of everything which depends on them. This includes decisions affecting the design and acquisition of military weapon systems, and decisions directly affecting the Intelligence and Operations communities. The quality and effectiveness of these decisions are only as good as the information they are based upon.

## 1.2 THE PAPER: CONTENTS & PURPOSE

### 1.2.1 CONTENTS

This paper contains a Preliminary Analysis of the Operational Problem, and the Project Perspective on Solving the Operational Problem. The preliminary analysis is based on the text processing and analysis problem faced by Intelligence Analysts at the Foreign Technology Division (FTD) at Wright-Patterson AFB (WPAFB) in Dayton, Ohio.

### 1.2.2 PURPOSE OF PRELIMINARY ANALYSIS

The main purpose of the Analysis is to present a preliminary definition of the operational problem, and a preliminary specification of the requirements

necessary to solve it. The reasons for doing so are:

o to provide background information on the problem we are trying to solve to others who may be interested;

o to establish an initial problem definition which can be used as a starting point for iterative refinement. This "working-version" of the problem definition is a tool for obtaining constructive input and feedback on our current understanding of the operational problem. By incorporating corrective input and feedback into the problem definition, we are able to evolve a more accurate understanding and definition of the operational problem. The better we understand the problem, and the more accurately we define it, the more likely we are to develop a system to solve the real problem, and to do so at less cost over the long-run.

Our overall goal is to solve the real operational problem, over a series of efforts, in as cost-efficient and intelligent a manner as possible. This is a first step towards defining the Operational Problem, and what is required to solve it. It is to be used as a working problem definition, a means of improving our understanding of the problem, and what is required to solve it. As such, constructive input and feedback are both welcomed and encouraged. This paper is NOT to be interpreted as a concrete or final statement of the operational problem.

### 1.2.3 PURPOSE OF PROJECT PERSPECTIVE

The Project Perspective presents the results of an analysis, by the Government, on how best to approach solving the problem over the long-term. It is included to help people understand what the Government's approach is, and why this approach is being taken. Constructive feedback is welcomed on this, as well.

# 2 THE OPERATIONAL PROBLEM

## 2.1 SCOPE

### 2.1.1 GENERAL SCOPE

The Operational Problem deals with the Processing and Analysis of free-text for Intelligence applications (see Figure 1).

### 2.1.2 SPECIFIC SCOPE

#### 2.1.2.1 SCOPE OF THE "BIG" PROBLEM

The long-term goal is to be able to deal with a broad domain of input text, from whatever sources (and in whatever forms) the analyst needs to deal with. Depending on the particular analyst's function, it was determined that the text input could be anything from hardcopy books, newsletters and reports, to digital message traffic or digital data from external databases. However, each of these types of input imposes its own particular problems and requirements on the job of finding, accessing, and utilizing data from text. It became apparent that the scope of the near-term problem would mushroom to unwieldy proportions if all these types and sources of input were considered.

#### 2.1.2.2 SCOPE OF THIS EFFORT

In an attempt to scope the near-term problem down to something more feasible, the focus has been restricted to the processing and analysis of text which is already in digital form. The primary input of interest will be digital message traffic, and possibly some reports in digital form. The subject domain currently under consideration is that of "Perestroika Watch". "Perestroika Watch" is information having to do with changes in the Soviet Union associated with Perestroika, such as changes in doctrine, economic restructuring, openness, etc. If necessary, further scoping down of the near-term problem will be considered in order to assure a realistic near-term goal.

## 2.2 THE ESSENTIAL (FUNCTIONAL) PROCESSES FOR SOLVING THE OPERATIONAL PROBLEM

Any type of processing can be boiled down to its essential, implementation-independent functions. This is a useful means of describing a problem, for the very reason that it IS independent of any particular implementation. This allows the people who are solving the problem to use their expertise in determining the best implementation to meet the needs and requirements of the user.

Specifying the system in terms of implementation-independent functions also supports evolution of the system's functional components. As technology evolves, new implementations may be found that are more suitable for performing a particular function. When this occurs, the new implementation of the function can be interchanged for the old one, as long as all interfaces are well defined.

### 2.2.1 FREE-TEXT PROCESSING

An attempt was made to specify the high-level essential functions for Processing free-text. They are illustrated in Figure 2. They include:

o Finding potential text-sources of data (on topics of interest); (Another functional interpretation of this step might be reducing the input volume of free-text to a manageable volume for the rest of the processing steps);

o Determining the specific topics(s) of the text;

o Determining if the topic of the text is the topic of interest; and

o Extracting the essential data from the text, for Analysis.

### 2.2.2 INTELLIGENCE DATA ANALYSIS

Intelligence Data Analysis involves reasoning about data on topics of interest in order to obtain useful intelligence information. See Figure 2. This may include analyzing patterns and relationships in the data, reasoning about its significance, and identifying/deducing important information. The specific essential functions comprising Analysis have not yet been pinpointed.

THE CURRENT METHODS FOR SOLVING THE OPERATIONAL PROBLEM, & THE PROBLEMS IATED WITH THEM

FREE-TEXT PROCESSING

.1 THE CURRENT METHOD

ırrent method of implementing the essential functions for the Processing nalysis of free-text is illustrated in Figure 3. Keyword Profiles are to send the Analyst text which might contain data of interest. The t is to reduce the input volume of the text they must look through to a manageable size. The remaining functions -determining the topic of the ext, determining if the topic of the text is the topic of interest, and cting the essential data from text- are all done manually, by the ligence Analyst.

.2 PROBLEMS ASSOCIATED WITH THE CURRENT METHOD, & THEIR SIGNIFICANCE

OO MUCH TEXT FOR ANALYST TO PROCESS MANUALLY; TOO TIME-CONSUMING

are several problems currently associated with Processing free-text Figure 4). One of the biggest problems is the large volume of free-text sts must read through, in order to find the data they are interested in nalysis. The time they spend doing this severely cuts into the time have left to actually Analyze the data.

KEYWORD SEARCH; MISSED INFORMATION vs TOO MUCH TEXT FOR ANALYST TO SS

rd searches try to reduce the incoming volume of text the Analyst must at by only forwarding messages on topics of interest to the analyst. is done by finding and forwarding only those messages containing ords" preselected by the Analyst.

tunately, the goal of keyword searches is not well achieved due to a r of problems in how they are implemented. Broad, general keyword les result in too large a volume of text for the Analyst to handle, ting its whole purpose. And since text of ANY topic which matches the rds gets forwarded to the analyst, they end up with a large amount of licable text. The other option is to specify a narrow, very specific rd profile. The analyst may use this type of profile to try to reduce nput volume, and to try to limit the text he does receive to his topic nterest. However, in so doing the Analyst runs the risk of missing cal data on his topic of interest, just because it does not contain particular keywords.

INTELLIGENCE DATA ANALYSIS

.1 THE CURRENT METHOD

the Analysts have found the data they are interested in, they Analyze it y to glean useful information (see Figure 1). This may involve looking he data tc study patterns and relationships, reasoning about the

potential significance of the data, and trying to ID/deduce important information. Currently, the Analyst must devise his or her own methods to do this.

### 2.3.2.2 PROBLEMS ASSOCIATED WITH THE CURRENT METHOD, & THEIR SIGNIFICANCE

The current method of Analysis does not exploit technological opportunities which it could be taking advantage of. Graphics utilities, such as utilities to plot data on timelines or maps, could greatly facilitate the Analyst in studying data. It is much easier to conceptualize patterns and relationships in data when they are represented pictorially. And it would be much easier for the Analyst to reason about the significance of the data if they could visualize it in a more meaningful form. Unfortunately, since the data output from Processing is currently not put into a computer-usable form, graphics applications cannot be utilized for Intelligence Analysis. Neither can any other types of application SW which could support the Analysis of the data. Figure 4 overviews the problems with the current method of Analysis.

### 2.3.3 PROBLEM ANALYSIS OUTLINE

Appendix A is a Problem Analysis Outline. It summarizes the essential functions composing the operational problem, the current methods for implementing these functions, and the problems associated with them.

## 2.4 REQUIREMENTS FOR SOLVING THE OPERATIONAL PROBLEM

Appendix B contains an outline of the Requirements for solving the Operational Problem, based on our preliminary Analysis. These are the long-term Requirements we are striving to achieve over successive efforts. Again, this is an initial, working specification of the requirements. It is not definitive.

## 2.5 A POTENTIAL NEW METHOD FOR SOLVING THE OPERATIONAL PROBLEM

### 2.5.1 THE NEW METHOD

Figure 5 overviews a potential new method for solving the Operational Problem, and our objectives in suggesting this approach. The proposed system will contain an Automatic Data Base Generation (ADBG) capability to process the free-text. This will help take the manual message processing burden off of the Analyst, and will put the data in a structured format other SW applications can utilize. And since the data output from Processing would now be in a computer-usable form, graphics utilities could be developed to support Intelligence Data Analysis.

It should be noted that the term Automatic Data Base Generation is specifically used, as opposed to NLU, to stress the functional requirement which needs to be achieved without implying how that functional requirement should be implemented. In other words, we would like the contractor designing the system to determine what technology best meets this functional requirement, and how best to meet the performance requirements associated with it.

### 2.5.2 CHARACTERISTICS OF THE NEW METHOD & HOW THEY MEET THE OPERATIONAL REQUIREMENTS

Figure 6 illustrates how the essential functions for Processing and Analyzing text would be implemented in the new method. The Processing function would now be implemented via ADBG, and by Intel Data Processing Utilities, such as utilities to access data on topics of interest. These would replace the old keyword searches, and most of the manual work done by the Analyst. Analysis would now be supported by utilities to create graphical representations of the data, and utilities allowing the analyst to access and configure data interactively. Figures 7 and 8 show how the new method for solving the operational problem would meet our operational requirements.

# 3 PROJECT PERSPECTIVE (ON SOLVING THE LONG-TERM OPERATIONAL PROBLEM)

Rome Laboratory and FTD are currently initiating an effort, Intelligence Analyst Associate (IAA), which will address the text Analysis and Processing problem. The purpose of this section is to communicate the larger context that IAA is a part of, in our long-term effort to solve the Operational Problem. It also explains why the current effort's goals are important to us.

## 3.1 LONG-TERM (OVERALL) PROJECT PERSPECTIVE

### 3.1.1 THE NEED FOR A LONG-TERM APPROACH ("WHY" NEEDED)

a. ADBG TECHNOLOGY NOT YET MATURE - While the functional concept for achieving the Operational Requirements seems like a good approach, we do not think that the full concept can be achieved in the near-term. The biggest reason for this is the state-of-the-art of the technologies involved; in particular, ADBG technology. To the best of our knowledge, ADBG technology is still in the developmental phase; its full, hoped for, potential will not be achieved until some future time. As a result, systems such as IAA, which utilize ADBG in their design approach, are really counting on being able to capitalize on the FUTURE functional and performance characteristics ANTICIPATED from the technology. Near-term systems cannot expect to achieve ADBG's full potential at this time.

b. ADBG TECHNOLOGY STILL CHANGING RAPIDLY - In addition, the fact that the technology has not yet matured implies that it is still very dynamic, still changing. This is very true for ADBG technology. And since the ADBG problem has not yet been solved, several different approaches exist, and are still being developed, to try to solve the functional problem. The implication is that any system that "hard-wires" an existing ADBG capability into itself is inviting near-term obsolescence.

It became clear that a long-term approach to solving the Operational Problem was necessary in order to avoid an unplanned, and overly expensive, long-term approach. Such an approach could result in our developing a series of successive, technologically short-lived systems, at undue cost to the Government.

### 3.1.2 FACTORS & CRITERIA INFLUENCING THE LONG-TERM APPROACH

In planning the long-term design approach, there are various factors and criteria which have to be considered in order to achieve long-term success. These include cost criteria, such as cost-efficiency in solving the Operational Problem over the long-term, and technological criteria, such as making sure that our long-term approach minimizes any risk of failure. Other technological factors which have to be considered include anticipating growth/evolution of the technologies involved; anticipating expansion of system subcomponents and the system itself; anticipating iterative change/evolution in the Operational Requirements, and system design specifications; and anticipating the fact that several different contractors may be involved in the effort over time. As can be seen, these criteria strongly influence our requirements for a long-term design approach.

### 3.1.3 LONG-TERM APPROACH REQUIREMENTS ("WHAT" IS NEEDED)

Given the important factors influencing IAA's long-term approach, the high-level requirements for successfully achieving our long-term goals were defined.

#### a. COST-EFFICIENCY OVER THE LONG-TERM: MINIMIZE UNPRODUCTIVE REDUNDANCY BETWEEN EFFORTS/PROMOTE REUSABILITY

Since we anticipate the need for several successive efforts to achieve our long-term Operational Requirements, and since we need to be cost-efficient in solving the Operational Problem over the long-term, it seems apparent that we should avoid unnecessary duplication of work from one effort to the next. In other words, we want to avoid making expenditures which do not productively bring us closer to solving the Operational Problem. The requirement which comes out of this is to promote planned-for reusability from one effort to the next. This includes promoting the reuse of knowledge from one effort to the next (reusable requirements, functional specs, etc.), and promoting the reuse of as much of the system as possible/practicable from effort to effort. Why redevelop portions of the system over and over again, which could be stabilized now (e.g., the User Interface)? This will also allow us to focus money and effort more productively in future efforts, by freeing us to concentrate more on advancing specific component technologies, such as ADBG.

#### b. GROWTH REQUIREMENTS: STRESS FLEXIBILITY OF SYSTEM DESIGN

Other technological criteria include anticipating growth/evolution of the technologies involved, anticipating expansion of system subcomponents and the system itself, anticipating iterative change/evolution in our requirements and designs, and anticipating the fact that several different contractors may be involved in the effort over time. All of these factors stress the need for a near-term design specifically geared to be flexible, and specifically designed to facilitate such change.

#### c. MINIMIZE RISK IN LONG-TERM APPROACH: NEED FOR A TECHNOLOGICALLY SOUND & FEASIBLE LONG-TERM APPROACH

Another criteria is to minimize the risk involved in the long-term approach to solving our Operational Requirements. Two of the biggest risks of a long-term approach (and requirements FOR a long-term approach) are making sure that the approach is both technologically sound and feasible. We decided that the best way to minimize this risk is to address it head-on; develop a long-term approach which specifically identifies those factors which need to be addressed and resolved in the long-term design approach of the system, so that the long-term plan is both feasible, and technologically sound. For example, what design standards need to be specified for the system now, up front, to make sure that it is truly evolvable?

### 3.1.4 LONG-TERM SOLUTION APPROACH ("HOW" TO MEET APPROACH REQS.)

Given "what" the long-term project requirements are, the question then becomes "how" best to achieve them. This process is analogous to system design, in that the requirements defined at one level (in this case, the "Over-All Projects Level") are allocated to different subcomponents at the next lower level (in this case, individual projects at the Project level). Figure 9 illustrates this concept.

## 3.2 NEAR-TERM PROJECT PERSPECTIVE

### 3.2.1 REQUIREMENTS ALLOCATED TO THE NEAR-TERM PROJECT ("WHAT" THE REQUIREMENTS ARE)

As discussed in the previous section, the Near-Term Project inherited various requirements associated with the long-term plan for achieving our Operational Requirements. In addition, the Near-Term System has several Near-Term requirements which it must meet, independent of any long-term considerations. These Near-Term Requirements include obtaining a Near-Term System capable of demonstrating end-to-end system functionality (the system concept), and having the system installed in the Operational Environment at FTD. Appendix C specifies the Near-Term System Requirements.

### 3.2.2 NEAR-TERM PROJECT TASKS

Given a long-term design approach, the Near-Term considerations for solving the Operational Problem become twofold:

1) Long-term Design Considerations - How best to approach our long-term goals, knowing that it may take several successive efforts; and

2) Near-Term Design Considerations - How best to meet our Near-Term Requirements, such as reducing the manual message processing burden now carried by the Analysts, given that a full-blown solution may be several years down the road?

In order to satisfy the requirements which need to met in the Near-Term Project, two Near-Term Tasks were defined; the Near-Term System Development Task, and a Long-Term Planning Task. Figure 10 illustrates the allocation of the Near-Term Project Requirements to these tasks.

NEAR-TERM PROJECT TASK: SYSTEM DEVELOPMENT ("HOW" we'll meet the reqs.)

IAA has two major design objectives, designed to satisfy the two types of design requirements it must meet; both the Long-Term System Requirements and the Near-Term System Requirements.

a. DESIGN OBJECTIVE 1 (TO MEET LONG-TERM SYSTEM REQUIREMENTS):

The system's long-term design requirements include cost-efficiency over the long-term; flexibility of system design; minimizing unproductive redundancy between efforts; and promoting reusability between efforts. Based on these requirements, it was determined that the Near-Term System should specifically be designed to be an evolvable system framework; a foundational system that is designed to evolve over successive efforts.

b. DESIGN OBJECTIVE 2 (TO MEET NEAR-TERM SYSTEM REQUIREMENTS):

In order to satisfy our Near-Term requirements, our second design objective is to use the system framework to develop our first prototype system. This system will have end-to-end functionality (full breadth), but will not have the full-depth of functionality or input scope anticipated for the long-term system. In terms of Performance Requirements, the main goal of the Near-Term system is to reduce the Intelligence Analyst's manual free-text processing/analysis burden, in terms of the time and effort they spend (see Fig. 11).

NEAR-TERM PROJECT TASK: LONG-TERM PLAN ("HOW" we'll meet the reqs.)

The second Near-Term Project Task was defined in order to meet our Long-Term System Planning requirements. Since the people most qualified to do long-term planning for the system are the people with the most expertise in the system's technological areas, we decided that it would be a good idea to actually make this part of the Near-Term effort. The criteria for the Long-Term Plan are specified in Appendix D. It will be up to the Contractor to identify the specific issues which need to be addressed in the long-term plan.

# 4 CONCLUSION

The free-text Processing and Analysis problem affecting our Intelligence Analysts has serious consequences, and its impact is widespread. Analysts need help manipulating text, finding and accessing text on topics of interest, and putting the data from text into a structured form that will enable the development of application SW for Analysis.

A full solution to this problem is probably several years down the road; applicable technologies have not yet matured, and the overall problem is large in scope. However, near-term solutions could be designed which would ease the Analysts' burden in the interim. An evolutionary approach would enable us to help the analyst while we develop the long-term solution, rather than leaving them high and dry until a full-blown, final solution has been developed. If our true goal is to help the analyst, it is imperative that a well-designed, evolutionary approach be adopted.

APPENDIX A: PROBLEM ANALYSIS OUTLINE

DESCRIPTION: This Appendix describes the current implementation of the essential functions, & the problems associated with them

**********

1.0 ESSENTIAL FUNCTION: PROCESSING
- Find & extract specific data on Topics of Interest for Analysis

A. CURRENT IMPLEMENTATION: (see implementations of subfunctions)

B. PROBLEMS/UNEXPLOITED OPPORTUNITIES:
- o Largely unautomated
- o Results not in computer-usable form. So application software cannot be developed for Analysis, which needs the resultant data for input.
- o Too time-consuming; takes time from Analysis
- o Volume of input text is still overwhelming; causes Analysts to miss important intelligence data

1.1 ESSENTIAL FUNCTION: FIND POTENTIAL TEXT-SOURCES OF DATA (ON TOPICS OF INTEREST)
- Find & acquire potential sources of Data of Interest (Reduce volume of input text to a more manageable subset for subsequent processing steps)

A. CURRENT IMPLEMENTATION: KEYWORD PROFILE
Analyst establishes a profile containing "keywords". Any messages containing any of these keywords (exactly as spelled) are sent to the analyst.

B. PROBLEMS/UNEXPLOITED OPPORTUNITIES

1. KEYWORD PROFILE RESULTS IN AN UNDESIRABLE "CATCH 22" TRADEOFF, DEPENDING ON HOW THE PROFILE IS SPECIFIED:

- o BROAD (MORE GENERAL) KEYWORD PROFILE -
  (miss less important data) BUT
  (get higher resultant volume of text AND
  higher % unapplicable text)

vs

- o NARROW (MORE SPECIFIC) KEYWORD PROFILE -
  (get lower resultant volume of text AND
  lower % unapplicable text)
  BUT (miss more important data)

2. PROBLEMS ASSOCIATED WITH HIGH RESULTANT VOLUMES OF TEXT and HIGH % UNAPPLICABLE INFO:

- o LARGE VOLUMES OF TEXT OVERWHELM ANALYST - Analyst knows he/she cannot possibly process all text thoroughly; discouraging to those who want to do a good job.

APPENDIX A: PROBLEM ANALYSIS OUTLINE, cont'd.

1.1 ESSENTIAL FUNCTION: FIND POTENTIAL TEXT-SOURCES OF DATA (ON TOPICS OF INTEREST), cont'd.

B. PROBLEMS/UNEXPLOITED OPPORTUNITIES, cont'd.

2. PROBLEMS ASSOCIATED WITH HIGH RESULTANT VOLUMES OF TEXT and HIGH % UNAPPLICABLE INFO, cont'd.:

- o TOO TIME-CONSUMING - Analyst does not have time to read it all
- o MISS IMPORTANT DATA - Like looking for a "needle in a haystack"
- o REDUCES TIME ANALYST HAS LEFT TO ANALYZE DATA
  - o Directly impacts the intelligence and operations communities
  - o Impacts Analysts' knowledge of the threat.
  - o Impacts military systems' acquisitions, which are based on Analysts' reports of threats.
  - o Impacts effectiveness of military systems' design: Analysts' knowledge of threat is used to evaluate the effectiveness of military system designs under consideration.

3. MISSING important data
   - o Can directly impact the intelligence and operations communities.

1.2 ESSENTIAL FUNCTION: DETERMINE SPECIFIC TOPIC(S) OF TEXT (i.e., the Essential Data composing the text)

A. CURRENT IMPLEMENTATION: ANALYST
- The analyst reads or scans through the subset of free text intelligence messages resulting from the profile to determine topic of text

B. PROBLEMS/UNEXPLOITED OPPORTUNITIES
- o INPUT VOLUME TOO LARGE FOR HUMAN (ANALYST) TO PROCESS IN AVAILABLE TIME
- o TOO TIME-CONSUMING - Since Analyst does not have time to read all the text, he/she may have to skip over some text which is not obviously interesting, or scan it very quickly.
- o MISS IMPORTANT DATA - Like looking for a "needle in a haystack".
- o INTRODUCES "HUMAN ERROR" - Humans can accidently overlook data, become fatigued and miss data, etc.

1.3 ESSENTIAL FUNCTION: DETERMINE IF TOPIC OF TEXT = TOPIC OF INTEREST (find specific data on topic of interest)

A. CURRENT IMPLEMENTATION: ANALYST
- The analyst determines if the text contains data on the topic of interest.

B. PROBLEMS/UNEXPLOITED OPPORTUNITIES
o INTRODUCES "HUMAN ERROR" - Humans are good at this type of judgement, for the most part. The biggest problem might be that humans can become fatigued and make mistakes, or just make an error in judgement.

1.4 ESSENTIAL FUNCTION: EXTRACT ESSENTIAL DATA ON TOPIC OF INTEREST FROM TEXT
- Extract Data on topics of interest from text, for Analysis

A. CURRENT IMPLEMENTATION: MANUALLY, BY ANALYST (IF DONE AT ALL).
- The individual Analyst may (or may not) save text or data of interest in a file for future reference or Analysis.

B. PROBLEMS/UNEXPLOITED OPPORTUNITIES
o DATA ON A GIVEN TOPIC MAY BE DIFFICULT TO FIND OR ACCESS OVER TIME
o RESULTS NOT IN COMPUTER-USABLE FORM - As a result, application software cannot be developed to utilize the data, both for Analysis, and for finding data on topics of interest.

APPENDIX A: PROBLEM ANALYSIS OUTLINE, cont'd.

2.0 ESSENTIAL FUNCTION: ANALYSIS
- The process of obtaining useful intelligence information from data on topics of interest. Analysis of data may include:
o Reasoning about the significance of data; identifying/deducing important information
o Identifying patterns and relationships in the data; reasoning about their significance
o Identifying changes in the norm; reasoning about their significance

A. CURRENT IMPLEMENTATION: NO ESTABLISHED METHOD; UP TO THE INDIVIDUAL ANALYST.
- No automated tools are currently available to support the analysis of intelligence data (beyond rudimentary text processing SW).

B. PROBLEMS/UNEXPLOITED OPPORTUNITIES

o MISSING THE OPPORTUNITY TO UTILIZE CURRENT GRAPHICS TECHNOLOGY
- Current graphics technology exists which would allow data, in computer-usable format such as DB records, to be represented in various pictorial forms. Graphic representations of intel data would:

o MAKE IT EASIER TO INTERPRET THE SIGNIFICANCE OF THE DATA (e.g., plotting data on maps, or timelines)

o IMPROVE ANALYSTS' ABILITY TO CONCEPTUALIZE PATTERNS AND RELATIONSHIPS IN THE DATA, AND REASON ABOUT THEIR SIGNIFICANCE.

o FACILITATE ANALYST'S ABILITY TO SPOT CHANGES IN DATA - Changes in data may not be as obvious in textual format as they would be if represented graphically.

o MISSING OPPORTUNITY TO DEVELOP & UTILIZE APPLICATION SW WHICH COULD FACILITATE INTEL DATA ANALYSIS - For example, there are Analysts at FTD interested in Cybernetic Modeling of data. But SW tools supporting such applications cannot be developed and utilized until the intel data is put into a format that is usable by computer software.

o MISSING THE OPPORTUNITY TO UTILIZE CURRENT SW TECHNOLOGY TO ALLOW INTERACTIVE MANIPULATION AND ANALYSIS OF DATA (automate the process of finding data, configuring data, displaying data, etc.)

o MISSING THE OPPORTUNITY TO DEVELOP APPLICATION SW CAPABLE OF IDENTIFYING CHANGES FROM THE NORM.
- If intel data was represented in computer usable format, application software could be developed to help look for changes in the data that the analyst might not spot otherwise.

APPENDIX B: REQUIREMENTS FOR SOLVING THE OPERATIONAL PROBLEM (LONG-TERM)

SYSTEM SCOPE:

I. INPUT SCOPE
   A. TYPE: Unformatted free-text.
      - o message traffic
      - o newsletters
      - o reports
      - o books
      - o data from external databases (e.g., newswire databases, libraries, etc.)
   B. FORM:
      - o digital
      - o hardcopy
   C. SUBJECT DOMAIN: Broad-domain (not restricted to one particular subject area).

II. OUTPUT SCOPE:
   A. Raw, unprocessed text
   B. Graphical Displays (e.g., timelines, maps)
   C. Graphical Displays Correlated with Intelligence Data
   D. Displays/Files created interactively by the User
   E. Data Base Records (i.e., the output of the ADBG subcomponent) containing the Essential Data from the input text
   F. Error Handling Information

APPENDIX B: REQUIREMENTS FOR SOLVING THE OPERATIONAL PROBLEM (LONG-TERM), cont'd.

SYSTEM REQUIREMENTS:

I. FUNCTIONAL REQUIREMENTS:

o ABILITY TO FIND DATA ON TOPICS IF INTEREST - Ability to find, access, and retrieve data on topics of interest to the Analyst, over time. I.e., analyst should be able to access data on a topic of interest, which covers (was received or created) over some (TBD) span of time.

o ABILITY TO FIND DATA ON CONCEPTS OF INTEREST - This would be a conceptually higher-level search than that for a topic, in that a concept may incorporate several topics, and may imply certain relationships. E.g., the TOPIC "bombing" is only one topic which may fall under the CONCEPT such as "terrorist attack".

o FACILITATE THE ANALYSTS ABILITY TO ASSIMILATE MORE DATA, MORE QUICKLY (e.g., by utilizing graphics)

o ABILITY TO CREATE GRAPHICAL DISPLAYS OF THE DATA - The displays shall present the data in a form which is more meaningful and useful to the Analyst than just text (e.g., timelines, maps, Cybernetic Modeling tables).

o FACILITATE ANALYST'S ABILITY TO ANALYZE INTEL DATA (make their job easier, less time-consuming, and improve their ability to do it well). I.e., FACILITATE ANALYST'S ABILITY TO:
  - o CONCEPTUALIZE data
  - o REASON about the meaning and significance of data
  - o INTERPRET the potential significance of data
  - o FIND/RECOGNIZE POTENTIALLY IMPORTANT PATTERNS AND RELATIONSHIPS IN THE DATA
  - o ID CHANGES in the DATA from the norm (i.e., ID potentially important info);

o FACILITATE ANALYSTS' ABILITY TO MANIPULATE TEXT AND INTEL DATA

o INTERACTIVE USER CONFIGURATION OF TEXT, DATA, AND DISPLAYS.

o USER CONTROL OF PROCESSING.

o ENABLE THE DEVELOPMENT & UTILIZATION OF APPLICATION SW FOR ANALYSIS OF INTEL DATA - I.e., structure & store data on topics of interest in a computer-usable format which can easily be used by other application SW.

o GRACEFUL ERROR RECOVERY - Ability to gracefully handle and recover from erroneous or unexpected inputs, and processing problems.

o POTENTIAL REQ.: FACILITATE ANALYST'S ABILITY TO SPOT POTENTIALLY IMPORTANT, BUT UNEXPECTED INFO (i.e., on topics that analyst may not know to look for).

APPENDIX B: REQUIREMENTS FOR SOLVING THE OPERATIONAL PROBLEM (LONG-TERM)

I. FUNCTIONAL REQUIREMENTS, cont'd.

o POTENTIAL REQ.: DEVELOP APPLICATION SW TO ID CHANGES IN DATA FROM THE NORM (i.e., potentially important, and possibly unexpected, info.)

APPENDIX B: REQUIREMENTS FOR SOLVING THE OPERATIONAL PROBLEM (LONG-TERM), cont'd.

II. PERFORMANCE REQUIREMENTS:

o REDUCE PROCESSING TIME - Significantly reduce the time spent by the Analyst in processing the text, looking for data on topics of interest, so that the analysts have more time for Analysis of the data.

o REDUCE VOLUME OF TEXT WHICH MUST BE MANUALLY PROCESSED - Reduce the volume of text that the analyst must look through to find data of interest for analysis.

o OPTIMIZE RECALL - Significantly improve the analyst's ability to find/access ALL data on a topic of interest (reduce missed information).

o OPTIMIZE PRECISION - Significantly improve the analyst's ability to find/access ONLY data on their topic of interest; reduce the volume of excessive, non-applicable text.

APPENDIX B: REQUIREMENTS FOR SOLVING THE OPERATIONAL PROBLEM (LONG-TERM), cont'd.

SYSTEM REQUIREMENTS:

III. DESIGN REQUIREMENTS:

o A man-in-the-loop IS acceptable, as long as the other requirements are met.

o PORTABLE

GROWTH REQUIREMENTS

o MODULAR SYSTEM DESIGN

o TRANSPORTABLE SUBCOMPONENTS - The system design shall facilitate the replacement of functional subcomponents, so that as new technology advances and improves in performance, it can replace existing subcomponent which do not perform as well.

o EXPANDABLE SUBCOMPONENTS - The system design must anticipate the expansion of its subcomponents, and develop a design to facilitate subcomponent expansion. Expansion of subcomponents is anticipated in order to add functionality to subcomponents (e.g., we anticipate the addition of graphical utilities to the Intel-Data Analysis Utilities Subcomponent over time), and to increase the input scope of the system in future efforts (e.g., we anticipate the need to expand the ADBG Subcomponent).

o SYSTEM EXPANDABILITY - The system design must anticipate extension, and be designed to facilitate this process. For example, it is anticipated that subsequent efforts will want to utilize output from the ADBG subcomponent as input to other special-purpose applications.

o FLEXIBILITY OF DESIGN - The need to change the system and its subcomponents is anticipated and expected. The system shall be designed to facilitate such changes (make as painless as possible).

o EVOLVABLE DESIGN FOR SYSTEM FRAMEWORK and SYSTEM SUBCOMPONENTS - The system design must anticipate the fact that the system and its subcomponents' designs are going to evolve over time, as both their respective technologies evolve, and as feedback from this initial operational capability produces refinement of the system requirements. Therefore, it is critical that the system be DESIGNED to anticipate and facilitate the evolutionary development of the system. For example, use of formal methodology in specifying the system, such as structured analysis, would probably help make the design easier to evolve.

APPENDIX B: REQUIREMENTS FOR SOLVING THE OPERATIONAL PROBLEM (LONG-TERM), cont'd.

III. DESIGN REQUIREMENTS, cont'd.

GROWTH REQUIREMENTS, cont'd.

o SPECIFICATION AND USE OF A FORMAL SYSTEM ANALYSIS, SPECIFICATION AND DESIGN METHODOLOGY - Use of a formal System Analysis & Specification Methodology should help encourage a well-thought out and strong system design. It supports several of the systems design requirements, such as modularity of system design, evolvability of the design, etc.

o REUSABILITY OF SYSTEM FRAMEWORK AND SUBCOMPONENTS - Since one of our big goals is cost-efficiency over the long-run, we want to design as much of the system as possible (and as makes sense) to be useful in subsequent projects, so we do not unnecessarily duplicate work done from project to project. I.e.:

  o PROMOTE REUSABILITY OF STABLE PARTS of SYSTEM in future efforts, so future efforts can consider these parts of the system as "givens" (e.g., the system framework and user interface). This will allow future efforts to focus on the R&D of the more unstable or technologically immature components (e.g., the ADBG Subcomponent), and on the expansion of the system scope (in terms of input, output, and processing).

  o MINIMIZE PORTIONS OF THE SYSTEM WHICH WILL NEED TO BE REDESIGNED IN FUTURE ITERATIONS. Identify and isolate unstable functions of system from the more stable portions.

o PROMOTE REUSABILITY AND EVOLVABILITY OF KNOWLEDGE gained in this effort, in terms of all the requirements and design specifications. For example, use of a formal system specification methodology to develop and communicate system design specifications would result in specification-level "blueprints" which could be reused from project to project.

FOR ADBG SUBCOMPONENT:

o DESIGN WHICH SUPPORTS INCREMENTAL & ITERATIVE DEVELOPMENT OF A BROAD-COVERAGE ADBG CAPABILITY - I.E., an expandable, upward compatible ADBG design approach.

FOR INTERACTIVE USER INTERFACE

o "LIKABILITY" - anticipate and design for a user who does NOT have a background in linguistics or computer science.

o EASY TO LEARN/EASE OF USE - anticipate and design for a user who does NOT have a background in linguistics or computer science.

APPENDIX C: NEAR-TERM SYSTEM REQUIREMENTS

## OVERALL SYSTEM REQUIREMENTS & DESIGN REQUIREMENTS

Develop a System that will:

1. SERVE AS THE INITIAL, BUILD 1 BASELINE FOR FUTURE SYSTEM BUILDS, in the evolutionary development of a system capable of meeting our long-term goals. The Near-Term System will provide the foundational system structure and functions for future efforts to build upon.
WHY: Want to obtain a sound, foundational system, which is designed to facilitate the iterative [and incremental] development of its functional subcomponents. This will allow future efforts to focus on evolving the individual subcomponents, without having to redevelop the framework and more stable subcomponents of the system each project.

AS THE BUILD 1 BASELINE, THE NEAR-TERM SYSTEM SHALL:

A. IMPLEMENT AND STABILIZE THOSE PARTS OF THE NEAR-TERM SYSTEM WHICH CAN BE STABILIZED (whose functional designs and technologies are more stable). E.g., The System Framework, the User Interface, those portions of the Imagery/Intelligence Data Display Subcomponent which get implemented, etc.
WHY: In order to reduce the number of subcomponents which need to be worked on or redeveloped in subsequent projects.
(See Functional Requirements for details on which subcomponents are to be stabilized this Near-Term effort).

B. SERVE AS A STABLE PROTOTYPING FRAMEWORK FOR THE ADBG SUBCOMPONENT, AND ITS FUNCTIONAL SUBCOMPONENTS.
Since ADBG technology is still evolving and changing, an objective of the system is to provide a stable structure which facilitates the investigation of alternative ADBG technologies and approaches, without having to redesign the rest of the system over and over again.

2. WILL BE INSTALLED and RUNNING AT FTD, IN THE OPERATIONAL ENVIRONMENT.

3. WILL DEMONSTRATE END-TO-END FUNCTIONAL PROCESSING.
The system will perform complete, end-to-end functional processing in an integrated framework. The Build 1 system will have the total BREADTH of functionality we are aiming for in the long-term system, but only a subset of its anticipated DEPTH of functionality, and input and output scope.

4. WILL DEMONSTRATE THE USEFULNESS OF THE SYSTEM CONCEPT in meeting our operational requirements.

APPENDIX C: NEAR-TERM SYSTEM REQUIREMENTS, cont'd.

SYSTEM SCOPE:

I. INPUT SCOPE:
   A. TYPE: Unformatted free-text.
      o message traffic
      o (possibly) reports
   B. FORM: Digital
   C. SUBJECT DOMAIN: "Perestroika Watch". Information having to do with changes in the Soviet Union associated with Perestroika, such as:
      o changes in doctrine
      o economic restructuring
      o openness, etc.
      If the input domain is too large, contractor and Government representatives will work together to designate a feasible, but non-trivial, subset of the input scope.

II. OUTPUT SCOPE
   A. Raw, unprocessed text
   B. Graphical Displays (e.g., timelines, maps)
   C. Graphical Displays Correlated with Intelligence Data
   D. Displays/Files created interactively by the User
   E. Data Base Records (i.e., the output of the ADBG subcomponent) containing the Essential Data from the input text
   F. Error Handling Information

APPENDIX C: NEAR-TERM SYSTEM REQUIREMENTS, cont'd.

SYSTEM REQUIREMENTS:

I. FUNCTIONAL REQUIREMENTS:

o ABILITY TO FIND DATA ON TOPICS IF INTEREST - Ability to find, access, and retrieve data on topics of interest to the Analyst, over time. I.e., analyst should be able to access data on a topic of interest, which covers (was received or created) over some (TBD) span of time.

o ABILITY TO FIND DATA ON CONCEPTS OF INTEREST - Desirable, but not essential for Near-Term effort.

o FACILITATE THE ANALYSTS ABILITY TO ASSIMILATE MORE DATA, MORE QUICKLY (e.g., by utilizing graphics)

o ABILITY TO CREATE GRAPHICAL DISPLAYS OF THE DATA - The displays shall present the data in a form which is more meaningful and useful to the Analyst than just text (e.g., timelines, maps, Cybernetic Modeling tables).

o FACILITATE ANALYST'S ABILITY TO ANALYZE INTEL DATA (make their job easier, less time-consuming, and improve their ability to do it well). I.e., FACILITATE ANALYST'S ABILITY TO:
  - o CONCEPTUALIZE data
  - o REASON about the meaning and significance of data
  - o INTERPRET the potential significance of data
  - o FIND/RECOGNIZE POTENTIALLY IMPORTANT PATTERNS AND RELATIONSHIPS IN THE DATA
  - o ID CHANGES in the DATA from the norm (i.e., ID potentially important info);

o FACILITATE ANALYSTS' ABILITY TO MANIPULATE TEXT AND INTEL DATA

o INTERACTIVE USER CONFIGURATION OF TEXT, DATA, AND DISPLAYS.

o USER CONTROL OF PROCESSING.

o ENABLE THE DEVELOPMENT & UTILIZATION OF APPLICATION SW FOR ANALYSIS OF INTEL DATA - I.e., structure & store data on topics of interest in a computer-usable format which can easily be used by other application SW.

o GRACEFUL ERROR RECOVERY - Ability to gracefully handle and recover from erroneous or unexpected inputs, and processing problems.

o POTENTIAL REQ.: FACILITATE ANALYST'S ABILITY TO SPOT POTENTIALLY IMPORTANT, BUT UNEXPECTED INFO (i.e., on topics that analyst may not know to look for).
- Not required for Near-Term effort

APPENDIX C: NEAR-TERM SYSTEM REQUIREMENTS, cont'd

I. FUNCTIONAL REQUIREMENTS, cont'd.

o POTENTIAL REQ.: DEVELOP APPLICATION SW TO ID CHANGES IN DATA FROM THE NORM (i.e., potentially important, and possibly unexpected, info.)
- Not required for Near-Term effort

APPENDIX C: NEAR-TERM SYSTEM REQUIREMENTS, cont'd.

II. PERFORMANCE REQUIREMENTS:

o REDUCE PROCESSING TIME - Reduce the time spent by the Analyst in processing the text, looking for data on topics of interest, so that the analysts have more time for Analysis of the data.
For Near-Term System: Should, at a minimum, be faster than if the Analyst performed the Processing and Analysis manually.

o REDUCE VOLUME OF TEXT WHICH MUST BE MANUALLY PROCESSED - Reduce the volume of text that the analyst must look through to find data on topics of interest for analysis.

o IMPROVE RECALL - Improve the analyst's current ability to find/access ALL data on a topic of interest (reduce missed information).

o IMPROVE PRECISION - Improve the analyst's ability to find/access ONLY data on their topic of interest; reduce the volume of excessive, non-applicable text.

APPENDIX C: NEAR-TERM SYSTEM REQUIREMENTS, cont'd.

III. DESIGN REQUIREMENTS

o A man-in-the-loop IS acceptable, as long as the other requirements are met.

o PORTABLE.

o MINIMIZE RISK - The main purposes of the Near-Term effort are to design and develop a strong system architecture for future efforts, and to obtain the capability to demonstrate the system's functional concept. It is NOT the goal of this particular effort to push the state-of-the-art in any of the individual functional areas of the system. The goal this effort is to apply existing, less risky technology to our application, not to develop new technology.

o UTILIZE EXISTING TECHNOLOGY, AND COMMERCIAL-OFF-THE-SHELF (COTS) PRODUCTS, TO THE EXTENT POSSIBLE AND FEASIBLE.

o MAXIMIZES THE USEFULNESS OF NEAR-TERM EFFORT to the extent that will not jeopardize achieving our long-term goals. More specifically:

  o NARROW THE LARGE GULF OF TIME & EFFORT, which the analyst currently must span, in order to find and extract data from free-text for Analysis.

  o REDUCE THE MANUAL MESSAGE-PROCESSING BURDEN currently experienced by the analyst, in finding and extracting data on topics of interest from text.

GROWTH REQUIREMENTS

o MODULAR SYSTEM DESIGN

o TRANSPORTABLE SUBCOMPONENTS - The system design shall facilitate the replacement of functional subcomponents, so that as new technology advances and improves in performance, it can replace existing subcomponent which do not perform as well.

o EXPANDABLE SUBCOMPONENTS - The system design must anticipate the expansion of its subcomponents, and develop a design to facilitate subcomponent expansion. Expansion of subcomponents is anticipated in order to add functionality to subcomponents (e.g., we anticipate the addition of graphical utilities to the Intel-Data Analysis Utilities Subcomponent over time), and to increase the input scope of the system in future efforts (e.g., we anticipate the need to expand the ADBG Subcomponent).

o SYSTEM EXPANDABILITY - The system design must anticipate extension, and be designed to facilitate this process. For example, it is anticipated that subsequent efforts will want to utilize output from the ADBG subcomponent as input to other special-purpose applications.

III. DESIGN REQUIREMENTS, cont'd.

GROWTH REQUIREMENTS, cont'd.

o FLEXIBILITY OF DESIGN - The need to change the system and its subcomponents is anticipated and expected. The system shall be designed to facilitate such changes (make as painless as possible).

o EVOLVABLE DESIGN FOR SYSTEM FRAMEWORK and SYSTEM SUBCOMPONENTS - The system design must anticipate the fact that the system and its subcomponents' designs are going to evolve over time, as both their respective technologies evolve, and as feedback from this initial operational capability produces refinement of the system requirements. Therefore, it is critical that the system be DESIGNED to anticipate and facilitate the evolutionary development of the system. For example, use of formal methodology in specifying the system, such as structured analysis, would probably help make the design easier to evolve.

o SPECIFICATION AND USE OF A FORMAL SYSTEM ANALYSIS, SPECIFICATION AND DESIGN METHODOLOGY - Use of a formal System Analysis & Specification Methodology should help encourage a well-thought out and strong system design. It supports several of the systems design requirements, such as modularity of system design, evolvability of the design, etc.

o REUSABILITY OF SYSTEM FRAMEWORK AND SUBCOMPONENTS - Since one of our big goals is cost-efficiency over the long-run, we want to design as much of the system as possible (and as makes sense) to be useful in subsequent projects, so we do not unnecessarily duplicate work done from project to project. I.e.:

  - o PROMOTE REUSABILITY OF STABLE PARTS of SYSTEM in future efforts, so future efforts can consider these parts of the system as "givens" (e.g., the system framework and user interface). This will allow future efforts to focus on the R&D of the more unstable or technologically immature components (e.g., the ADBG Subcomponent), and on the expansion of the system scope (in terms of input, output, and processing).

  - o MINIMIZE PORTIONS OF THE SYSTEM WHICH WILL NEED TO BE REDESIGNED IN FUTURE ITERATIONS. Identify and isolate unstable functions of system from the more stable portions.

o PROMOTE REUSABILITY AND EVOLVABILTY OF KNOWLEDGE gained in this effort, in terms of all the requirements and design specifications. For example, use of a formal system specification methodology to develop and communicate system design specifications would result in specification-level "blueprints" which could be reused from project to project.

APPENDIX C: NEAR-TERM SYSTEM REQUIREMENTS, cont'd.

III. DESIGN REQUIREMENTS, cont'd.

GROWTH REQUIREMENTS, cont'd.

FOR ADBG SUBCOMPONENT:

o DESIGN WHICH SUPPORTS INCREMENTAL & ITERATIVE DEVELOPMENT OF A BROAD-COVERAGE ADBG CAPABILITY - I.e., an expandable, upward compatible ADBG design approach.

FOR INTERACTIVE USER INTERFACE

o "LIKABILITY" - anticipate and design for a user who does NOT have a background in linguistics or computer science.

o EASY TO LEARN/EASE OF USE - anticipate and design for a user who does NOT have a background in linguistics or computer science.

APPENDIX D: LONG-TERM STRATEGIC PLAN

OVERVIEW
---

Develop a strategic plan (i.e., design approach) for HOW to achieve the long-term functional & performance goals, over the course of successive efforts. This should be done in a manner that is cost-efficient over the long-run, and which maximizes the usefulness of the Near-Term and Interim efforts to the extent possible.

DESIGN REQS:
---

Several design criteria should be considered in developing the Long-Term Strategic Plan. These include:

o COST-EFFICIENT OVER THE LONG-RUN*
The long-term approach for solving the operational problem shall be designed to minimize unnecessary expenditures over the long-run. This includes:
  o MINIMIZING UNPRODUCTIVE REDUNDANCY in efforts over the long-run; and
  o PROMOTING PLANNED FOR REUSABILITY of the products of individual efforts, in succeeding efforts. E.g., Planning to obtain reusable knowledge: requirements documents, functional specs, etc; planning to reuse stable portions of the system design; etc.)

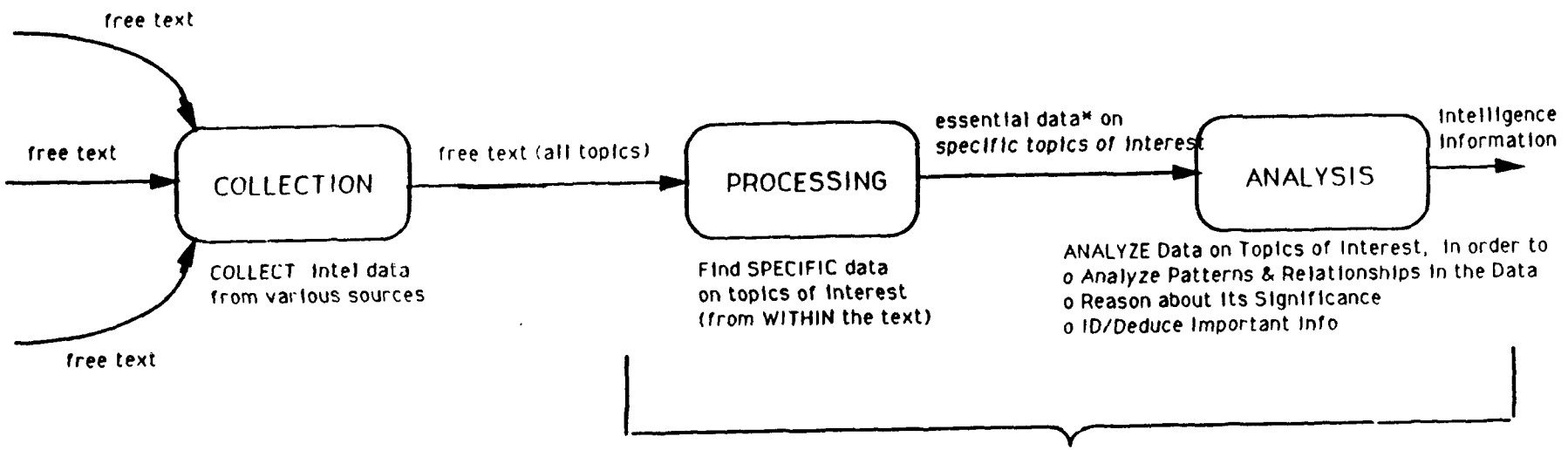
o RISK MINIMIZATION
The long-term strategy should be developed and approved by experts in related technical fields, in order to assure that it is both technically sound and feasible.

o MAXIMIZES THE USEFULNESS OF NEAR-TERM AND INTERIM EFFORTS
The Near-Term and Interim Solutions shall be designed to act as useful "bridges" between our current method of Processing free-text (and their problems), and our Long-Term goals for Processing free-text. The Near-Term and Interim Systems shall be designed to:
- NARROW THE LARGE GULF OF TIME & EFFORT, which the analyst currently must span, in order to find and extract data from free-text for Analysis. See Fig. 11.
- REDUCE THE MANUAL MESSAGE-PROCESSING BURDEN currently experienced by the analyst in processing and analyzing data from unformatted text.

* "Cost-efficient over the long run" means that $ spent on an effort should result in an equitably useful return, in terms of:
  o the advancement of our knowledge on HOW to proceed from Near-Term to Long-Term capabilities; or
  o improved or expanded capabilities in line with our Long-Term goals;
(in other words, we don't spend $ unless it gets us closer to our long-term goals).

# FIG. 1 OPERATIONAL PROBLEM SCOPE

## 3 ESSENTIAL STEPS IN OBTAINING INTEL INFORMATION FROM TEXT; COLLECTION, PROCESSING AND ANALYSIS

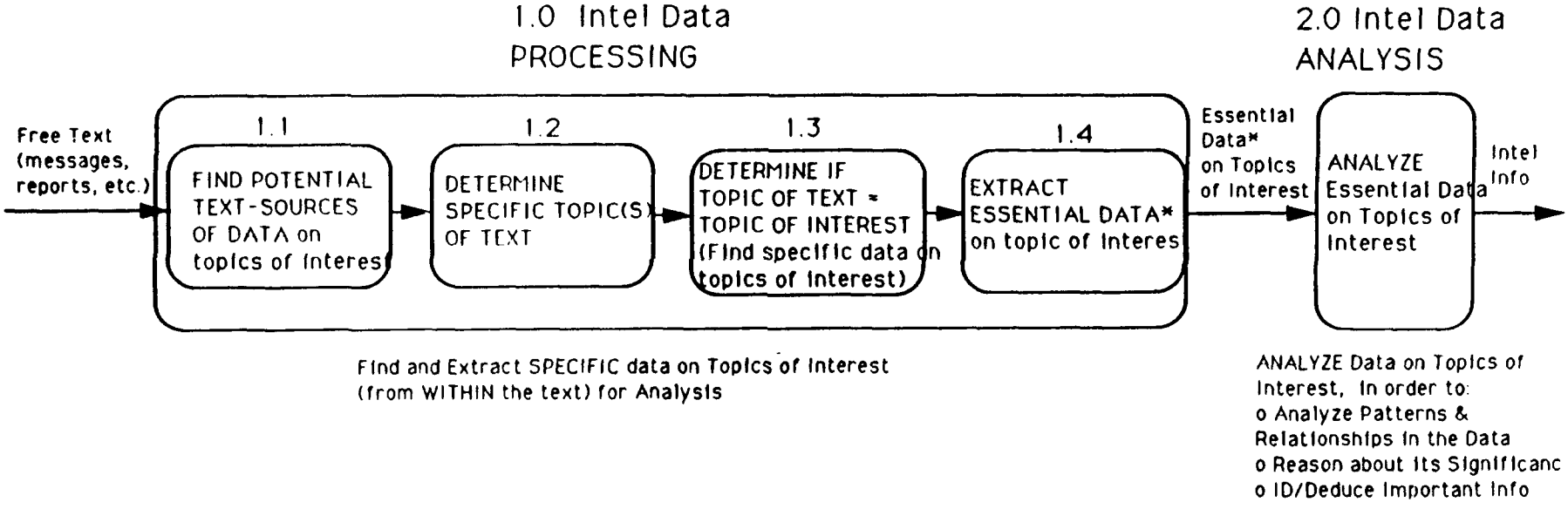



IAA OPERATIONAL PROBLEM SCOPE;
PROCESSING and ANALYSIS of TEXT

* essential data =
who, what, where, when, how, etc.

# FIG.2: THE ESSENTIAL FUNCTIONAL STEPS FOR SOLVING THE OPERATIONAL PROBLEM

## WHAT needs to be done: Implementation-Independent

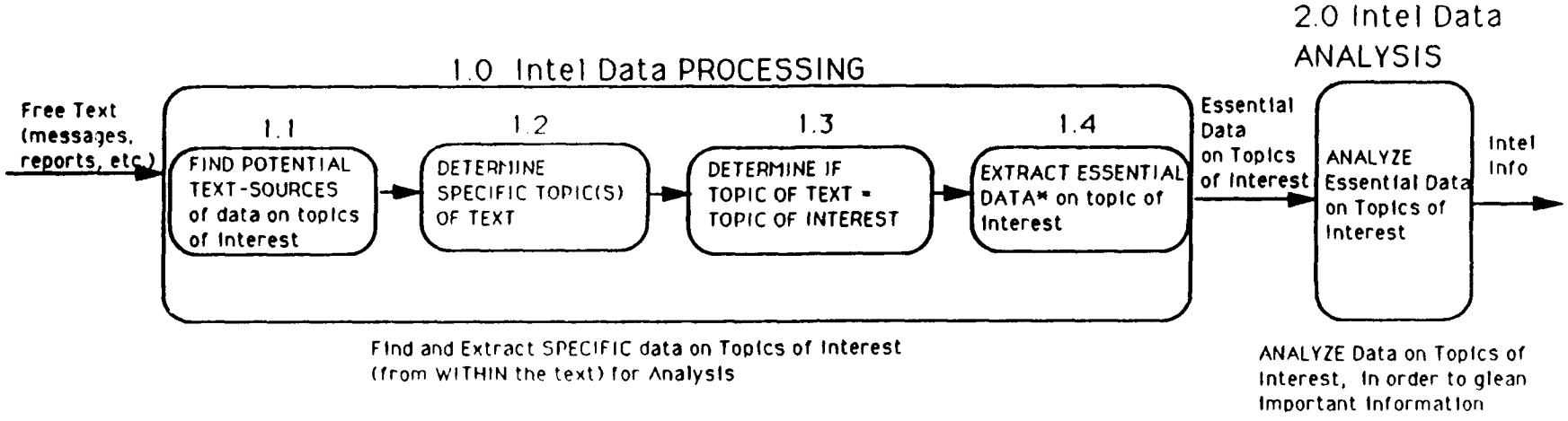



Note: Not all implementations do all functions (e.g., some ADBG systems do not implement 1.1), and some implementations do the functions in different order.

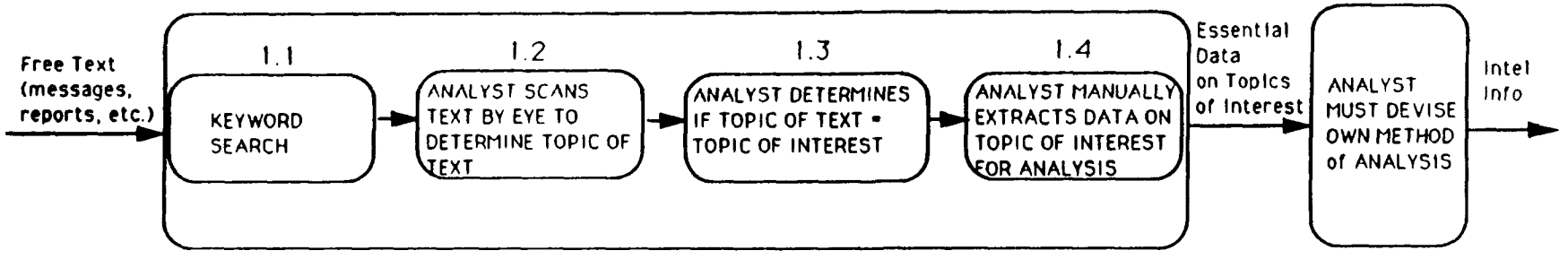
* essential data = who, what, where, when, how, etc.

# FIG. 3: THE CURRENT METHOD OF SOLVING THE OPERATIONAL PROBLEM

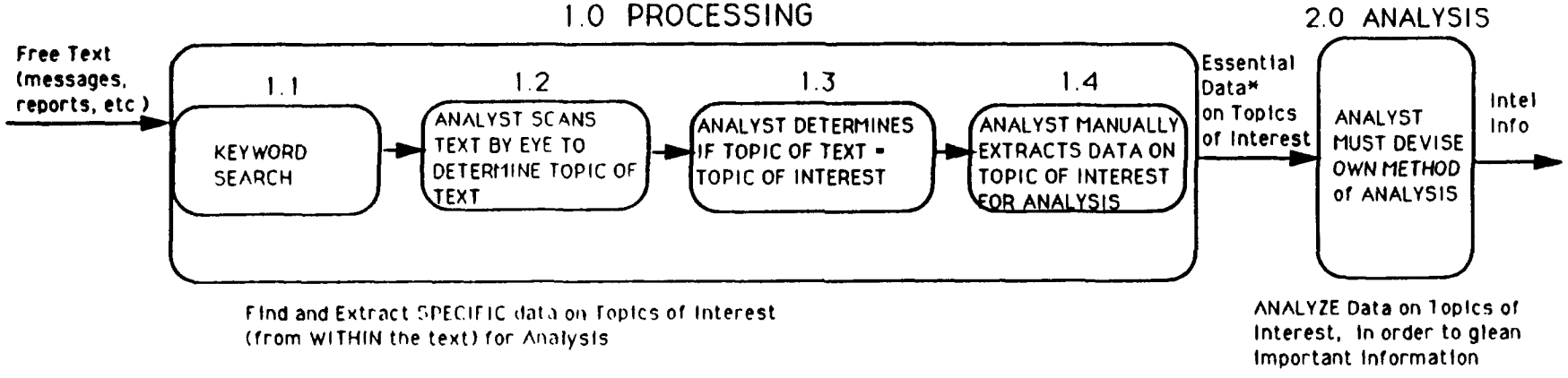
The Essential Functional Steps
- WHAT needs to be done (Implementation-Independent)




THE CURRENT METHOD for SOLVING the OPERATIONAL PROBLEM
- HOW the essential functions are implemented



# Fig. 4: Problems associated with the Current Method of Solving the Operational Problem




**PROBLEMS with Current Method of *PROCESSING*:**

**LARGELY UNAUTOMATED**

**KEWORD SEARCH RESULTS IN UNDESIRABLE, "CATCH 22", RESULTS:**
o BROAD KEYWORD PROFILE
Find Most Data, but High Volume of Text
vs
o NARROW KEYWORD SEARCH
Low Volume of Text, but Miss Important Data

**LARGE VOLUMES of INFORMATION:**
o TOO TIME-CONSUMING to process all of it (takes too much time from actual analysis)
o MISS IMPORTANT INTELLIGENCE DATA ("needle in a haystack")

**RESULTS NOT IN COMPUTER-USABLE FORM**

**PROBLEMS with Current Method of *ANALYSIS*:**

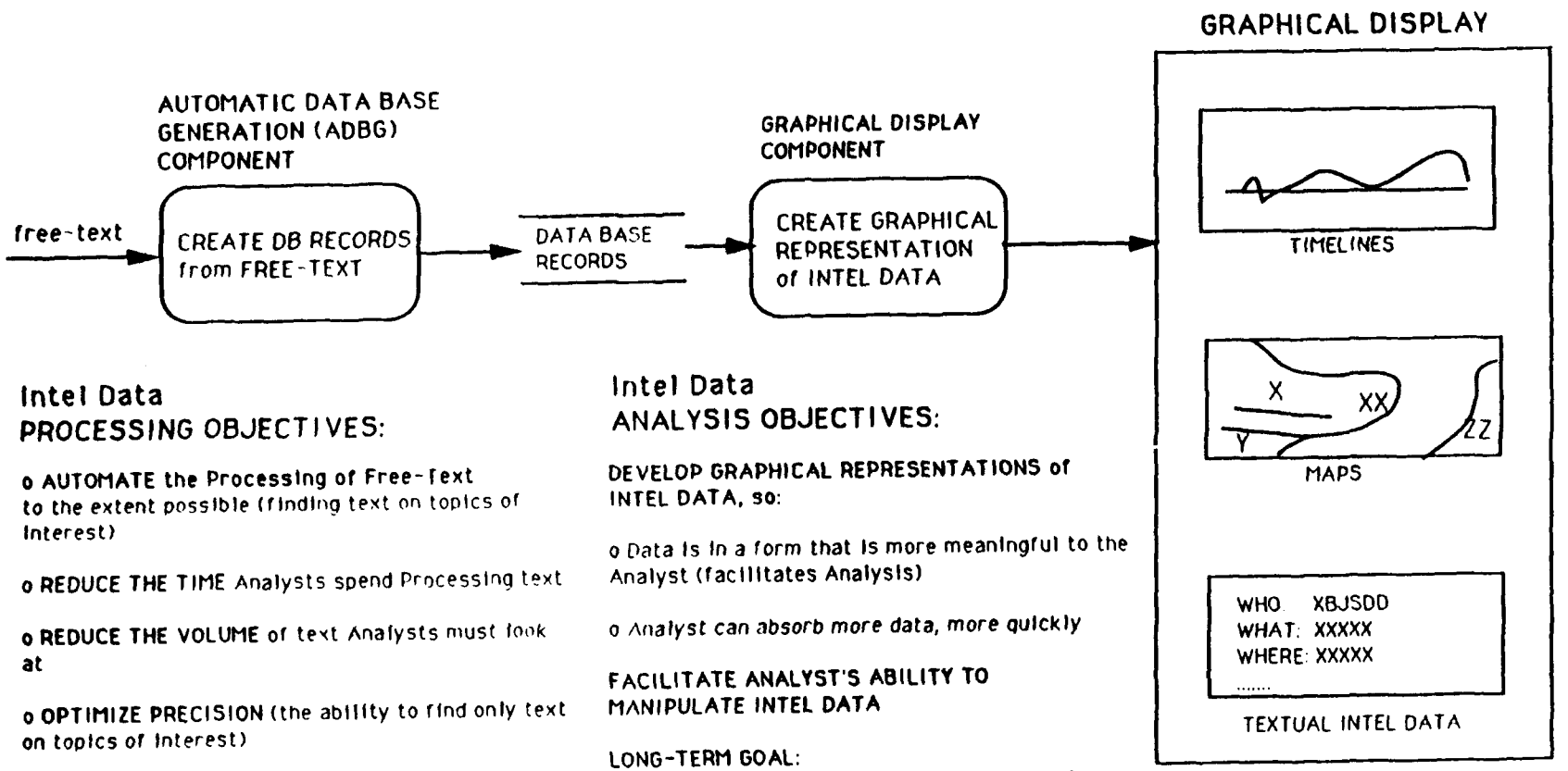
**UNEXPLOITED OPPORTUNITIES (due to the Input to Analysis NOT being in Computer-Usable Form):**

**o UNABLE TO TAKE ADVANTAGE of GRAPHICS to FACILITATE ANALYSIS**

**o UNABLE TO TAKE ADVANTAGE of SW UTILITIES to FACILITATE ANALYSIS**

* essential data = who, what, where, when, how, etc.

# Fig. 5: THE NEW METHOD for SOLVING the OPERATIONAL PROBLEM: High-Level Overview & Approach Objectives




**Intel Data PROCESSING OBJECTIVES:**

- **AUTOMATE the Processing of Free-Text** to the extent possible (finding text on topics of interest)
- **REDUCE THE TIME** Analysts spend Processing text
- **REDUCE THE VOLUME** of text Analysts must look at
- **OPTIMIZE PRECISION** (the ability to find only text on topics of interest)
- **OPTIMIZE RECALL** (the ability to find all text on topics of interest)
- **PUT ESSENTIAL DATA extracted from text in COMPUTER-USABLE FORM** (Enable the use of SW utilities for Analysis)

**Intel Data ANALYSIS OBJECTIVES:**

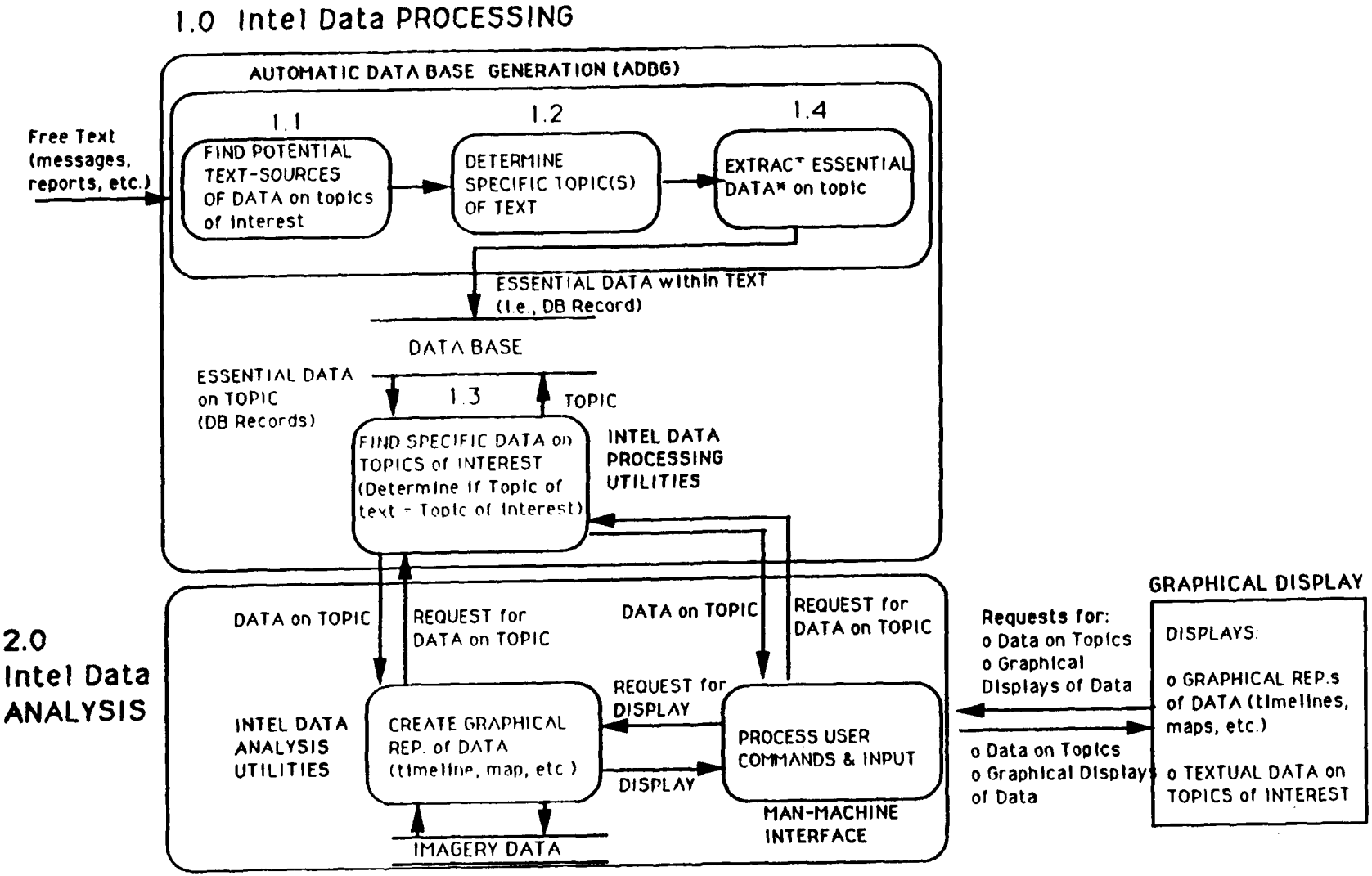
**DEVELOP GRAPHICAL REPRESENTATIONS of INTEL DATA, so:**

- Data is in a form that is more meaningful to the Analyst (facilitates Analysis)
- Analyst can absorb more data, more quickly

**FACILITATE ANALYST'S ABILITY TO MANIPULATE INTEL DATA**

**LONG-TERM GOAL:**
**ABILITY TO DEVELOP & USE SW UTILITIES TO FACILITATE ANALYSIS (e.g., Cybernetic Modelling).** This is one of the main reasons that the data from Processing needs to be in a structured, computer-usable form.

# Fig. 6 THE NEW METHOD for SOLVING the OPERATIONAL PROBLEM - HOW the essential functions are implemented



# Fig. 7: THE NEW METHOD for SOLVING the OPERATIONAL PROBLEM & HOW IT WILL MEET OUR OPERATIONAL REQUIREMENTS

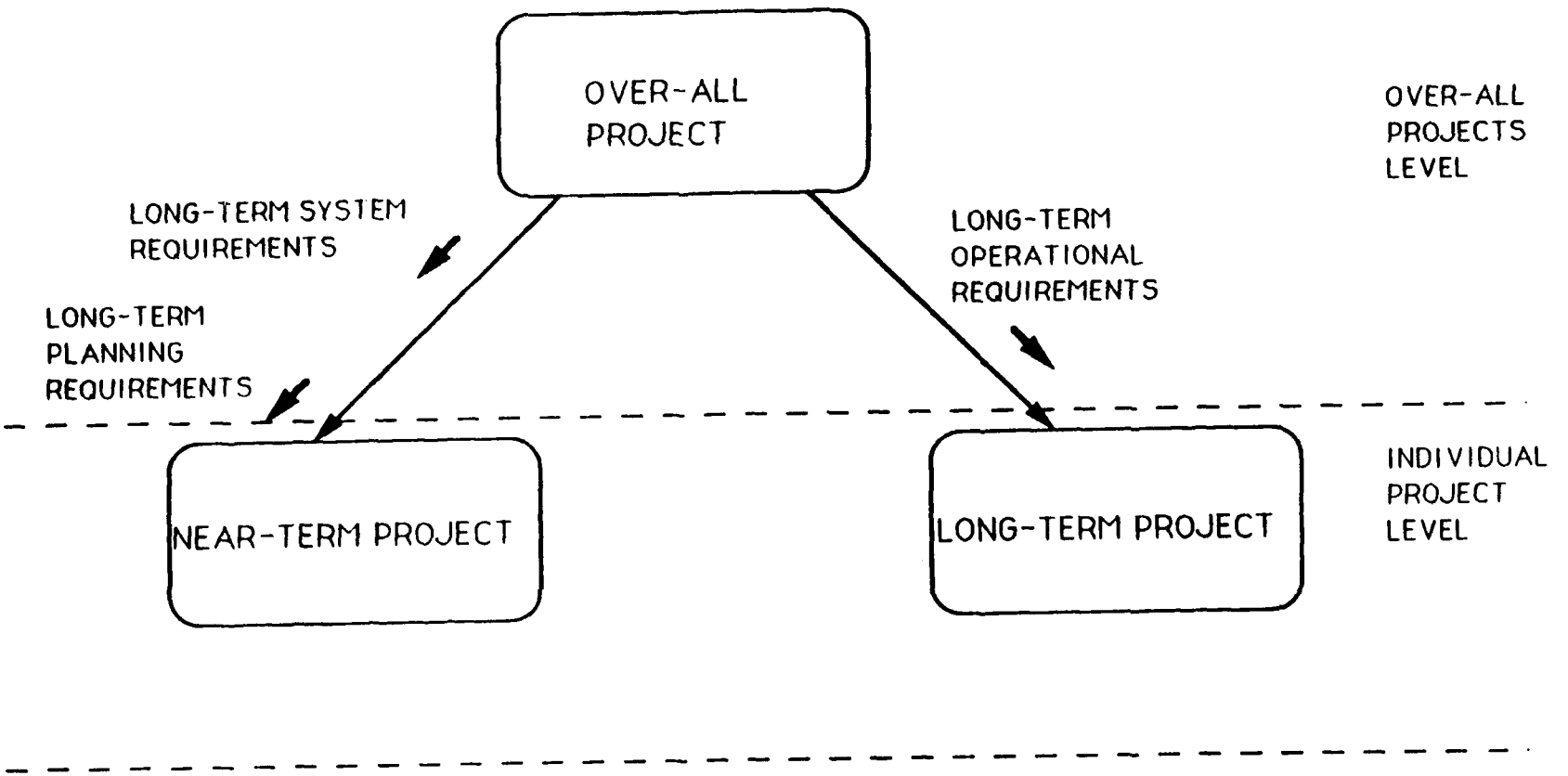
## 1.0 INTEL DATA PROCESSING

| IAA SUBCOMPONENTS / MAJOR REQUIREMENTS | AUTOMATIC DATA BASE GENERATION ADBG | INTEL-DATA PROCESSING UTILITIES | INTEL DATA ANALYSIS UTILITIES (e.g., CREATE GRAPHICAL DISPLAY) | MAN-MACHINE INTERFACE |
|---|---|---|---|---|
| REDUCE THE TIME ANALYSTS SPEND PROCESSING TEXT | X | | | X |
| REDUCE THE VOLUME OF TEXT ANALYSTS HAVE TO LOOK AT | X | X | X | |
| OPTIMIZE PRECISION (Ability to ONLY Find data on topics of interest) | X | X | | |
| OPTIMIZE RECALL (Ability to find ALL data on topics of interest) | X | X | | |

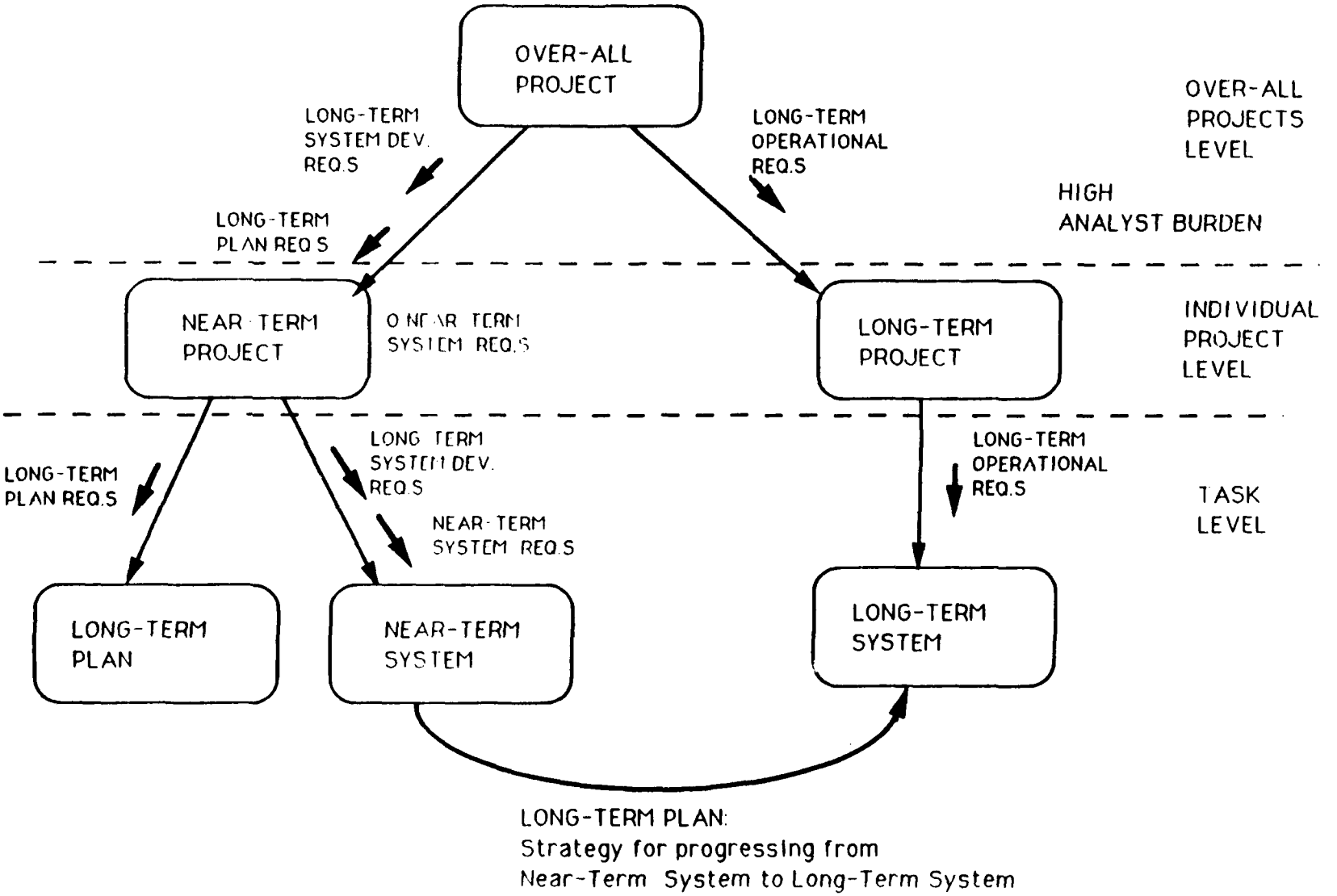
# Fig. 8: THE NEW METHOD for SOLVING the OPERATIONAL PROBLEM & HOW IT WILL MEET OUR OPERATIONAL REQUIREMENTS

| 2.0 INTEL DATA ANALYSIS | | | | |
|---|---|---|---|---|
| IAA SUBCOMPONENTS / MAJOR REQUIREMENTS | AUTOMATIC DATA BASE GENERATION ADBG | INTEL-DATA PROCESSING UTILITIES | INTEL DATA ANALYSIS UTILITIES (e.g., CREATE GRAPHICAL DISPLAY) | MAN-MACHINE INTERFACE |
| ENABLE the USE of APPLICATION SW & GRAPHICS to FACILITATE ANALYSIS | X | | | |
| FACILITATE ANALYSTS' ABILITY TO ABSORB MORE DATA, MORE QUICKLY | | | X | |
| REPRESENT DATA in a FORM that FACILITATES ANALYSIS | | | X | |
| FACILITATE ANALYSTS' ABILITY to MANIPULATE INTEL DATA | | | X | X |

# Fig.9 ALLOCATION OF LONG-TERM REQUIREMENTS TO INDIVIDUAL PROJECTS



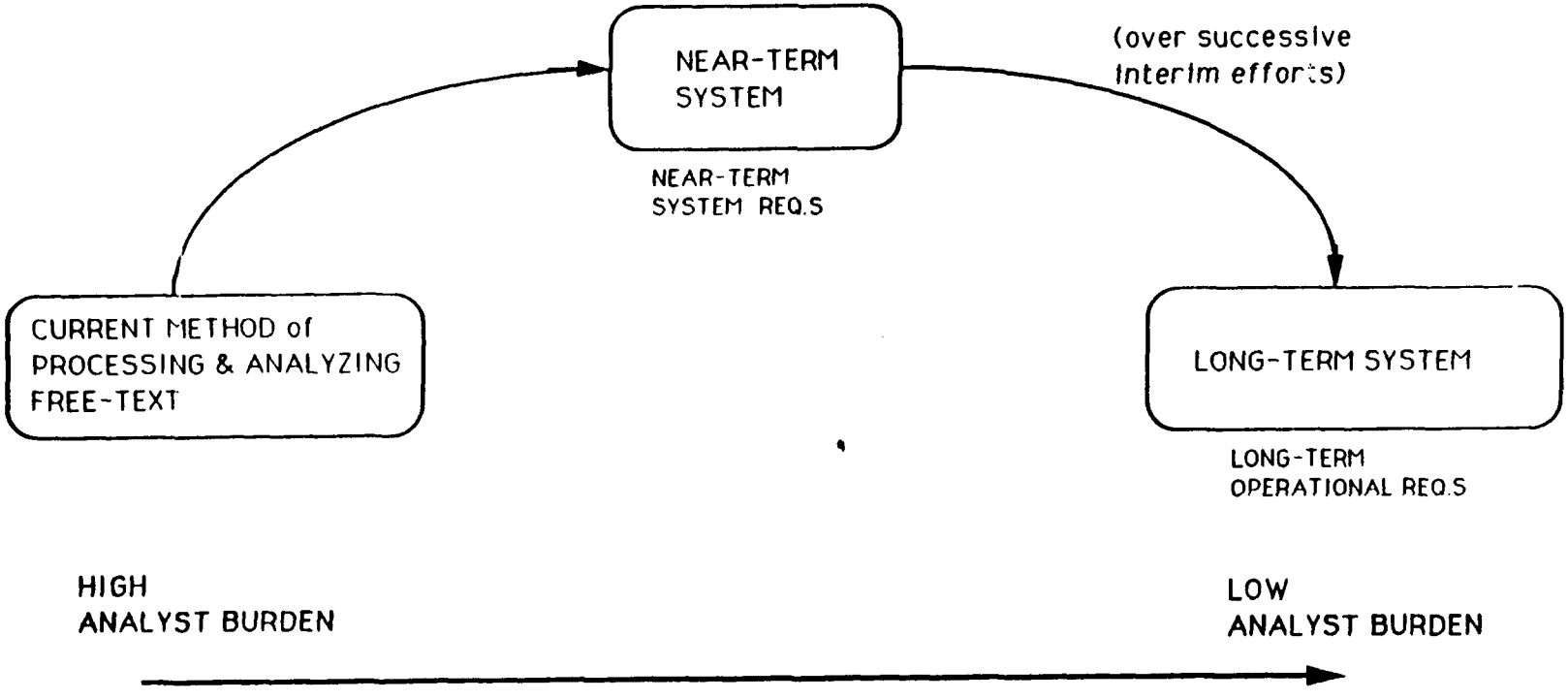
# Fig.10 ALLOCATION OF LONG-TERM REQUIREMENTS TO INDIVIDUAL TASKS




LONG-TERM PLAN:
Strategy for progressing from
Near-Term System to Long-Term System

# FIG 11 NEAR-TERM SYSTEM CONCEPT:

- a BRIDGE BETWEEN the CURRENT OPERATIONAL CAPABILITY and a system meeting the LONG-TERM OPERATIONAL REQUIREMENTS



## *MISSION*

## *OF*

## *ROME LABORATORY*

*Rome Laboratory plans and executes an interdisciplinary program in research, development, test, and technology transition in support of Air Force Command, Control, Communications and Intelligence ($C^3I$) activities for all Air Force platforms. It also executes selected acquisition programs in several areas of expertise. Technical and engineering support within areas of competence is provided to ESD Program Offices (POs) and other ESD elements to perform effective acquisition of $C^3I$ systems. In addition, Rome Laboratory's technology supports other AFSC Product Divisions, the Air Force user community, and other DOD and non-DOD agencies. Rome Laboratory maintains technical competence and research programs in areas including, but not limited to, communications, command and control, battle management, intelligence information processing, computational sciences and software producibility, wide area surveillance/sensors, signal processing, solid state sciences, photonics, electromagnetic technology, superconductivity, and electronic reliability/maintainability and testability.*