CATALYST

CATALYST: A Concept for an Integrated Computing Environment for Analysis

SW 89-10052
October 1989

CATALYST: A Concept for an Integrated Computing Environment for Analysis

This paper was prepared by Diane Q. Webb, Office of Scientific and Weapons Research. Comments and queries are welcome, and may be directed to the Chief, CATALYST Development Staff, on 874-0400 or secure 70423.

October 1989

Scope Note

This paper builds upon previous thinking within the Intelligence Community about the use of computers in intelligence analysis. It suggests an overall framework in which many different computer tools can be evaluated and compared. The paper discusses the concept of an integrated system of computer tools for assisting the all-source analyst, particularly the analyst who must cope with large amounts of text. Many of the individual ideas presented in this paper have been discussed at length in the Intelligence Community, but we believe that this paper provides a unique framework for the application and integration of these ideas.

The concept described in this paper grew out of the CATALYST (Computer-Aided Tools for Analysis of Science and Technology) project in the Office of Scientific and Weapons Research. The project is aimed at developing tools to assist in the analysis of foreign science and technology in the hope that past successes in computer-aided signals analysis can be repeated in the more complex area of all-source technical analysis. The CATALYST concept has been useful to OSWR in helping to plan and prioritize new investments in computer-based tools. Other members of the Community briefed on the concept have indicated a potentially broader applicability across many analytic organizations. With that expectation, this paper is presented to help stimulate research and focus attention on a wide range of initiatives that developers of computer systems may provide to the analytic community.

Contractor use of the graphics and concepts discussed in this paper is generally encouraged, with the stipulation that attribution be made to the US Government.

CATALYST: A Concept for an Integrated Computing Environment for Analysis

Summary

To meet the intelligence demands of the next century, the Intelligence Community needs to develop powerful tools for extracting more analytic insight from increasing volumes of all-source data. ***The CATALYST (Computer-Aided Tools for Analysis of Science and Technology) concept provides a functional specification for the analytic computing environment that we believe will be needed.***

The foundation of the CATALYST concept is the Intelligence Community's longstanding objective of helping analysts identify critical information without having to read through impossibly large amounts of raw data. We propose a set of automated tools managed by the analyst to process both incoming and historical data — sorting, retrieving, comparing, and counting data; performing basic statistical analyses; and looking for trends, anomalies, and missing data — with sufficient built-in knowledge to operate on large amounts of data automatically.

These tools would broaden the information under the analyst's control. They would draw immediate attention to information or patterns of activity that the analyst has specified as having potentially high significance. Such tools are essential to enable analysts to focus their time on interpreting rather than searching for the best available data.

Enhancing overall analyst productivity requires a system of interconnected computer capabilities, so that efficiency gains in one part of the analytic process can be exploited in all other stages of analysis. The analytic computing environment proposed in this paper reflects the wide variety of real analytic activities. ***It consists of an integrated set of computer tools needed by the analyst at each step of the analysis process — from the review of computer-highlighted data and hypotheses, and the retrieval and analysis of additional data, to the production of finished intelligence.***

We propose a set of automated tools managed by the analyst to process both incoming and historical data . . .

The major elements of the CATALYST environment are reflected in the six levels of the computer-assisted analysis pyramid. . .

The major elements of the CATALYST environment are reflected in the six levels of the computer-assisted analysis pyramid (see figure):

- Automatic entry of information into online data bases — integrating text, signal, and image data from diverse sources — and making this data accessible in the computer environment
- Automatic highlighting of potential alerts, anomalies, trends, and complex relationships
- Computer tools to assist the analyst in evaluating data and patterns and in exploring alternative scenarios
- Organization of ideas and supporting data using tools that help the analyst put ideas and data into the framework of a finished analysis product
- Production of textual and multimedia reports and briefings
- Timely and efficient revision and coordination of finished intelligence.

Implementing CATALYST

The successful implementation of the CATALYST concept will require meeting significant challenges in hardware and software development, user interfaces, training, and computer support. The key challenge to implementing CATALYST is the development and integration of a wide range of software for each block of the pyramid. Although significant software development will be required, existing commercial products, if properly integrated and implemented, could fulfill many of the requirements of the analytic computing environment.

The analyst must be able to access and use all elements of the analytic computing environment from a single workstation. The concept of distributed computing is critical to the successful implementation of the analytic computing environment. Careful attention to user interfaces, as well as extensive online training — which suggests to the analyst what to do as well as how to do it — must be integral components of the CATALYST environment.

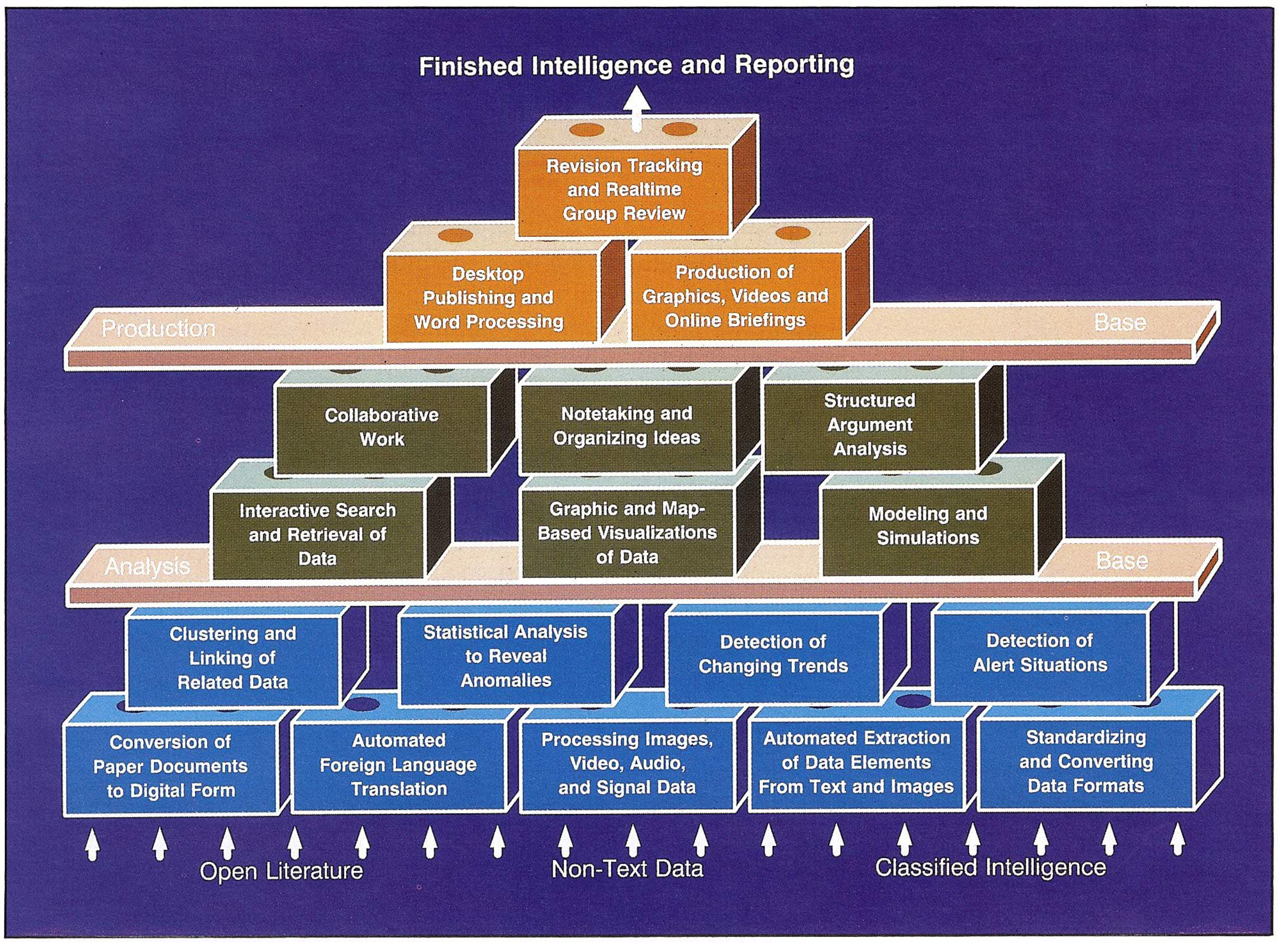


Figure ***The Key Blocks of the CATALYST Analytic Computing Environment***

The CATALYST environment can be visualized as a computer-assisted analysis pyramid in which interconnected computer tools are available at every step of the analysis process.

Finally, the effective use of powerful analytic tools will depend on a support organization that ensures reliable system performance and assists the analyst in the use and modification of a diverse and constantly changing set of tools.

Benefits of CATALYST to the Intelligence Community

The implementation of an integrated system of tools. . . would enable analysts to base their assessments on a far wider range of data. . .

The implementation of an integrated system of tools for computer-aided analysis would enable analysts to base their assessments on a far richer set of data. The ability to monitor massive volumes of all-source data would allow effective tracking of a much greater range of potential threats. ***By focusing analysts' efforts on the problems requiring intuition and judgment, the Intelligence Community could use existing analytic resources far more effectively.***

Furthermore, the benefits of an analytic computing environment would go beyond improving the analyst's use of data. An integrated analytic environment would enable production components to task collection systems more efficiently by allowing more of the existing data to be used in defining collection requirements.

The CATALYST environment also would provide an excellent means to preserve institutional memory. Analysis and computer-aided processing could be based not just on what today's analyst has read and can remember but on all the data and analysis contained in the system. New analysts could more easily build on the expertise of their predecessors. ***As a result, the Intelligence Community would be able to provide a significantly better product to policymakers.***

Contents

Introduction

Throughout the Intelligence Community, questions are being posed about potential next steps in the application of computers to help solve analytic problems. What do analysts need most from computers? Is there a common vision of what an effective system of computer-based analysis tools might look like? Should the many disparate computer projects in the Intelligence Community share a common underlying strategy, even though they deal with different kinds of analysis?

Some definite ideas about these issues have grown out of the exploration of advanced computer tools in one particular analytic environment, the Office of Scientific and Weapons Research (OSWR). Although these ideas were conceived in the context of scientific and technical analysis, we believe they may be extensible to many other areas of analysis.

The potential of computers to address the needs of analysis has been understood for years. Nevertheless, the Intelligence Community has had difficulty getting the most out of computer technology in the analytic environment. Analysts have been provided with useful word-processing and text-retrieval tools, but the anticipated dramatic gains in analyst capabilities have been elusive. ***We believe that the following factors have inhibited the development of effective analytic computing tools to date***:

- ***A focus on individual tools aimed at specialized aspects of the analytic process.*** Because analysis involves many different kinds of activities in a typical day, the analyst needs a wide-ranging system of interconnected computer tools to be effective. No single tool can substantially enhance overall analyst productivity by itself. Unless a tool acts as part of a system, much of the value of automation at any one stage of analysis is likely to be lost at adjacent steps in the analysis process.

- ***Insufficient or unsuccessful involvement of analysts in the development of computer tools.*** It has been hard to find developers and potential users of computer tools who can comprehend each others' capabilities, limitations, goals, and aspirations; yet, effective communication of this sort is essential.

CATALYST Terminology

In writing this paper, we have tried to minimize the use of technical terminology. Rather than using terms such as "software," "word processor," or "data base application," we talk generically of "tools," which are combined hardware and software applications provided by the computer to the analyst.

The use of the word "tool" is deliberately intended to provoke comparison to a hammer or screwdriver. Just as the right screwdriver allows one to remove a screw quickly and efficiently, so we believe that the proper computer tools will greatly improve aspects of the analyst's job.

To describe both the complete set of tools needed by analysts and the operation, support, and training aspects of making these tools effective, we use the word "environment." The analytic computing environment (or the CATALYST environment) therefore describes all aspects of the computer as it relates to improving analysis.

The Problem — Significant Data Often Goes Unseen or Unexploited

The Intelligence Community today faces a serious and widely recognized problem—too few analysts to effectively exploit the massive volumes of data being collected to address an ever-increasing number of intelligence issues. Analysts do not have enough time to read and assimilate each new day's collection, let alone compare the new data to large holdings from prior collection.

The volume of data relevant to wide-ranging issues concerning many nations will certainly continue to expand. ***There are simply not enough analysts to constantly monitor the data on each of these potential threats.***

- *Immature computer hardware and software technology.* We believe that, until recently, information processing technology has simply not been capable of supporting the development of a system of tools that could be readily exploited by most analysts.

Because of advances in technology and experiences in the Community, we believe that future analytic tool development can be markedly more successful than in the past. From our exploration of computer-based tools for scientific and technical analysis, we have developed a functional specification for a comprehensive analytic computing environment that can be achieved in the near term.

The foundation of this environment is a set of automated tools to operate on enormous sets of new and old data to provide the analyst tipoffs to potentially significant links, trends, and anomalies. In addition, we have identified a broad set of computer tools likely to be needed by the analyst throughout the analysis process — from retrieval of data to evaluation of computer-generated hypotheses to production of finished intelligence.

The Only Practical Solution — Computer-Based Analytic Tools

In principle, computer tools should be able to do what a limited number of analysts never will — constantly monitor the totality of incoming data and compare it with data received in the past. People have long envisioned systems of tools that would provide the analyst tipoffs to relationships in the data — trends, patterns, correlations, and anomalies — that could indicate important intelligence events. The major benefits of such an analytic environment would be:

- More effective monitoring of worldwide events using comprehensive, all-source data.
- Creation of an institutional base enabling the analyst to build on the experience and knowledge of other analysts.
- Ability to quickly begin assessment of new intelligence events from a base of online, preanalyzed data.
- Freeing the analyst to spend more time doing analysis — evaluation and prediction of events — rather than routine data management.

The concept described in this paper grew out of the CATALYST (Computer-Aided Tools for Analysis of Science and Technology) project in OSWR, which is aimed at developing tools to assist in the analysis of foreign science and technology (S&T). Although many of the individual concepts and ideas presented in this paper have been discussed at length in the Intelligence Community, we believe that this paper provides a unique framework for the application and integration of computer tools for all-source analysis.

A Plan for the Analytic Computing Environment

The Analytic Process Today

The intelligence process has four major components: collection, processing, analysis, and production. Clearly, computer-based tools can be applied to all aspects of this process. In this paper, we focus on the last three stages of the intelligence process, which require the intelligence analyst to manipulate large amounts of data. [1]

Today, the analyst is heavily involved in all three stages. In the processing stage, the analyst finds, reads, sorts, and files all the data that seem to pertain to his areas of responsibility. Having gathered data together, the analyst has already begun the analysis stage — noticing patterns, making hypotheses, and focusing additional collection. From this research, the analyst organizes his thoughts and supporting data into drafts of finished intelligence, which are typed, reviewed, and revised.

This way of doing analysis has its strengths and weaknesses. Its strongest point is that analysts are intimately familiar with each piece of data on which their analysis is based. The problem is that we are faced with far more incoming data than analysts can assimilate. Because analysts can read only a limited amount per day, their judgments necessarily reflect an increasingly limited sample of the total available data on a given issue. ***But what about the collected data that is never read? The analyst must hope that the unread data is consistent with the conclusions reached.***

How can computers help? Computer tools can improve the process whereby analysts choose the raw data on which to focus their attention. Because computers are good at doing rapid and repetitive processing on large volumes of data, modern automated tools could tirelessly process both incoming and historical data — sorting; retrieving; counting; performing basic statistical analyses; looking for trends, anomalies, and missing data — and alert the analyst only to items or patterns of potentially high significance.

The analyst need not personally review or remember each piece of data that might conceivably pertain to his research topic. Instead, he can examine a richer set of data — the information that the computer highlights in response to his instructions, plus additional data that the analyst calls for in the process of investigating the associated hypotheses.

[1] The analyst's involvement begins at the point where raw intelligence is handed off by the collection and processing elements to the analysis components and ends with the delivery of the analyst's coordinated intelligence product to the production and dissemination components. However, the analyst is continuously involved with all elements of the intelligence process by providing guidance and feedback to collectors and by maintaining close ties with consumers.

The Concept of an Analytic Computing Environment

The CATALYST concept is based on computer tools that operate unattended, looking for relationships and trends in huge amounts of raw, all-source data. As indicated above, this automated approach is essential for analysts to base their conclusions on a much broader and deeper set of data than they could ever read unaided. Automated tools are needed if intelligence consumers are to get the most value out of the raw data collected on any given issue.

The analytic computing environment should not be merely a collection of computer-based tools; rather, it should be a system of interconnected capabilities designed to aid the analysis process. Because the analysis process can be broadly conceived as winnowing large amounts of raw data into a focused finished intelligence product, ***we can best understand the CATALYST environment as a computer-assisted analysis pyramid, in which interconnected computer tools are available at every step of the analysis process*** (see figure 1).

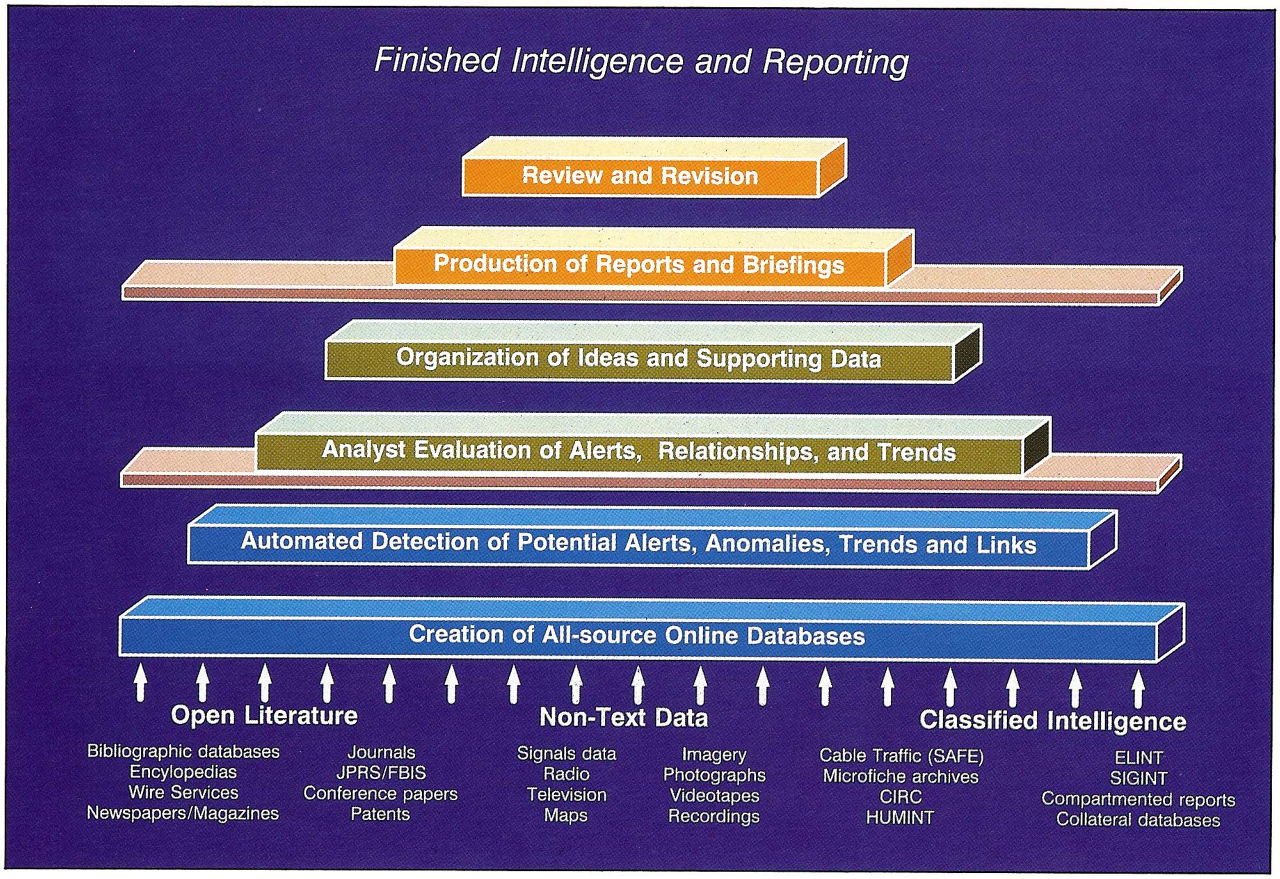


Figure 1 ***The Steps in the Computer-Assisted Analysis Pyramid***

Analysts need computer tools at each step in the analysis process, from data processing to analysis to the production of finished intelligence.

Included in the analytic computing environment would be tools for analysts to manipulate data and ideas, evaluate and test hypotheses, identify patterns in the data, and forecast future developments. The analyst must have direct and rapid access to any tool or piece of the data held in the system, including reference data bases, institutional knowledge, models, and simulations. ***Our experience indicates that the analyst needs such capabilities throughout the analytic process, from the outset of a research project to the final production of finished intelligence.***

The Steps of the Computer Assisted Analysis Pyramid

The Data Processing Steps

At the bottom of the pyramid, data flows in from a wide range of sources. Regardless of the source of the data, the ***first step is to automatically build integrated, all-source, online data bases*** on which computer tools can operate. Some of this data is already in electronic form, but much of the data exists only on paper. Clearly, computer tools can operate only on data that exists in the computer environment, so conversion is a key step for successful operation of subsequent tools.

The ***second step is the automated detection of potential alerts, anomalies, trends, and links.*** This step is at the heart of the analytic computing environment, as it provides the enhanced selection ability needed to cope with massive volumes of past and present all-source data. This step would consist of collections of computer "agents" (see box) that would search both incoming and historical data for indicators of potentially significant activity, such as a link between two previously unassociated persons or rising/falling activity at one of a thousand facilities. These agents, operating automatically, would constantly draw on a knowledge base to provide the analyst with tipoffs to potentially important intelligence events.

The first two steps in the pyramid create the analysis base, representing a large volume of data that has been automatically previewed and organized on behalf of the analyst. The analyst at this point has been involved only in guiding data collection and in making critical decisions about the raw data to be fed into the system. The analyst has also played a key role in defining the items and patterns of interest that guide the automated review of new and old data. But thus far, the analyst has not yet seen any of the individual reports.

The Analyst's Agents and Demons

In his recent book *The Society of Mind*, Marvin Minsky, a leading thinker in artificial intelligence, suggests that intelligent minds are made up of many smaller processes, which he calls "agents." Each of these agents carries out simple, nonintelligent functions; it is the sum of the activity of many agents that approaches intelligent behavior.

Minsky also talks of "demons," which are recognition agents that lurk silently in the background, watching for certain conditions and then carrying out predefined actions. We envision the smart tools in the second layer of the pyramid as being composed of such agents and demons — each carrying out simple statistical, rule-based or pattern-matching processes — the sum of which would begin to approximate the data monitoring capability of human analysts.

Will Analysts Accept the Concept of Automated Data Evaluation Tools?

Whether analysts would welcome such a change in approach depends critically on the tools provided at the third level. It seems clear, for example, that analysts will justifiably reject any system that does not provide direct access to the original source data. Analysts will also be concerned that an autonomous computer previewing and organizing the raw intelligence will fail to recognize subtle but significant leads that a human could exploit.

Nevertheless, we believe analysts will conclude that even imperfect tools operating on orders of magnitude more data will yield many more leads for an analyst to pursue. Furthermore, organizing these tipoffs along with all other pertinent information into a well-structured data base will increase the overall chances for analysts to find the more subtle leads. ***We expect the analyst's improved control over data to outweigh any concern over imperfect computer processing.***

The Analysis Steps

The analyst thus enters the process on a day-to-day basis at the ***third step, the evaluation of the potential alerts, links, trends, and anomalies*** that have been identified by the automated detection tools. At this point, the analyst would be presented with computer tipoffs to any developments defined to be of possible interest. To test the validity and significance of these computer-generated hypotheses, the analyst would have many tools at his disposal, such as interactive retrieval tools, graphical display and analysis capabilities, and modeling and simulation packages. These tools would allow the analyst to view data relationships from different points of view and to quickly find information related to the hypothesis under evaluation.

The fourth step involves the organization of ideas and supporting data. Here, the analyst organizes the results of analysis done in the third step into hypotheses and well-structured arguments, and then organizes these arguments into the outline of a finished intelligence product. At this step, the analyst would have tools to facilitate the organization of notes and data, outlining, and the incorporation of group work.

For the analyst, these two analysis steps represent the core of the analytic computing environment. Using the time previously spent trying to monitor and organize incoming data, the analyst devotes his expertise to the more abstract tasks of interpreting foreign developments, relating them to important intelligence issues, testing alternative hypotheses, and making predictions. The third and fourth steps together result in structured and tested analytic judgments which constitute the intelligence production base. At this point, an intelligence story is ready to be told.

The Production Steps

The fifth step is the production of reports and briefings. The analyst uses simple, flexible, and interactive word processing and graphics tools to present analytic judgments in a draft of finished intelligence. In the ***sixth and final step, online interactive review and revision tools*** assist the tracking and incorporation of the comments and changes of the multiple parties involved in coordination and production. This process is a source of great frustration to analysts today. Computer tools have potential for reducing the confusion and uncertainty in the production cycle, thus facilitating the difficult path towards consensus.

The Finished Intelligence Product

The final result is ***finished intelligence*** available for hard-copy or electronic dissemination. The finished intelligence product could take many forms: the traditional report, slides or viewgraphs for a briefing, videotape, or even a multimedia online presentation. We believe that, in the not-too-distant future, selected consumers will require online access to finished intelligence as a natural consequence of an increasing reliance upon near-real-time intelligence.

An on-line dissemination environment may at times entail the spreading of certain current intelligence responsibilities across shifts of analysts. Enormous importance will be placed on continuity of knowledge and a capability for rapid comparison of new data to previous information and judgments. Clearly, a CATALYST-like system would be essential in an era of electronic intelligence support to consumers.

The production of finished intelligence in an integrated computer environment would open great possibilities for improving consumer access to Community thinking. We could provide not only today's concise judgments, but, at the consumer's touch of a button, a more detailed presentation of the thinking and data underlying that finished intelligence. Eventually, electronic dissemination should facilitate more rapid and effective feedback between the analytic and policymaking communities.

The Building Blocks of the Computer-Assisted Analysis Pyramid

Figure 2 (see next page) illustrates the building blocks of computer tools that would be needed to build an effective analytic computing environment. The specific tools are described in more detail in the Appendix.

We recognize that the analytic environment presented in figure 2, with data flowing in at the bottom and finished intelligence flowing out at the top, is a simplified view of the analysis process. In the real process of analysis, which often entails following many different problems at the same time, the analyst will move back and forth between tools at different levels in the analytic process. For example, while putting the finishing touches on a report, the analyst may need to go back to the raw data, or a new piece of data might suddenly provide the missing element of an hypothesis the analyst was developing in the notetaking process.

The tools that make up the computer-assisted analysis pyramid must support the chaotic reality of analysis, allowing the analyst to jump back and forth between tools to follow up chains of thought as they occur. To support this, the analyst should be able to move data and ideas smoothly from one tool to another. Further, the analyst should be able to retrace his path — or follow the path of another analyst — through the system from the finished intelligence product, to the generation of the analyst's argument, to the computer-generated hypotheses, and finally back to the original data.

To be effective, an analytic computing environment must combine the individual tools at every level of the analysis process into a unified

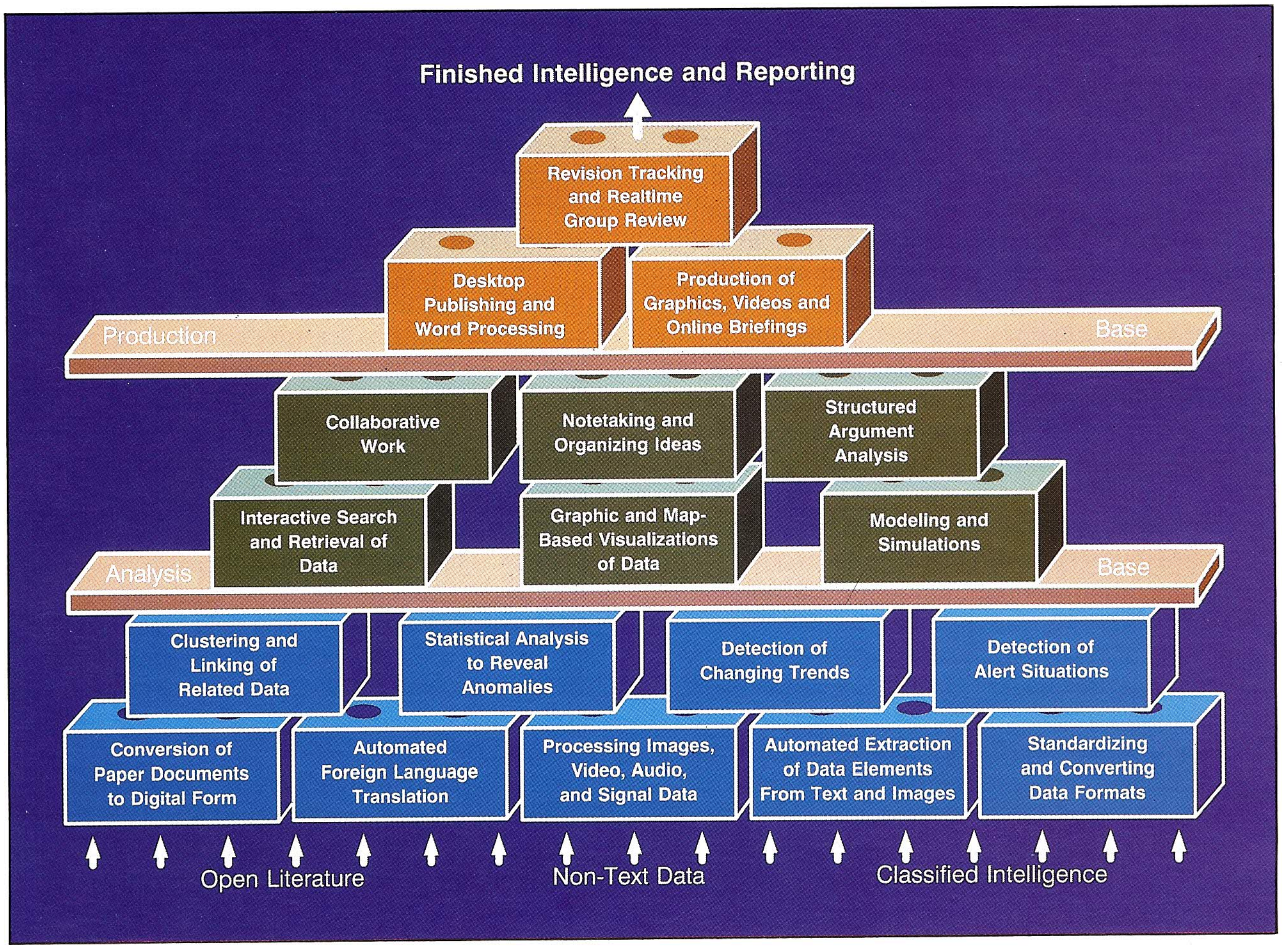


Figure 2 ***The Key Blocks of the CATALYST Analytic Computing Environment***

The CATALYST environment can be visualized as a computer-assisted analysis pyramid in which interconnected computer tools are available at every step of the analysis process.

system. This is because the analytic process consists of many varied, interdependent activities. We believe that the lack of a systems approach has contributed to the disappointing results of several past computer efforts — the projects were technically successful but there was no set of surrounding capabilities to make the new tool fully useful to analysts.

For example, disconnected projects that build online data bases contribute to a strong foundation for analysis, but we need to build on this foundation by providing automated tools that guide the analyst to significant pieces of data. Likewise, stand-alone word processors facilitate intelligence drafting, but they lack a strong foundation of data and analysis capabilities. ***To make analysts more productive, developers of the analytic computing environment must provide a wide range of computer capabilities.***

Is CATALYST Useful Beyond S&T Analysis?

How extensible is this concept to areas beyond S&T analysis? ***Although the CATALYST concept was originally conceived for long-term S&T analysis problems, we suspect it is applicable to virtually all types of analysis.*** Clearly, the particular computer tools depicted in figure 2 will differ for different types of analysis; additional capabilities will be required, and some of the tools useful for S&T analysis may not be as useful for other applications.

We believe that all analysts must be able to process incoming information, compare it with what is already known, formulate and test hypotheses, and render judgments in a finished intelligence format. In the case of long-term strategic analysis, this process might take months; in the case of tactical analysis, it might occur in minutes. Indeed, analysts responsible for quick-response analysis would appear to be in even greater need of tools to help them process information rapidly and efficiently.

Keys to Implementing the Analytic Computing Environment

S***uccessful implementation of the CATALYST concept requires new initiatives in computer hardware and software development, user interfaces, training, and computer support.*** Significant challenges in each of these areas must be seriously addressed if analysts are to be provided an effective analytic computing environment.

The key challenge to implementing CATALYST is the development and integration of a wide range of software for each block of the pyramid. Although significant software development will be required (see box), existing commercial products, if properly integrated and implemented, could fulfill many of the requirements of the analytic computing environment.

Much greater analyst involvement in software development and modification is needed. Although the analytic computing environment will have components shared by large groups of analysts, the exact set of tools useful to any individual analyst will be heavily shaped by his own intelligence assignment. Clearly, an analyst covering Third World agriculture requires some different computer tools than the analyst tracking changes in weapons design would use.

Given a basic set of tools, the analyst should be able to select, design, and customize a complete system of computer-based tools appropriate for his analytic problem. Providing this essential flexibility will certainly require the replacement of today's software interfaces (which are designed for developers) with more straightforward interfaces suitable for the typical analyst.

Requirements for the CATALYST Workstation

One consequence of the need for a system of interconnected tools is that they all must operate from one analyst workstation. Analysts cannot benefit from a rich set of synergistic tools unless they can access them easily and flexibly in a single integrated environment, preferably a single workstation on each analyst's desk. We anticipate that this analyst workstation would have the following characteristics:

- Extremely fast processors and large internal memory.
- Multitasking and multiprocessing.
- Multiwindow graphical user interface.
- Large-screen, high-resolution color and/or grayscale display with the ability to rapidly process multimedia graphics.
- Ability to interface with a variety of computer systems, including mainframes, workstations, and personal computers.
- Ability to add special-purpose processors (for example, signal processing and image manipulation).
- Ultra high bandwidth data communication paths.

Although it is impossible to anticipate the exact requirements of the future analyst workstation, it is clear that current personal computer technology falls short of the demands of the analytic

Selected Technologies Warranting Further R&D

A wide range of computer technologies are needed to construct the six levels of a computer-assisted analysis pyramid. Although many of these technologies have reached at least an early prototype stage of development, considerable improvement appears to be realizable. Technologies that warrant further R&D include the following:

- Optical character recognition, particularly for typeset text
- Natural language understanding (for both English and foreign languages)
- Interactive expert systems and smart data bases
- Parallel processing, neural networks, and pattern recognition techniques
- Object-oriented databases and programming languages
- Hypermedia and hypertext (the storage and manipulation of multimedia data as a network of nodes connected by links)
- Graphical presentation techniques, including computer simulation, animation, and geographic information systems
- Automated image recognition, image processing, and 3-D image transformation

computing environment. The computer required to implement the CATALYST concept will probably more closely resemble today's advanced multipurpose engineering workstations.

Distributed Computing

We do not propose that all the complex processing involved in the analytic computing environment take place on the analyst's workstation. The data processing tools represented by the lower layers of the computer-assisted analysis pyramid could clearly operate on supercomputers or massively parallel processors networked to the analyst's workstation. Indeed, we foresee the networking of powerful special-purpose processors, analyst workstations, peripheral storage, and other devices as central to the analytic computing environment. ***The concept of distributed computing is critical to the successful implementation of the CATALYST concept.***

Making Tools Easier to Learn and Use — Interfaces and Training

Another key to successfully implementing the CATALYST concept will be careful attention to the user interface. Today, analysts are overwhelmed by a plethora of diverse and difficult-to-use software products. Given the complex range of capabilities we hope to make available to the analyst, it will be critical that all tools be straightforward and easy to use. Adopting a visual and user-friendly interface that is consistent from one tool to another is essential to allow the analyst to focus on using the computer-based tools rather than on learning them.

A more difficult task will be to guide analysts to tools and techniques with which they are not familiar. ***Online training*** — which suggests to the analyst what to do as well as how to do it — should be an integral component of the analytic computing environment. One approach is to gradually introduce new tools to the analyst so that expertise is acquired in a logical and phased sequence.

We envision a system that would remember successful strategies used by analysts and suggest additional tools and techniques the analyst might explore. Or, if an analyst failed to apply certain tools, or used them inefficiently, the system might offer an online tutorial or refer the analyst to colleagues who are already using those tools well.

Supporting the CATALYST Environment

The complexity and scope of the analytic computing environment will also require a new approach to computer support. The effective use of powerful analytic tools will depend on a support organization that ensures the reliable performance of the analytic computing environment and assists the analyst in the use and modification of a diverse and constantly changing set of tools.

Computer support personnel will need to work closely with analysts so that the unique requirements of each organization can be understood and met. ***The diversity of tools required for analysts covering different intelligence problems and the need for distributed computing resources will compel a distributed approach to computer support as well.***

Potential Benefits to the Intelligence Community

The implementation of a complete set of tools for computer-aided analysis would enable analysts to base their assessments on a far wider range of data. Such analysis would more closely approach the oft-mentioned goal of all-source analysis, enabling intelligence analysts to base their conclusions on a more complete set of data from both classified and open sources.

The ability to monitor massive volumes of all-source data would allow the Intelligence Community to effectively track a much greater range

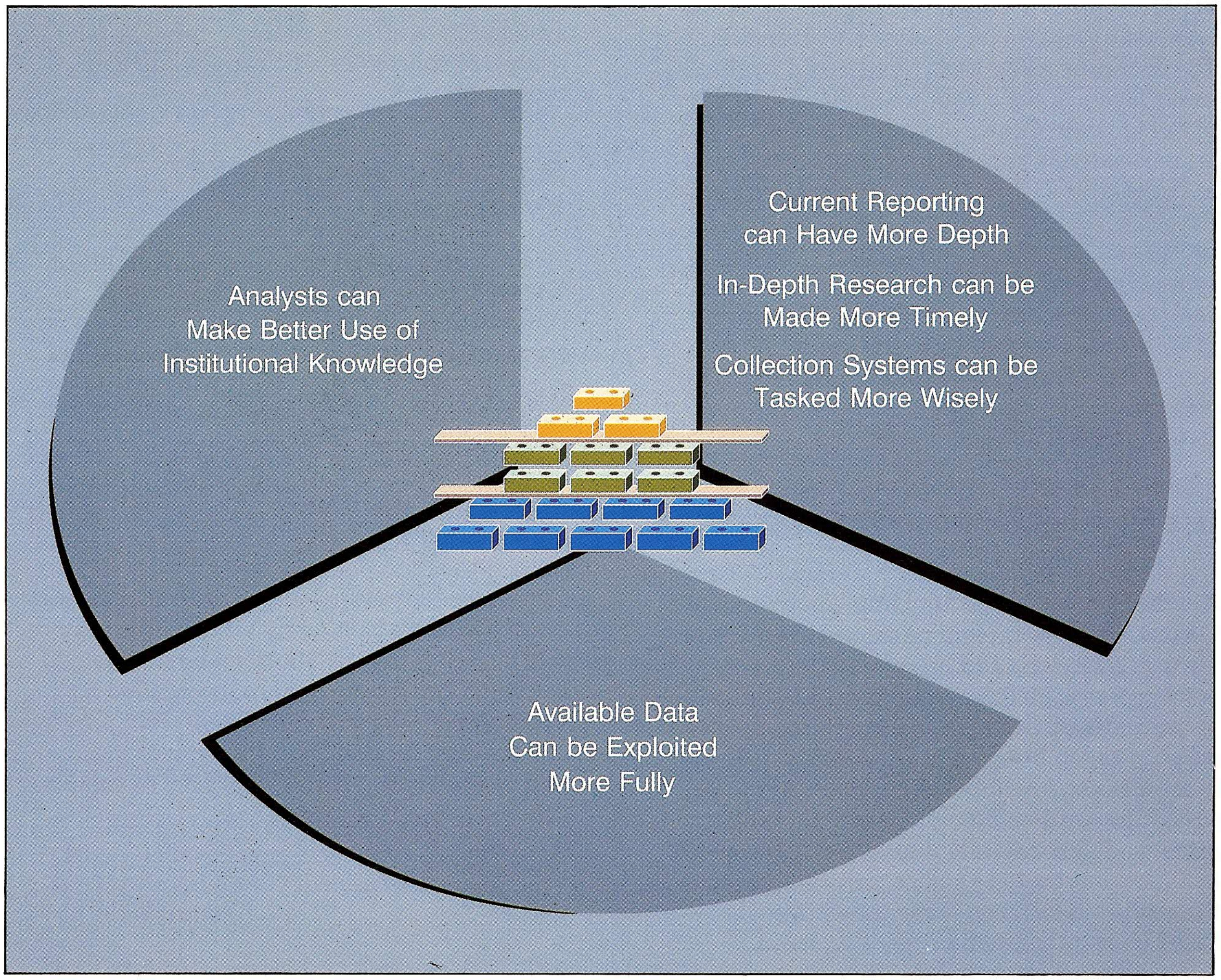


Figure 3 ***Benefits of a Modern Analytic Computing Environment***

By covering a wider range of potential threats more thoroughly, the Intelligence Community would be able to provide policymakers with a significantly better product in terms of both quantity and quality.

of potential threats. By focusing analysts' efforts on the problems requiring intuition and judgment, the Intelligence Community could task existing analytic resources far more effectively.

The benefits of an analytic computing environment would go beyond improving analysts' use of data (see figure 3). In particular, an integrated analytic computing environment would enable analytic components to more efficiently guide future data collection. With data from many sources being used more effectively, analysts could focus their tasking more precisely.

In addition, the analytic computing environment could be exploited to determine how data was actually being used by the analysts. Demonstrably valuable data sources could be thus targeted for increased exploitation. Identifying poorly utilized data sources could also lead to improved exploitation or perhaps to a determination that particular data sources have limited intelligence value. ***The net result would be a better coupling between collection and analysis, and a better allocation of scarce collection resources.***

The CATALYST environment would also provide an excellent means to preserve institutional memory. Analysis and computer-aided processing could be based not just on what the analyst has read and can remember but on all the data and analysis contained in the system. Not only could raw data be stored where analysts could rapidly access it, but a record of analytic judgments and supporting evidence would be available to successive generations of analysts. New analysts could more easily build on the expertise of their predecessors rather than having to laboriously rebuild data collections and personal knowledge.

The benefits of the CATALYST approach, already needed in today's analytic environment, will become even more important as the Intelligence Community evolves toward providing real-time intelligence to key consumers on line. . .

The CATALYST Program

The Office of Scientific and Weapons Research (OSWR), in cooperation with the Office of Information Resources (OIR) and the Office of Research and Development (ORD), is working to develop key elements of the CATALYST environment. Some projects currently underway include:

- Pinnacle — Digital conversion of paper documents and full text search and retrieval (OSWR)
- ATLAS — All Source Trends and Linkages Analysis System (OSWR)
- ALERT — Extraction of data elements from text, natural language processing (OIR)
- CATALYST Hypermedia — Notetaking and organization of ideas (ORD)
- BEACON — clustering of related data items using co-citation analysis (OSWR)

of potential threats. By focusing analysts' efforts on the problems requiring intuition and judgment, the Intelligence Community could task existing analytic resources far more effectively.

The benefits of an analytic computing environment would go beyond improving analysts' use of data (see figure 3). In particular, an integrated analytic computing environment would enable analytic components to more efficiently guide future data collection. With data from many sources being used more effectively, analysts could focus their tasking more precisely.

In addition, the analytic computing environment could be exploited to determine how data was actually being used by the analysts. Demonstrably valuable data sources could be thus targeted for increased exploitation. Identifying poorly utilized data sources could also lead to improved exploitation or perhaps to a determination that particular data sources have limited intelligence value. ***The net result would be a better coupling between collection and analysis, and a better allocation of scarce collection resources.***

The CATALYST environment would also provide an excellent means to preserve institutional memory. Analysis and computer-aided processing could be based not just on what the analyst has read and can remember but on all the data and analysis contained in the system. Not only could raw data be stored where analysts could rapidly access it, but a record of analytic judgments and supporting evidence would be available to successive generations of analysts. New analysts could more easily build on the expertise of their predecessors rather than having to laboriously rebuild data collections and personal knowledge.

The benefits of the CATALYST approach, already needed in today's analytic environment, will become even more important as the Intelligence Community evolves toward providing real-time intelligence to key consumers on line. . .

The CATALYST Program

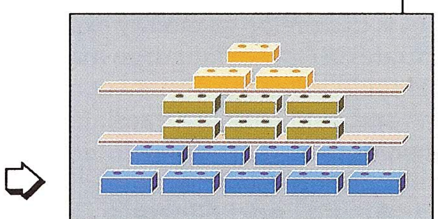
The Office of Scientific and Weapons Research (OSWR), in cooperation with the Office of Information Resources (OIR) and the Office of Research and Development (ORD), is working to develop key elements of the CATALYST environment. Some projects currently underway include:

- Pinnacle — Digital conversion of paper documents and full text search and retrieval (OSWR)
- ATLAS — All Source Trends and Linkages Analysis System (OSWR)
- ALERT — Extraction of data elements from text, natural language processing (OIR)
- CATALYST Hypermedia — Notetaking and organization of ideas (ORD)
- BEACON — clustering of related data items using co-citation analysis (OSWR)

Appendix

Some of the Computer Tools Needed To Implement CATALYST

This appendix is intended to provide illustrative examples of some of the key capabilities required for the implementation of the CATALYST system. Each of the sections below corresponds to a block in the CATALYST pyramid (see figure 2). This discussion is not intended to definitively describe all analysts' requirements or the technical approaches best suited to meeting those requirements. Rather it is intended to stimulate thinking and discussion among both analysts and information planners.



Building Online Databases

Conversion of Paper Documents to Digital Form

The conversion of nonelectronic information — both incoming and historical files — into digital form is clearly one of the major components of the computer-assisted analysis pyramid. We have identified two major types of conversion requirements:

- The conversion of massive historical files (such as 20 years of key technical publications)

- The day-to-day conversion of small amounts of information that the analyst finds important to his assigned tasks. The first type of conversion should obviously be done with analysts' encouragement but not actual involvement. For the second type, however, the analyst needs a mechanism for getting important information into the computer environment quickly (in hours, not days).

Existing technology to address this problem includes optical character readers (OCRs), such as those developed by the DEST, Kurzweil, and Calera corporations. These scanners convert printed and sometimes even handwritten text into digital representations. However, the existing technology has several major limitations, including inability to scan foreign character sets, difficulty in dealing with equations and figures, and the need for significant amounts of human quality control following the recognition process. At present, the high cost and low throughput of OCRs makes scanning large volumes of text prohibitive. There is a critical need for faster, more accurate OCRs, as well as for automated context analyzers that correct some of the OCR errors.

Automated Foreign Language Translation

Merely getting data into the electronic environment is not sufficient; it must be present in a useful form. Much of the information relevant to intelligence issues exists only in foreign languages, and the analytic computing environment must incorporate strategies for dealing with such material. The Intelligence Community has undertaken several efforts to develop automated foreign language translation capabilities, which produce either rough "gisting" of documents or more polished translations. Clearly, the Intelligence Community should continue to strongly support efforts to translate documents from one language to another.

An important parallel requirement is to develop or modify expert systems and statistical tools to operate on foreign language material as easily as on English. Such tools could select the documents of maximal importance to the analyst and automatically translate only the documents that the analyst chooses to read.

Processing Images, Video, Audio, and Signals Data

Nontextual information — including imagery, photographs, line drawings and diagrams, audio sources, television, and videotapes — is an important source of intelligence information. The challenge is to integrate this information with textual sources so that the analyst can make effective use of both. Existing technology allows the relatively easy digitizing of images, which can be linked to text through hypermedia approaches. Increasingly, computer workstations are providing the ability to deal with high-resolution color and grayscale graphical images, as well as incorporating digitized sound and live video capabilities. A major problem today, however, is meeting the massive data storage and high-bandwidth communications requirements of these very information intensive data sources.

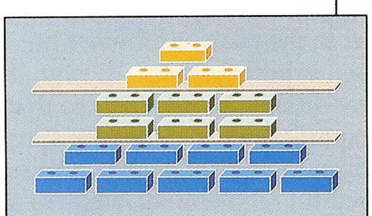
Automated Extraction of Data Elements from Text and Images

Another key step in getting data into a useful format is to automatically recognize and extract potentially important data elements from full text and images. Data embedded in superfluous prose (text) or surrounding scenery (images) is rarely useful; it must be put into a standard format so that it can be compared and contrasted with other structured data. In the case of text, this requires the ability to recognize data elements such as the names of people, facilities, programs, pieces of equipment, and weapons systems along with the context in which they appear in a document. The data elements can then be stored in a structured form (such as a relational data base). This step is an automated version of the labor-intensive manual data entry procedures in use today.

The automated extraction of data from images and text requires significant advances in the fields of natural language understanding and automated image recognition (two subfields of artificial intelligence). In addition to the traditional artificial intelligence techniques being explored today, new technology in parallel processing and neural nets may be applicable to these problems. A further requirement will be the construction of end-user interfaces, which allow the analyst to instruct and modify the automated extraction process.

Standardizing and Converting Data Formats

Although the Intelligence Community should continue to encourage the adoption of standard formats for storing all sorts of data, many data sources will continue to be created outside of the Intelligence Community. For the foreseeable future, important data will exist in many incompatible electronic formats. To effectively use data from many different sources, the Intelligence Community will need to develop a capability to rapidly convert data from one format into another.



Automated Detection of Potential Alerts, Anomalies, Trends, and Links

Clustering and Linking of Related Data

One key to identifying significant data relationships is grouping together, or clustering, related pieces of data. This might involve identifying all items dealing with a particular scientific research topic or those about a certain country. Once data clusters have been identified, the computer can alert the analyst to clusters with unusual characteristics or to data that do not fit into standard clusters.

Several techniques have been developed for clustering full-text information, including co-citation and co-author clustering, co-term analysis, N-grams, and more traditional statistical clustering methods. Given the massive amounts of data that intelligence problems often involve, it is particularly important to develop techniques for rapidly clustering very large volumes of data.

Statistical Analysis To Reveal Anomalies

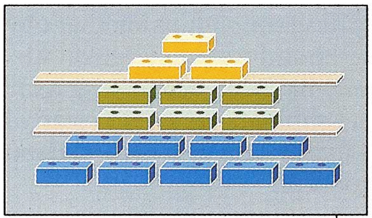
Given the extraction of structured data items (such as names, places, and project names), traditional statistical analysis techniques can be powerfully applied to revealing unusual patterns in the data. Data elements that do not fit preexisting data patterns may indicate phenomena of interest to the analyst. For example, the sudden appearance of research activity at a previously unknown facility may indicate the appearance of a new research program.

Detection of Changing Trends

Statistical techniques can also be applied to spotting unusual trends in data. By watching rising and falling levels of activity in a particular area, for example, the analyst could detect changing priorities and shifts of emphasis. These statistical techniques could range from simple counting methods to more sophisticated statistical analyses. As with other tools in this level, the most effective tools will be those that provide interfaces allowing the analyst to easily modify the statistical analyses and how they are presented to the analysts.

Detection of Alert Situations

The detection of events of interest to the analyst provides an excellent application for rule-based expert systems. Ideally, these expert systems would be easily modifiable by analysts so that they could interactively instruct the expert systems to look for specific events, or combinations of events, in the data. These expert systems could operate on top of powerful relational data bases, which would be populated by the tools needed to build online data bases.



Analyst Evaluation of Alerts, Relationships, and Trends

Interactive Search and Retrieval of Related Data

It is essential that the analyst be able to access any data in the CATALYST system (including raw data) rapidly and easily. This requirement can be addressed with today's state-of-art tools for full-text and structured data base search and retrieval. The key requirements for such interactive retrieval systems are rapid response against very large data bases and retrieval techniques that are transparent to the analyst. Existing technology to search full-text data bases includes preindexed text search (implemented in SAFE and the OSWR Pinnacle system) and hardware-based pattern matching systems such as the Fast Data Finder. Natural language interfaces that allow the analyst to query data bases with simple English sentences and icon-based or graphical query systems have great potential to facilitate rapid and interactive analyst retrievals.

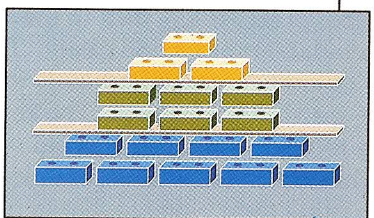
Graphic and Map-Based Visualizations of Data

Pictorial or graphical representations of data often have more meaning to humans than do the same data presented in traditional text or tabular forms. Such graphical representations of trends or relationships can help the analyst to recognize subtle patterns in data. Geographic displays are particularly useful for many intelligence applications because the analyst is often interested in following activity in a particular country or region. Effective use of geographic displays requires sophisticated geographic information systems that are based on either digital or analog maps. Graphical displays also provide an effective user interface, allowing the analyst to move between different levels of detail in data structures.

Modeling and Simulations

Particular aspects of the analysis process lend themselves well to modeling and simulation techniques. Models and simulations allow the analyst to determine the significance and impact of new data to their understanding of an intelligence problem, as well as providing ways to test "what if" scenarios. If the analyst can define a fairly well-behaved structure or process (for example, the predicted performance of a weapons system), new and historical data — even hypothetical data — can be plugged into the model to assess their implications.

The need for modeling and simulation constitutes another obvious requirement for artificial intelligence and expert systems. Although the Intelligence Community has been developing related technologies for years, models and simulations would clearly be even more effective if embedded in a system of tools and data as represented by the CATALYST concept.



Organization of Ideas and Supporting Data

Notetaking and Organization of Ideas

One of the most needed elements of the analytic computing environment is a tool that allows the analyst to capture and organize ideas and supporting data. One look at the existing paper files, boxes of index cards, and notebooks that analysts keep today reveals both the need and the enormous potential for a flexible analyst notetaking environment.

We believe that this notetaking requirement is closely met by the "hypertext/hypermedia" concept (the storage and retrieval of multimedia data with complex data relationships and connections), which has recently become popular in the computer community. Xerox's NoteCards and Apple's Hypercard are examples of existing hypermedia products.

A hypermedia system could allow the analyst to create and modify a complex network of data and hypotheses and the relationships between them. The analyst could store ideas and the data relating to them in packages that could be organized at any time into the more structured arguments that would underlie a finished intelligence product. Furthermore, the analyst could build alternative scenarios, or frameworks, and could reject or accept them as future information suggested. Such a tool might also allow a more focused communication of ideas among analysts and managers concerning the status and substance of proposed research papers.

The requirements for a notetaking environment are:

- The analyst must be able to quickly and easily incorporate information in the system into analyst notes.
- The notetaker must provide traceable links back to original source documents.
- The analyst must be able to quickly visualize the structure of the network of notes and reorganize it if necessary.

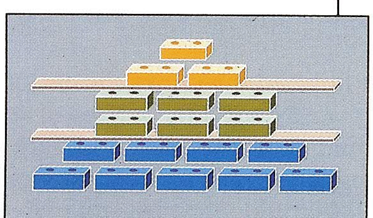
- The notetaking environment must be closely integrated with the finished intelligence production tools at the next level of the CATALYST pyramid.

Collaborative Work

An analyst's ideas and related data should be embedded in an environment where this information can be easily shared with other analysts. Many intelligence applications require that analysts work together, with each analyst contributing some piece of the solution or with several analysts providing different points of view. At present, much time is wasted in meetings, mailing drafts back and forth, and otherwise transmitting information that then must be incorporated into each person's work file. Providing capabilities for groups of analysts to work with a shared set of data and ideas, with each person's contribution clearly visible, would greatly assist collaborative work efforts.

Structured Argument Analysis

Because the analyst is ultimately responsible for constructing a logical and supportable argument, analysis could benefit from tools that apply formal logic and reasoning techniques to the arguments the analyst has constructed. Such techniques could clarify the reasoning the analyst has used to reach a conclusion; point out where the analyst lacked supporting data, or perhaps had found contradictory data; and enable the analyst to rapidly explore the impact of new data on analytic conclusions or test alternative scenarios. Research is already under way at the Xerox Palo Alto Research Center to apply formal reasoning techniques in the NoteCards environment.



Production of Reports and Briefings

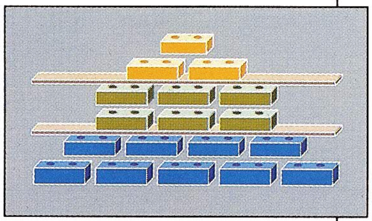
Desktop Publishing and Word Processing

The Intelligence Community has, like many other computer users, recognized the value of word processing and desktop publishing tools in the creation of high-quality finished intelligence products. Prior to the availability of these tools, analysts had to write drafts that were then typed by a secretary. Changes were difficult to make, as they typically required retyping the entire draft. By contrast, a flexible and easy-to-use word processing system allows the analyst to quickly create draft reports and easily modify these drafts, giving the analyst more control over the creation of his finished intelligence product. Desktop publishing tools provide the additional capability (when appropriate) of rapidly producing high-quality finished products. Advanced desktop publishing packages allow the analyst to use attractive layout techniques to mix text and graphics to communicate his message to the reader.

Although a variety of desktop publishing tools are already in use, we believe these tools would prove even more useful when embedded in a fuller set of capabilities in the computer-assisted analysis pyramid. For example, rather than creating merely a document, the finished intelligence product could retain links to all the ideas and data that supported it. Such finished intelligence packages would be far more valuable (particularly to successive analysts) than the disembodied intelligence reports produced today.

Production of Graphics, Videos, and Online Briefings

In addition to finished intelligence reports, the analyst frequently presents analytic judgments in the form of multimedia briefings. Tools are badly needed that would enable the analyst to quickly produce colorful, informative briefing aids showing the results of analysis. Such tools would include flexible and powerful graphics editors and the ability to send system displays to high-quality printers, film recorders, and videotape. In addition, the system should include a capability to create computer-based briefings that could be disseminated and displayed on line.



Review and Revision

Revision Tracking and Real-Time Group Review

The coordination and revision of finished intelligence is one of the key aspects of the analysis process and one badly in need of assistance. We envision tools that would allow the analyst to track and incorporate the comments and suggested revisions of many reviewers quickly and flexibly. These tools could allow several reviewers to simultaneously edit a single document rather than relying on the time-consuming, back-and-forth process used today. The analyst could instruct the system to incorporate a reviewer's suggested changes, review the impact of the changes, and reverse them if necessary. Revised documents would clearly indicate the source of revisions and retain the editing history of the document. A reviewer could tag particular suggested revisions as critical, so that both he and the analyst would have to deal with the specified issue before the document could proceed through the process. These revision and review tools, although certainly not eliminating the friction of the coordination process, could certainly streamline it and relieve the analyst of much of the tedious burden of revision.