Top Ten Needs for Intelligence Analysis Tool Development

Richard V. Badalamente and Frank L. Greitzer
Battelle—Pacific Northwest Division
P.O. Box 999
Richland, WA, 99352, USA
richard.badalamente@pnl.gov *frank.greitzer@pnl.gov*

Keywords: Fusion, Information Sharing/Collaboration, Usability/Habitability.

Abstract

A workshop was held to generate ideas about future enhancements to software systems designed to aid intelligence analysts in the analysis process. Workshop participants were working analysts, working members of the Pacific Northwest National Laboratory Field Intelligence Element. An analyst from the Counter-Intelligence Program also participated in the workshop. This paper summarizes the results of the workshop and discusses implications for intelligence analysis software tool development.

1. Introduction and Background

Intelligence analysis (IA) professionals are faced with daily challenges to meet high demands for rapid, accurate assessments that require discovery and marshalling of evidence, integration and synthesis of data from disparate sources, interpreting and evaluating information that is constantly changing, and providing documentation and recommendations (intelligence products) that meet the customer's reporting requirements. Ongoing research efforts seek technology-based solutions to reduce the analyst's workload and improve the throughput and quality of IA products. Some of the motivation for tool development is based on informal "corporate knowledge and experience." For example, a particularly influential book by Heuer (1999) provides such a perspective, informed by principles of psychology and human information processing. There is a limited amount of empirical research to guide tool development (e.g., Patterson, Roth, and Woods, 2001), and there is a growing body of work based on cognitive engineering efforts such as interviews, observations, and cognitive task analysis (e.g., Patterson, Woods, Tinapple, and Roth, 2001; Hutchins, Pirolli, and Card, in press). There is a continuing need to organize, motivate, and better define requirements in order to ensure that research and development products provide effective solutions to critical problems.

The Pacific Northwest National Laboratory (PNNL) has for some time been involved in both the development of software tools meant to aid in intelligence analysis (IA), and in actually performing intelligence analysis as part of the Department of Energy's Field Intelligence Element (FIE). In an attempt to better inform the research and development (R&D) of IA tools, we conducted a workshop with intelligence analysts to generate ideas about how to enhance software systems designed to aid in the analysis process. The purpose of this paper is to report on the results of that workshop and to discuss implications for future IA tool development and/or enhancement.

2. Approach

We conducted a one-day workshop with analysts employed at PNNL. Most of the analysts worked in the PNNL FIE and, for the most part, performed a class of analysis known as scientific and technical intelligence. One of the analysts worked in the PNNL counterintelligence (CI) program and conducted CI and counter terrorism (CT) analysis.

The workshop, which was conducted in July 2002, started at 8:00 in the morning and lasted 5 and 1/2 hours. A facilitator guided workshop proceedings. Two human factors specialists participated as catalysts and scribes. Nine working analysts participated in the workshop, eight from the PNNL FIE, and one from the PNNL CI Program.

The workshop commenced with a presentation of an Intelligence Analysis Process Model developed as a part of a PNNL internal R&D Project (Figure 1). This multistage process model was developed from interactions we had from practicing intelligence analysts in open source operations as well as from the analysts working in the PNNL FIE.

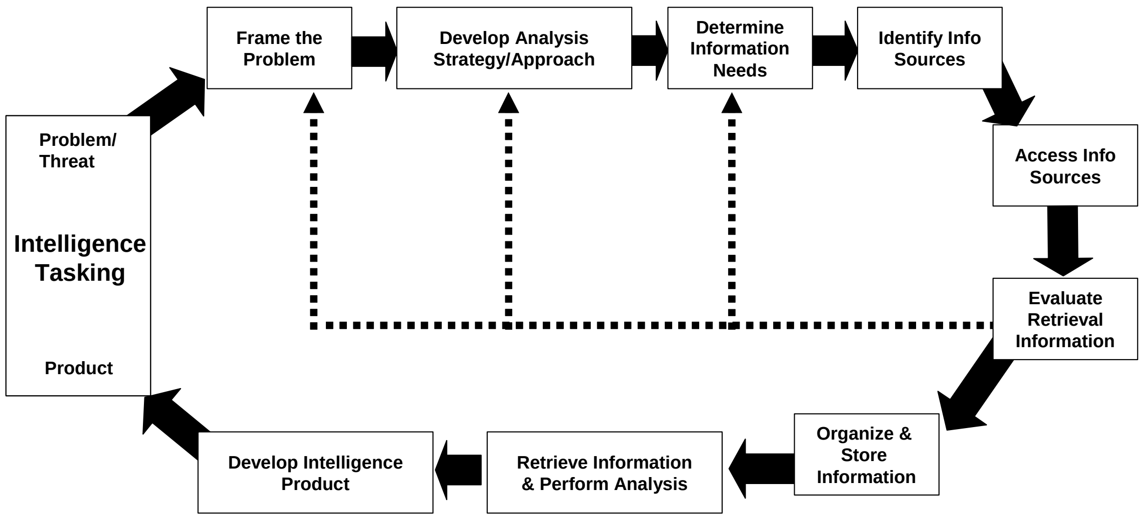

Figure 1. Intelligence Analysis Process Model used to provide a background and common reference point to the analysts participating in the workshop.

The analysts involved in our workshop perform functions in all the steps illustrated in Figure 1. This diagram therefore reflects the process from the perspective of the analysts we interviewed. It also served our research purposes well as a hypothetical construct to guide our attempts at capturing data from analysts across the full spectrum of their analysis functions.

Our discussions with analysts reinforced our own experience with regard to the task of describing and understanding a process that is essentially cognitive in nature. That is, such processes are difficult to describe in words or diagrams. The analyst's thought processes as described to us are much more fluid, fuzzy, and opaque (even to the analysts) then our diagram would imply. We have attempted to illustrate the iterative and decidedly nonlinear nature of the process with the dotted lines from "Evaluate Retrieved Information" to "Frame the Problem," "Develop Analysis Strategy/Approach," and "Determine Information Needs," as well as the feedback loop embedded in the overall process. However, even this diagrammatic feature is inadequate to fully convey the fuzzy or non-sequential nature of the process.

After the presentation, participants broke out into two groups with instructions to use the model as a point of departure for identifying future software system developments/enhancements that would aid analysts in doing their job. The groups were given 2 hours to do this. Upon completion, they reassembled and each group briefed their findings. The participants then voted on the top ten developments and/or enhancements they would like to see occur.

3. Results

The two groups approached their tasks very differently. Group 1 followed the steps in the model and attempted to identify enhancements related to each step. In this sense, they took a systems or programmatic approach. Group 2 focused immediately on data processing facets of the intelligence analysis task and the kinds of automation support they felt would aid them in their job. Despite this, there was a great deal of commonality in the enhancements identified. The top ten enhancements identified during the workshop are described below and listed in Table 1, in the priority order decided upon by the two groups (highest priority first).

Seamless Data Access and Ingest [#1]

Analysts want the ability to access data seamlessly across multiple databases and bring (ingest) the merged data into whatever analysis tool (software application) they may be using. The term "seamless" refers to the ability of the user to move from one database to another without logging in and out of the multiple databases, or having to learn different query/retrieval routines and different graphical user interfaces (GUIs). The data ingest process needs to be transparent to the analyst/user. Invariably, software applications require that data be in a particular format before it can be processed by the application—e.g., Starlight (a visualization tool developed by PNNL) requires that text be converted from ACSI to XML. Whatever data conversions are required in order for an analyst to employ data of interest need to be accomplished in the background (much like the printer driver and buffer work) and not require analyst intervention – the process needs to be "transparent" to the analyst.

Table 1. Top Ten System Enhancements

#	Enhancement
1	Seamless Data Access and Ingest
2	Diverse Data Ingest and Fusion
3	Shared Electronic Folders for Collaborative Analysis
4	Hypothesis Generation & Tracking
5	Template for Analysis Strategy
6	Electronic Skills Inventory
7	Dynamic Data Processing and Visualization
8	Intelligent Tutor for Intelligence Product Development
9	Imagery Data Resources
10	Intelligence Analysis Knowledge Base

Diverse Data Ingest and Fusion [#2]
Analysts want the ability to access and work effectively on their computer desktop with a wide variety of data types: text, photographs, satellite images, maps/ / geographic information, signals, measurements and signatures, etc. They want software tools with the capability of treating the various types of data as "information units" so that data types can be "fused" in ways that facilitate synthesis.

Shared Electronic Folders for Collaborative Analysis [#3]
This enhancement would allow analysts to share information gathered during the course of multiple, wide ranging, but potentially related analysis projects. Meta data tagged to each file, standardized file formats, and a common ontology, and dynamic updating and notification would greatly facilitate collaboration of analysts and reduce redundant searches and analyses.

Hypothesis Generation and Tracking [#4]
Analysts believe that they would benefit greatly from a software tool that essentially "coached" them in the formulation and testing of hypotheses generated during analysis of important issues requiring the careful weighing of alternative explanations or key judgments. Since their hypotheses change and evolve as additional information is added, analysts also want the tool to keep track of what they've hypothesized and what new information may have relevance.

Template for Analysis Strategy [#5]
Analysts want a software tool that coaches them in the development of a strategy or plan of attack for conducting more complex, longer-term analyses. Such a tool might provide a template that covers the various facets of analysis and provides automated linkages to potential methods and sources.

Electronic Skills Inventory [#6]
Often, there is no easy way for analysts to identify staff members who have the skills, experience, and clearances needed to assist in the analysis task. Access to some form of electronic skills inventory would aid the analyst in quickly identifying staff and assembling an analysis task team. This sort of tool would be a critical component of an effective collaboration environment.

Dynamic Data Processing and Visualization [#7]
Analysts want the ability to deal effectively with data streams as opposed to static data sets. As analysts develop ways to better comprehend the data space through user-controlled visualizations, they want their visualizations automatically refreshed with new data on a schedule dictated by the analyst.

Intelligent Tutor for Intelligence Product Development [#8]
Analysts are taught to develop their products and present their findings in a very structured manner [described in the CIA publication "Analytical Thinking & Presentation for Intelligence Producers" (FOUO)]. The keys to the process are sound analytic thinking and defensible judgments. Analysts feel that they would benefit from software that somehow instantiated the analytic process in the software tools used to produce the intelligence product. More than a template, such a tool would actually test the analyst's logic and comment on the defensibility of his/her conclusions.

Imagery Data Resources [#9]
Imagery is growing in importance as a source of data for intelligence analysis. Image data requires vast amounts of electronic storage and significant processing before "finished imagery" is presented to the analyst for use. Capabilities are needed to further compress imagery data in order to stretch limited storage space. In addition, tools are needed that assist the imagery/photo analyst to categorize, tag, organize, and process imagery data.

Intelligence Analysis Knowledge Base [#10]
Analysts have the sense that other analysts in the organization have addressed many of the problems they are addressing in one form or another, but their findings, or what they've learned in the process, are not always fully documented. Analysts feel that they would benefit from the development *and maintenance* of a "knowledge base" that not only captured results of prior analyses, but also

the knowledge gained by other analysts as they pursued their analysis goals and developed their products. This represents another ingredient of an effective collaboration support environment

4. Discussion

First, we point out an obvious limitation in the current study: the results are constrained by the fact that a limited number of analysts participated, all from the same organization, and nearly all engaged in one form of IA (all-source analysis focusing on science and technology). Nevertheless, we believe that findings obtained in this study are instructive for guiding further investment in research and development by and for the Intelligence Community (IC). Ideas and findings of this study are entirely consistent with observations from other representatives of the IC—such as the application of psychological science to intelligence analysis in general by Heuer (1999), writings of senior analysts and practitioners in the field (e.g., Bodnar, 2003; Grabo, 2002), and discussions that we have had with analysts in other domains (e.g., cybersecurity and information security).

It is not surprising that the analysts in the present study assigned the highest priorities to the need for seamless data access/ingest (#1) and the ability to ingest diverse data types (#2). Our interviews with analysts, both in the group studied and reported here as well as with other analysts within our laboratory and elsewhere points to this need. A great deal of time is required to prepare data for ingest, and this often requires specialized knowledge (such as familiarity with XML and computer science expertise), which is not part of the repertoire of a typical analyst. Tools are needed to support the ingest and fusion functions. Until such tools are developed or suitable functionality is added to existing ingest software, IA departments should consider the feasibility of employing "data ingest technicians" who can perform such functions in a support role, in much the same way that system administration staff provide support for computers and network communications.

The notions of shared electronic folders (#3) for analysis, templates for analysis strategies (#5), electronic skills inventories (#6), and an intelligence analysis knowledge base (#10) are examples of enhancements that would support work groups and collaborative work. By and large, information technology tools that have been developed or that are under development focus on the individual analyst rather than on teams of analysts working collaboratively. These ideas most likely represent only a portion of the whole picture that could emerge from a more focused study of requirements for collaborative intelligence analysis.

The need for support for hypothesis generation and tracking (#4) is not at all surprising. This relates to well-documented human information processing limitations that may be addressed in part by employing certain techniques designed to overcome such limitations, such as the analysis of competing hypotheses (e.g., Heuer, 1999; Jones, 1998). The Advanced Research and Development Activity's (ARDA) Novel Intelligence from Massive Data (NIMD) program has directed significant resources among the R&D community toward developing computer-based tools for supporting hypothesis generation and tracking. The current results provide additional justification for this effort.

Dynamic data processing and visualization (#7) and imagery data resources (#9) derive from needs for more effective visualization solutions and more dynamic ways of representing knowledge that is evolving. Solutions are needed that transcend what is typically described as "visualization" – in contrast to a predominantly "passive" relationship between the system that displays complex visualizations and the analyst who still must digest and interpret them. What is needed is a much more interactive and dynamic relationship in which the analyst is better able to explore the information within the visualization. With imagery data, the problem is more basic: solutions are needed to process imagery data and provide context/relationships or filtering functions that can reduce the load on the analyst and direct attention to salient features in the massive data.

The need for training is evident in enhancement #8, intelligent tutor for intelligence product development. It has been observed that a considerable amount of time is devoted to training new analysts on the production and structuring of intelligence analysis products. Requirements for such products tend to vary with the client. An important training requirement that receives too little attention is the need to train analysts on the analysis process. The IC needs to examine training needs with the aim of developing training solutions that will reduce the time required to train new analysts (largely a lengthy on-the-job process now) and that will establish more formal, rigorous practices for conducting analyses. The following quote provides a similar argument for devoting significantly more attention to training on the analysis process:

> For many years, the intelligence collection process has best been described as the task of trying to collect *everything* with the hope of finding *something*. This accounts in part for the sheer volume of information being gathered by the many agencies in the Intelligence Community. But throwing massive amounts of information at an intelligence analysis problem will

> not, by itself, solve this problem. What is so frequently left out of the equation is the process by which the information is analyzed….
>
> Any intelligence analysis task involves three major ingredients that must be generated or discovered by an analyst: hypotheses (possible explanations, predictions, or conclusions), evidence, and arguments linking evidence and hypotheses. (Hughes and Schum, 2003).

Frank Hughes, who teaches at the Joint Military Intelligence Center, provides training on a critical component of the analysis process, the process of evidence marshalling and inference. More training such as this—on the process in addition to the policies/procedures—is needed to raise the level of experience and expertise of the IC. Online training and blended training approaches (combinations of classroom and online training) can offer some effective solutions for training of the types of complex cognitive processes that are required.

5. Conclusions

The complexity, uncertainty, and ambiguity with which the analyst must deal in reaching judgments about future actions or events will likely remain beyond the capabilities of software tools for some time. Sherman Kent (1965) said, "… there can never be a time when the thoughtful man can be supplanted as the intelligence device supreme." Higher cognitive tasks (e.g., pattern recognition and what might be called "contextual awareness") are challenging to automate. Major programs supporting the IC, such as ARDA's NIMD and Advanced Question Answering for Intelligence (AQUAINT) programs, are aimed at addressing ambitious needs such as these. At the same time, however, many of the more routine tasks currently required of analysts in collecting, organizing, storing, and retrieving the data used in their analyses are amenable to automation; a number of software tools already aid in these aspects of the IA process. Somewhere in between are tasks required of the analyst for which automation can provide some relief, but which have somehow not yet been addressed. Many of the ideas reflected in Table 1 are examples of such enhancements.

Figure 2 illustrates some speculation in this regard. The figure provides, at a very rough level of approximation, a picture of various IT needs, characterized in terms of their potential impact on IA process (as reflected in the analysts' rankings) and their relative complexity from a design, development and/or deployment perspective. The needs identified in the present paper are shown in the cloud in the central area of the figure, reflecting their middle-ground positions with respect to their presumed impact and IT complexity. To help "anchor" the ten enhancements discussed here, we added (on the right side) some examples of more advanced enhancements such as pattern recognition systems or advanced question-

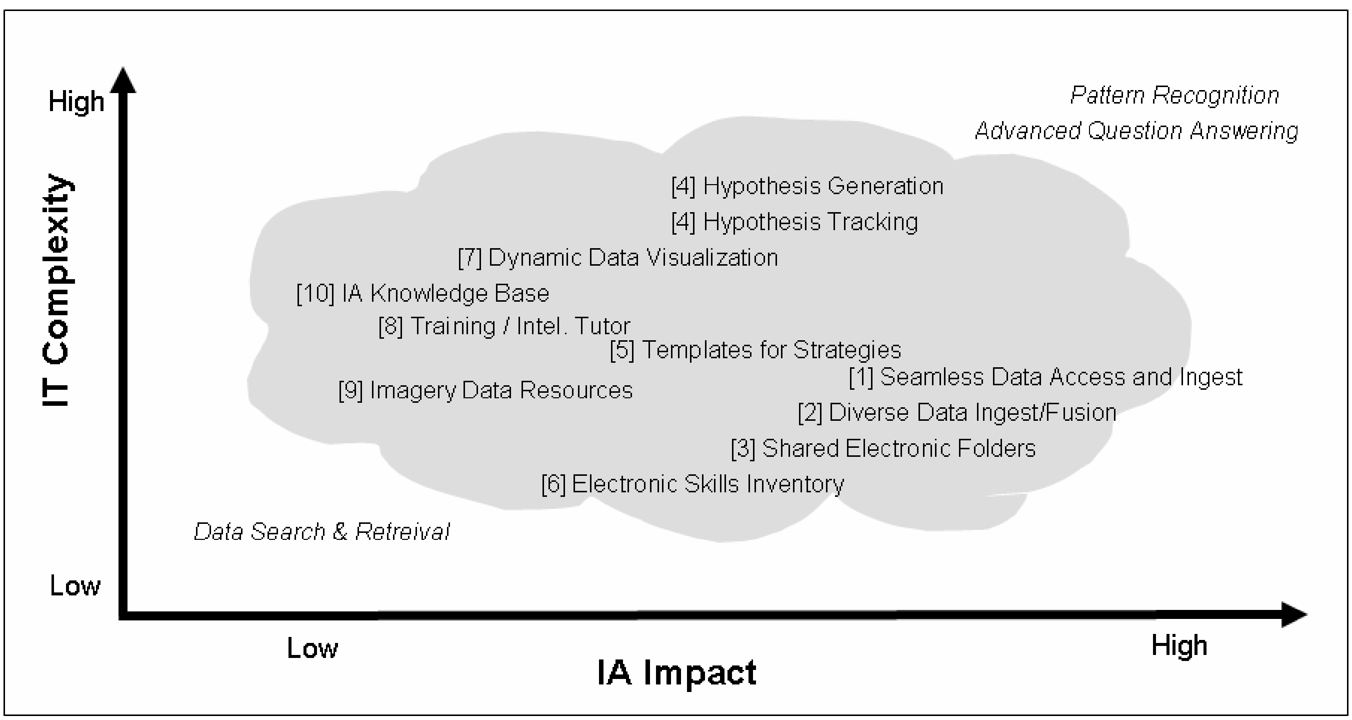

Figure 2. Illustrative Diagram Speculating On Relative Impact and IT Complexity of Top Ten Enhancements.

answering systems. Similarly, on the left side (and also located at the low end of the IT complexity scale), we included "data search and retrieval" systems, exemplified by some commercial search engines and other special purpose retrieval systems currently in use.

The enhancements shown in Figure 2 are represented in very rough relative positions, and the specific positions are certainly subject to some debate. The point we wish to make is that a more detailed exploration of these considerations, with additional consideration of possible development and lifecycle costs, may lead to the identification of "low hanging fruit" that can be "plucked" to achieve a more immediate return on investment.

6. References

Bodnar, J. W. 2003. *Warning Analysis for the Information Age: Rethinking the Intelligence Process.* Washington, D.C.: Joint Military Intelligence College's Center for Strategic Intelligence Research.

Grabo, C. M. 2002. *Anticipating Surprise: Analysis for Strategic Warning.* Ed. Jan Goldman. Washington, D.C.: Joint Military Intelligence College's Center for Strategic Intelligence Research.

Heuer, R. J. 19991. *Psychology of intelligence analysis.* Washington, D.C.: Central Intelligence Agency, Center for the Study of Intelligence.

Hughes, F. J., and Schum, D.A. (2003) Preparing for the future of intelligence analysis: Discovery—Proof—Choice. Unpublished Manuscript, Joint Military Intelligence College.

Hutchins, SG, P Pirolli, and SK Card (in press). What makes intelligence analysis difficult? A cognitive task analysis of intelligence analysts. In R. Hoffman (Ed.), *Expertise out of Context*, Hillsdale, NJ: Lawrence Erlbaum.

Kent, S. 1965. *Strategic Intelligence for American World Policy*. CIA: The Sherman Kent School for Intelligence Analysis.

Patterson, E. S., Roth, E. M., and Woods, D. D. 2001. Predicting Vulnerabilities in Computer-Supported Inferential Analysis Under Data Overload. *Cognition, Technology & Work, 3*, 224-237.

Patterson, E., S., Woods, D. D., Tinapple, D., and Roth, E. M. 2001. Using cognitive task analysis (CTA) to seed design concepts for intelligence analysts under data overload. In *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting.* Santa Monica, CA: HFES.