2014

Development and Application of an Analyst Process Model for a Search Task Scenario

Hendrickson K. Karl
Wright State University

DEVELOPMENT AND APPLICATION OF AN ANALYST PROCESS MODEL FOR A SEARCH TASK SCENARIO

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Engineering

By

Karl K. Hendrickson
B.S., Wright State University, 2012

2014
Wright State University

WRIGHT STATE UNIVERSITY

GRADUATE SCHOOL

14 May 2014

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Karl K. Hendrickson ENTITLED Development and Application of an Analyst Process Model for a Search Task Scenario BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science in Engineering

____
Mary E. Fendley, Ph.D
Thesis Director

____
Thomas N. Hangartner, Ph.D
Department Chair

Committee on
Final Examination

____
Mary E. Fendley, Ph.D

____
Subhashini Ganapathy, Ph.D

____
Yan Liu, Ph.D

____
Robert E.W. Fyffe, Ph.D
Vice President for Research and
Dean of the Graduate School

ABSTRACT

Hendrickson, Karl, K. M.S.Egr., Department of Biomedical, Industrial and Human Factors Engineering, Wright State University, 2014. *Development and Application of an Analyst Process Model for a Search Task Scenario*.

A key intelligence analyst role in open source search is the transformation of data into understanding. Better comprehension is needed of how new tools impact the analyst search process. The use of function analysis, heuristic analysis, and a usability study combine to provide the basis for developing an analyst process model, which affords the researcher with a structure to measure the impact of tools and expertise in performing a search task. The experiment utilized representative analyst scenario tasks in comparing baseline tools with the Geospatial Open Search Toolkit (GOST). The results show error rates increase when using a new toolset due to unfamiliarity with system affordances. Lack of toolset familiarity impacted participant output and time on task breakdown. Opportunities exist both for additional novice process training as well as more time for experts to acclimatize to new toolsets.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

ACKNOWLEDGEMENTS

This research was supported, in part, under the following Radiance Technologies contract:

HUMAN PERFORMANCE EVALUATION

For: GEOINT OPEN SOURCE TOOL (GOST) PHASE II

Radiance Contract No. FA8650-10-C-6113

Funding was also provided under the following Wright State University contract:

Neuroscience and Medical Imaging

Analyst Test Bed Contract No. FA8650-11-C-6157

The staff at Radiance Technologies provided invaluable support and this study would not have been possible without their help. I would also like to thank the staff at the Advanced Technical Intelligence Center (ATIC) in Beavercreek, OH, where the experiment was conducted, for their support throughout. I thank my committee advisors, Dr. Subashini Ganapathy and Dr. Yan Liu, along with everyone in the Wright State University College of Engineering and Computer Science who provided support throughout the study. In particular, I want to thank my thesis advisor, Dr. Mary Fendley, for her knowledge, guidance, and sustained optimism. Also, Bev Grundin, P.Stat., at the Wright State University Statistical Consulting Center provided valuable assistance in the data analysis. Finally, I could not have accomplished this without the support and encouragement of my family.

I. INTRODUCTION

The role of an intelligence analyst (IA) is to sift through large amounts of data to make quick, accurate assessments regarding the relevancy of available data through a process of search and retrieval, integration, and synthesis. A key IA role in open source search is the transformation of data into understanding. A better comprehension is needed of how new tools impact the analyst search process. The model developed through this research provides insight into the analyst process as well as a structure for inserting metrics which allow both the study of the process as well as the toolset being used. This allows for testing of toolsets as well as process developments.

Creating a mental model of an analyst search process requires sufficient background to provide context. This includes information about the intelligence analysts to understand their skills and job requirements. Analysts search for information and manipulate raw data into a coherent end product through a process of data transformation. As in the case of studying new tools such as the Geospatial Open Search Toolkit (GOST), it is important to be cognizant of the issues surrounding software development. Both the GOST system and the existing analyst tools are fundamentally decision support systems which allow the analyst to draw conclusions about the relevance of data being assessed. Investigating the role of the analyst in the context of this environment allows us to develop a model of the cognitive process. In turn, this allows us to insert appropriate metrics to measure the effectiveness, efficiency and ease of use of the system being studied.

The use of a function analysis, heuristic analysis, and a usability study combine to provide the basis for developing an analyst process model. The model is designed to afford the researcher with a structure for conducting an experiment to measure the impact of tools and expertise in performing a search task.

The goal of this research is to utilize representative analyst scenario tasks in comparing baseline tools with the Geospatial Open Search Toolkit (GOST). The study also compares the impact of expertise in order to assess the GOST system and provide a basis for developing appropriate training tools for novice analysts.

1.1 Overview and Problem Description

Analysts are inundated with data that needs to be assessed or analyzed in a short period of time. Tools are being developed to aid in the analysts tasking but are not always evaluated from a human factors perspective, and testing with real end users is not always possible during the development of these tools. Due to the varied experience levels of the users we will be looking at not only testing the new tool, but also understanding the impact on user groups that the tool aims to aid in task performance.

1.2 Research Questions

The research effort seeks to answer the following questions:

1. What are performance differences between expert and novice?
2. What are performance differences between systems, i.e., baseline and GOST?
3. Can a model be developed and validated that reflects the analyst search process?
4. Does the model provide an accurate description of the role of both human and system?

The results from question three lend insight into rationale for the differences that are found in questions one and two. Additional questions for discussion include determining the validity of measures of cognitive workload and measures of performance.

1.3 Research Objectives

The user evaluation objectives are: (1) Exercise the application under semi-controlled test conditions with representative users, (2) Establish baseline user performance and user-satisfaction levels of the user interface for future development and evaluation, (3) Develop and validate a model representative of the analyst search process, (4) Evaluate cognitive workload while using GOST, and (5) Identify potential design issues. This thesis outlines the methodology to evaluate and obtain results useful for further review and development of system capabilities.

The goal of the experiment is twofold. First, to evaluate cognitive workload of the participants while using GOST and compare that to the workload using baseline tools utilizing participants who have not previously been exposed to GOST. Second, to compare the performance of intelligence analysts acting as subject matter experts (SMEs), with novice users, each group using the toolsets to complete search tasks.

1.4 Hypotheses

This research effort seeks to test the following hypotheses:

H_0: Performance SME = performance novice

H_1: Performance SME $\neq$ performance novice

H_0: Performance GOST = performance baseline

H_1: Performance GOST $\neq$ performance baseline

II. LITERATURE REVIEW

In order to model analyst decision making it is necessary to understand the domain task requirements and system requirements. The following sections cover these and other topics essential to understanding the research methodology.

2.1 User Profile

The intelligence analyst (IA) is a primary focus of the research. The skills and knowledge of the analyst are of interest along with the varying levels of expertise demonstrated by participants.

2.1.1 Intelligence Analyst

Increasing amounts of available data makes determining relevancy more difficult for intelligence analysts. Analysts are expected to produce quick, accurate assessments that require high workloads through the process of search and retrieval, integration, and synthesis of data from multiple resources (Greitzer, 2005). Open source intelligence (OSINT) is derived from newspapers, journals, radio and television, and the Internet (Best & Cumming, 2007). The disparate sources are the basis for conflicting information and the reason for a human analyst in the decision making process. In addition, many of these sources present dynamic, time-critical, and often incomplete data. As described by Best & Cumming (2007):

> Definitions of 'open source information' have varied over time. Most simply, the term refers to information that is unclassified. It also has been defined to signify information that is derived from overt, non-clandestine or non-secret, rather than hidden or covert collection. The Intelligence Community defines open source information as that

information that is publicly available material that anyone can lawfully obtain by request, purchase, or observation.

Analysts are tasked with finding relevant data and creating an end product that conveys their understanding of a topic or scenario.

Geospatial Intelligence (GeoINT) and Open Source Intelligence (OSINT) analysts collect and analyze data in order to convey relevant information to customers who want a better understanding of selected events. While there is a significant overlap in skills and tasking, a GeoINT analyst has a stronger focus on geospatial relevancy, while the OSINT analyst is more likely to focus on data analysis (National Geospatial-Intelligence Agency, 2009). Both of these areas are addressed by the GOST system and consequently, both analyst types are the target users.

GeoINT analysts and OSINT analysts share some common attributes including an education background in cartography, geography, Geographic Information Systems (GIS), Physical Science, Applied Mathematics, Statistics, or a related discipline. They also share many technical skills, including experience with various remote sensing and geospatial systems. There is also common shared technical knowledge, including geospatial, sociopolitical, and security and mission-related (National Geospatial-Intelligence Agency, 2009).

There are many skills and requirements that differ between GeoINT and OSINT analysts. A GeoINT analyst focuses more on physical and spatial attributes whereas an OSINT analyst focuses on qualitative data which consists of attributes that distinguish or describe a given topic or geographic area. The GeoINT analyst relies on imagery, understanding of geography, spatial analysis, GIS, social and physical sciences. The GeoINT analyst will also need to have extensive technical skills and knowledge of geospatial systems and will be tasked with using these skills.

The OSINT analyst will need to have data query knowledge and skills along with a broad knowledge of world events. The OSINT analyst will be tasked with monitoring and reporting on many types of media sources and applying their knowledge of local history, customs and current events. The OSINT analyst is a data mining specialist that relies on expertise in identifying, acquiring, analyzing, and evaluating data sources.

Most of the user groups interviewed collected OSINT for use in (or as ancillary sources to) GeoINT products. Because of the overlap of tool usage between GeoINT and OSINT, the GOST system combines aspects of both. Consequently, the usability analysis needed to address both the ability to complete geospatial tasks along with the effectiveness of data mining tasks.

2.1.2 Expertise

As posited by Feltovich et al. (1997) and Kurland et al. (2006), experts demonstrate more skills and knowledge in a given domain than novices. While an individual participant may have expertise in a particular area, during the course of the experiment they will demonstrate their ability to apply expertise in the context and domain presented by the system and scenario task. As elucidated by Serfaty et al. (1997), the performance of experts is impacted by the working environment, task and domain of the problem being studied. These are the constraints of interest in the perception-action cycle outlined by Dainoff et al. (2012) and are important considerations in the evaluation of the system and the participant.

Expertise is the ability to apply knowledge or skill to produce concrete results in the context of a task in a particular field (Feltovich, Ford, & Hoffman, 1997; Oxford Dictionary, 2009; Ericsson et al., 2007). Experts working in their domain should demonstrate both speed and robustness, which Ericsson et al. (2007) call superior performance. Measurable discrimination and consistency are necessary to qualify as an expert (Shanteau et al., 2003; Weiss & Shanteau, 2005; Ericsson et al., 2007). While some contend that 10,000 hours of deliberate practice is required to develop expertise (Gladwell, 2009; Horn & Masunaga, 2006), this may vary by domain (Ericsson et al., 2007).

2.2 Search Task

The OSINT analyst is commonly tasked with searching for information which they distill and form into a cogent format to convey as relevant knowledge to interested parties. This process puts a temporal and cognitive burden on the analyst who is time constrained in their effort to transform raw data into understanding.

2.2.1 Temporal and Geospatial Search

The search process is based on performing a structured gathering of data in order to generate a report that analysts employ to convey the acquired knowledge. Search engines promote exploration, aggregation, and comparison of information along with the synthesis and evaluation that supports the investigation of a topic (Marchionini, 2006).

Keyword search involves a simple or advanced search, one that should be fast and accurate (Vu, Proctor, & Garcia, 2012). Search involves the three processes of exploring, enriching, and exploiting (Pirolli & Card, 2005). The exploring phase increases the span of information analysis. Enriching is the process of narrowing the collected information for analysis. Exploiting involves a more thorough evaluation of the documents. These three phases may be in conflict due to time constraints imposed by the task.

2.2.2 Data Transformation

Data transformation models can appear in various forms. Kuperman (1997) states that "The transformation of data into information is a value-adding process." In the geospatial domain, search tools are focused on structuring results based on physical location. Other factors that add value in the domain are the identification of temporal elements, named entities, and the language of origin. Figure 1 shows how, in the context of the GOST system, unstructured open source data is transformed to deliver understanding to the user. Following the funnel in the center of the diagram, the process starts with data from sources such as Google, Bing and Twitter. This geospatial, temporal, and topical information is organized as relevant knowledge to give the analyst predictive understanding. The boxes on the right show how GOST capabilities aid in transforming the data into understanding.

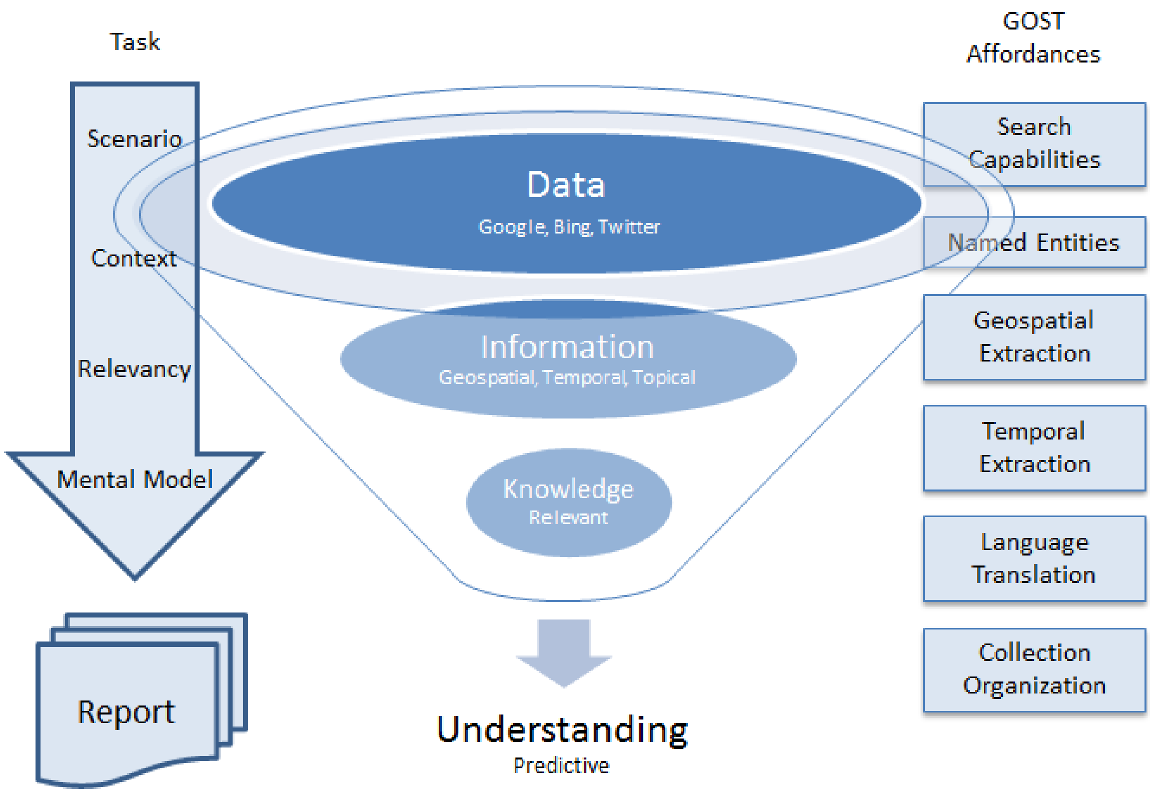


Figure 1: Data transformation into understanding (based on Kuperman, 1997)

As shown in Figure 1, when a scenario task or topic is introduced, the initial step is to search and filter data based on the scenario or topic. This applies contextual framing which structures the data and yields information. Information is then organized by relevancy to the topic via geospatial, temporal, and topical associations, which produce knowledge. Knowledge is accessed by means of a mental model which reflects the goals and constraints of the scenario and results in understanding. This understanding of the scenario or topic can then be used as a predictive tool (Marchionini 2006; Libicki & Johnson 1995).

As shown in the task flow of Figure 1, the process begins with a scenario task that drives the initial data collection. The scenario context provides a basis for determining the relevancy of retained information. The mental model then affords access to the accumulated knowledge which can be presented in report form. The associated GOST affordances are indicative of the ways in which the system supports the process, providing search capabilities, identification of named entities, as well as geospatial and temporal extraction of information. GOST also provides language translation and the ability to categorize information into relevant collections.

Currently, analysts use a basic toolset to evaluate data and determine relevancy (personal communication, January 2013). Understanding this process and creating a model allows researchers to more effectively study the procedure and aid software developers in creating new tools that allow analysts to do their jobs more efficiently and effectively (Spence, 2000; Crandall et al., 2006). Analysts are often required to assess geospatial and temporal information in order to ascertain contextual relevancy. In doing this, the analyst develops a mental model of the scenario being studied.

Salas & Klein (2001) maintain that schema is the "expert's memory structure for storing and retrieving relevant experience." Crandall et al. (2006) show that discovering meaning occurs when the focus shifts "from examining individual data records to more general characteristics of the data set as a whole." Both schema and meaning are integrated in the Pirolli & Card (2005) sensemaking loop and integral to the analysts' mental model. Pirolli and Card (2005) contend that as effort and structure are applied in the sensemaking loop, schemas are developed which allow conclusions to be drawn. Data is transduced from "its raw state to a form where expertise can apply." Hypotheses can be tested and a final representation can be formed to facilitate communication. While the Kuperman model (1997) focuses on the transformation of data, the sensemaking process includes the development of the analysts' mental model, or schema. Both schema and meaning are integrated in the Pirolli & Card sensemaking loop.

2.2.3 Information processing

The analyst task is strongly weighted toward the encoding and processing of both textual and visual information presented by the system. When viewed in context of an information processing model such as the one presented by Hollands & Wickens (1999), the ability to perform the task is influenced by user experience and consequently, long term memory will come into play to varying extents depending on user expertise. There is a relatively heavy burden on central processing as the human is required to make many decisions and constantly update their working memory as new information is presented by the system. Studies on running memory tasks have indicated that, while the typical memory span is less than five chunks, this can be expanded by domain expertise (Wickens & Carswell, 2012).

2.3 System Development and Profile

2.3.1 Software Development

Anselin (2012) provides a summary of the status of spatial data analysis software, giving an overview of the history of the available software and its development. Anselin highlights how spatial analysis has moved into the mainstream as well as becoming accessible and easy to use. There is also a developing awareness of the interdisciplinary nature of the area of Geographic Information Science and its importance in making use of the growing quantities of geospatial data (Blaschke et al., 2011).

Usability tools can be applied "at different stages of the software development process" (Horsky et al., 2010). Software development consists of at least five distinct stages, including the evolution of development which may include alpha and beta releases of the software to gain feedback from potential customers (Rajlich, 2000).

One way of tracking product maturity is by using Technology Readiness Levels (Mankins, 1995). The Technology Readiness Level (TRL) scale tracks product development through the use of nine levels which are grouped into Basic, Advanced, and Applied categories.

The Basic category, consisting of levels 1-3, is where the basic principles along with the concept and application are formulated, as well as identifying critical functions and characteristics. In the Advanced category, levels 4 and 5, the concept is validated in a laboratory and in a relevant environment. Finally, in the Applied category, levels 6-9, a prototype is demonstrated and the system is developed for mission operations. One role of the TRL scale is to reduce risk in implementing new technology (Graettinger et al., 2002). The TRL scale is a mechanism to better understand the risks and costs involved in system application (Moorhouse, 2002). A higher TRL score indicates reduced unknown risk in using the system along with a more accurate understanding of costs (Moorhouse, 2002). "Effective use of TRLs can reduce the risk associated with investing in immature technologies" (Graettinger et al., 2002).

"Much of the value of TRLs comes from the discussions between the stakeholders that go into negotiating the TRL value" (Graettinger et al., 2002). By using the TRL scale to track development, the software developer can more effectively address weaknesses in the system and concerns of the consumer.

2.3.2 Decision Support Systems

Both baseline analyst tools and GOST constitute Decision Support Systems which are being evaluated. How a decision is structured influences how value is apportioned to the task objectives (Clemen & Reilly, 2001). Because the task involves geospatial data, visualization plays a part and is not considered an independent task. Task components must be integrated with data management, decision support, task management, as well as content authoring and publishing (North, 2012; Shneiderman, 2002).

The Perception-Action Cycle provides the interaction framework with the Decision Support System (DSS). This cycle includes the Gulf of Execution which constitutes the user actions with the system and the Gulf of Evaluation which constitutes the user analysis of the change to the system (Norman, 2002; Spence, 2000). The process of taking action based on a goal is complemented by the evaluative process whereby the user contemplates the result of their action and the corresponding change in system state.

The Perception-Action Cycle considers the dynamic nature of visual change that occurs when interacting with the system (Spence, 2000). It provides both the basis for the mental model as well as additional structure for insertion of usability metrics. The Perception-Action Cycle also provides consideration for influencing factors, including organizational, environmental, individual, and task or scenario factors.

In the case of a DSS where the primary goal is information processing, presenting the user with more information is only helpful if the information is usefully structured. As is often the case, the human may be working under a time constraint where more information would be detrimental to making a decision. The goal of a DSS is to provide information that is structured and relevant to the problem. In the case of GOST, the system design being used is user-centered design (Czaja & Nair, 2012).

A DSS should be integrated with the decision process of the operator to enhance cognitive decision making capabilities (Fendley & Narayanan, 2012). The decision process model will allow identification of areas where GOST can impact and potentially improve decision making and analyst performance.

Information gathered from analysts through interviews indicated the system capabilities needed to perform search tasks. These requirements include the ability to rapidly generate new queries and tailor previous workflow and queries to new tasking. The search process is based on performing a structured gathering of data in order to generate a report that analysts employ to convey the acquired knowledge. Search engines promote exploration, aggregation, and comparison of information along with the synthesis and evaluation that supports the investigation of a topic (Marchionini, 2006). The analysts also need to consider the pedigree of source material and the completeness of an on-going analysis. They need to assess the uncertainty in an evolving product as well as the ability to generate timely intelligence products. Providing the capability for a less experienced analyst to rapidly adopt a standard workflow is also a benefit.

As shown in Figure 2, GOST is a web-based system that includes assisted search construction, scheduled searches, machine translation, and taxonomy building. It provides content management and interactive filtering along with geospatial and temporal visualization of results. The system identifies named entities such as people, places, and organizations to aid in information extraction. These "best of breed" tools are delivered in an easy-to-use interface.

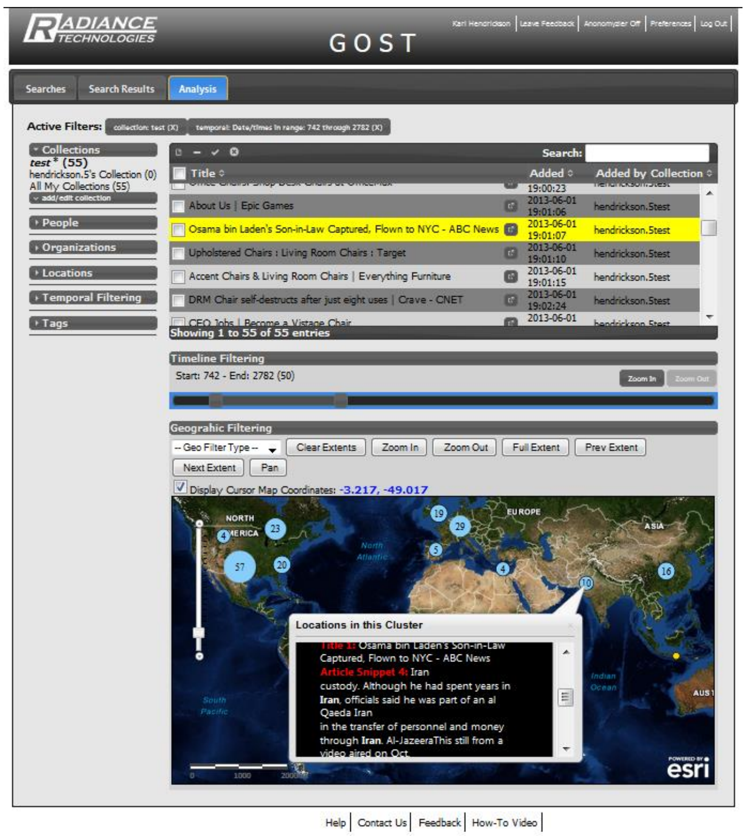


Figure 2: Geospatial Open Search Toolkit (GOST)

2.4 System Analysis and Mental Models

The purpose of the system analysis is to perform a comprehensive assessment of the system. This is comprised of the interactions between human and system, the dynamics of the system, and the analysis of the system in context of the task being performed (Woods & Hollnagel, 2006). From this basis, a mental model is created which can be used to track participants in the process of completing a relevant task.

2.4.1 System Analysis

The system analysis is comprised of three parts: function analysis, heuristic analysis, and usability analysis. The function analysis outlines the affordances of the system. The heuristic analysis studies the human-computer interface (HCI), potential user interaction with the system, and potential usability issues. The usability analysis provides a structured user interaction with the system, allowing a closer look at usability issues and elements of user interaction. These analysis elements combine to provide the basis for a model which integrates the user process with system affordances and a structure for metrics. As shown in Table 1, each analysis focuses on a different aspect of the human-computer interface.

Table 1: Elements of System Analysis

Analysis	**Focus**
Function Analysis	System Affordances
Heuristic Analysis	System Interface
Usability Analysis	User interaction with system

A function analysis uses simple logic in conjunction with task and function descriptions to identify significant relationships within the system (Homeland Security Institute, 2009; Meister, 2000). The objective is to understand the scope of the functions performed by the system (Jacko et al., 2012; Wickens et al., 2004). The function analysis (Appendix D) shows the system affordances and overall system structure. This provides both feature enumeration as well as elucidation of constraints on user mobility within the system. The function diagram attempts to strip away the graphic user interface and all related devices for user interaction. This provides a much broader, less prescriptive, view of the system's workings.

A heuristic analysis is a commonly used tool among usability professionals (Barnum, 2011) which has been shown to be effective when combined with other methods (Horsky et al., 2010). Gerhardt-Powels (1996) implemented a set of ten principles to enhance a human-computer interface design. Molich and Nielsen (1990) also produced a set of usability principles, which developed into Nielsen's ten usability principles and have been widely adopted. Both the Gerhardt-Powals and Nielsen principles have been shown to be effective (Hvannberg et al., 2006) and can be applied to highlight usability issues. In a usability analysis of software prototypes, Karahoca et al. (2010) showed that the Nielsen heuristic principles contributed to enhanced usability. It has also been shown (Alsumait et al., 2010) that the Nielsen heuristics can be effectively expanded to address new application domains. Heuristic principles can also be modified (Sivaji et al., 2011) or used as the basis for a usability assessment scheme (Horsky et al., 2010) which can be tailored to fit a particular need (de Kock et al., 2009) as they were in this study.

The Quesenbery 5E principles can be used to guide usability testing where the development goals are to create a system that is effective, efficient, and easy to use. The Quesenbery principles are: Effective, Efficient, Engaging, Error Tolerant, and Easy to Learn (Quesenbery, 2012). The Quesenbery principles are useful for doing an initial system evaluation and providing structure to the discussion. They can also be expanded and developed as the evaluation proceeds (Barnum, 2011).

As shown in Table 2, a merged list of Gerhardt-Powals and Nielsen cognitive design principles were developed as guidelines to be used in the development of software and were deemed appropriate for this evaluation. These were scored on a scale of 1-10, from weak to strong. While the GOST system scored well in many areas, low scores were of interest to inform further development. At the time of this analysis, the system was rated at TRL 5 which was indicative of a system in the development phase. Being cognizant of the heuristic analysis aided in structuring the usability analysis.

Table 2: Cognitive Design Principles grouped by score

	Cognitive Design Principle
Strong	Group data in consistently meaningful ways to decrease search time
	Match between system and the real world
	Aesthetic and minimalist design
Average	Reduce uncertainty
	Present new information with meaningful aids to interpretation
	User control and freedom
	Recognition rather than recall memory
	Flexibility and efficiency of use
Weak	Visibility of system status
	Helping users recognize, diagnose and recover from errors
	Automate unwanted workload

The framework proposed by McNeese et al. (1999) provides an appropriate structure for usability metrics. The goals of the study were associated with model development and given in the introduction. The experimental world of the study is a synthetic environment. Knowledge acquisition tools included interviews, questionnaires, and observation. Representation was both conceptual and computational, using function analysis along with the process model. Evaluation of both a quantitative and qualitative nature was used. Post-session questionnaires along with session recordings provided data to evaluate both qualitative as well as quantitative aspects of the user experience.

Crandall et al. (2006) address issues related to the cognitive demands created by information technology, which provide incentive to measure cognitive workload of the analyst while using the system. Spence (2000) also addresses the mental mapping that occurs when the user interacts with the software. System navigation is an important aspect of efficiency, effectiveness, and ease of use. As listed in Table 3, the primary questions posed by Spence (2000) during navigation can provide the basis for creating decision points and inserting metrics in the model.

Table 3: Navigation Decision Points (Spence, 2000)

Where am I?
Where can I go (from here)?
How do I get there?
What lies beyond?
Where can I usefully go?

Greitzer (2005) indicates that it is difficult to conduct true experiments in the Intelligence, Surveillance, and Reconnaissance (ISR) domain. One method of addressing this issue is through the use of structured and semi-structured tasks (Hammond & Hammond, 1966). The use of an autonomous task scenario that reflects the analyst ecology can be a useful tool to elicit both expertise and representative actions based on existing skills and mental models (Spath et al., 2012). Woods (1995) and Messick (1994) have shown the value of using scenarios as a "context-bound methodology which fosters a rich cognitive interaction between people and the system being studied" (Hammond, 2001). A realistic problem scenario provides a richer context than a fictional example (Spath et al., 2012).

As part of a usability analysis, the study performed interviews and background on the intelligence analyst as well as the role of analyst in search process. It also investigated the role and affordances of the GOST system. Four novice participants and six subject matter experts (SME) participated in a usability study which utilized a structured scenario task. The study looked at task completion, errors, time on task, and affordance utilization. Participants were given specific tasks relevant to an overall scenario. They were given verbal instructions explaining what was expected and guidance on how to accomplish the task within the system. They were evaluated on their ability to perform the task through the use of the system. The ability to complete tasks was 85% for novices and 86% for SMEs. Novices had a critical error mean of 4.0 with non-critical error mean of 7.3. A non-critical error was defined as a deviation from the task with the need for self-error recovery. A critical error was defined as a complete inability to perform the task due to a system error or the inability to find a system function, with the need for error recovery from an outside source, such as the help menu or administrator. Novice affordance utilization was 19% compared to 13.5% for SMEs. Error rates were not tracked for SMEs. Time on task data provided background data which was used to inform model metric insertion in the current study.

2.4.2 Mental Models

The purpose of the system analysis is to perform a comprehensive assessment of the system. This is comprised of the interactions between human and system, the dynamics of the system, and the analysis of system in context of the task being performed (Woods & Hollnagel, 2006). From this basis, a mental model is created which can be used to track participants in the process of completing a relevant task.

Mental models are used to gain insight into the process and to allow metrics to be applied in the study of how a participant performs in the context of a relevant task. Mental models provide a schema of dynamic systems, including system components, how the system works, and how it is used (Wickens et al., 2004). A mental model often represents how a system functions for a given task, incorporating user goals and action, as well as expectations about the system (Proctor & Vu, 2012). It may also provide a problem space to allow for more elaborate encoding of prior methods (Payne, 2009).

Mental models are useful for following the behavior of people executing a task or using a system. Combining the function, heuristic, and usability analyses with a relevant scenario task provides a basis for developing a model to provide a framework for metric insertion. Klein's Recognition-Primed Decision (RPD) model was chosen due to its incorporation of expertise, as well as its ability to integrate into a large model in an iterative structure. RPD was chosen for its representation of naturalistic decision making which looks at decisions in a real world context with an emphasis on the role of expertise (Klein & Klinger, 1991). This study looks at the role of the toolset and expertise and how they affect the performance of the participant.

Klein's Recognition-Primed Decision (RPD) model is based on situation recognition, serial option evaluation, and mental simulation (Klein et al., 1993). Klein and Klinger (1991) present three examples of the RPD model, from Simple to Complex. In the case of the Simple Match shown in Table 4, the situation is recognized and a course of action is implemented. In the Complex case shown in Table 5, a multifaceted process is involved where the decision maker may need to search for additional information and integrate this into their mental simulation of possible actions. Options are evaluated for workability and the process may iterate until a sufficiently workable course of action is identified. Both the Simple and Complex versions of this model were selected as components to be used in modeling the analyst search process.

Table 4: Simple Recognition-Primed Decision Model Elements with references to Perception-Action Cycle (based on Klein & Klinger, 1991; Norman, 2002)

Perception-Action Cycle	**Elements of Simple RPD Model**
Perception	Situation/perception in context
Interpret, Evaluate, Intention, & Action Plan	Situation Assessment & Activation from memory
Execute Action & Resulting Change in World	Implementation

Table 5: Complex Recognition-Primed Decision Model Elements with references to Perception-Action Cycle (based on Klein & Klinger, 1991; Norman, 2002)

Perception-Action Cycle	Elements of Complex RPD Model
Perception	Situation/perception in context
Interpret & Evaluate	Situation assessment & Activation from memory
Intention & Action Plan	Mental simulation review in context & plan feasibility determination
Execute Action & Resulting Change in World	Implementation

As shown by Dalinger & Ley (2011), RPD is an appropriate model for decision support systems and can be tailored to fit the area of interest. RPD has also been adopted as a computational decision model and for decision making in task networks (Ji et al., 2007; Leiden et al., 2001).

The RPD model provides a framework for inserting metrics. Each RPD model segment ends with "Implement" which indicates an action on the part of the participant. Tracking this action allows the researcher to follow the progress of the participant. As such, the development of this model addresses the need that the ISR community has identified to develop valid metrics to assess the usefulness and impact of tools and technologies that may aid analyst performance (Greitzer, 2005). This study incorporates a cognitive modeling methodology to aid in understanding the analyst's decision making process to better define metrics and design tools.

The model can be used to identify the delineation between human and computer in a Joint Cognitive System (JCS) such as with an analyst using the Geospatial Open Search Toolkit (GOST) system. How the decision process is structured influences how value is apportioned to the task objectives (Clemen & Reilly 2001). The model allows for the analysis of how tasks are apportioned and identification of JCS problem areas. The model is then used with a relevant scenario task for study of the system under semi-realistic conditions.

2.5 Measurement and Scoring

Qualitative and quantitative measures were used in the development and execution of this study. Qualitative methods are useful in conducting exploratory investigation, such as with interviews and questionnaires. The results of these methods are useful in forming hypotheses as well as structuring experimental methodology (Ravasio et al., 2004). Quantitative measures form a basis for agreement and certainty which can be discussed and supplemented with qualitative results (de Figueiredo, 2010).

2.5.1 Qualitative Measures

Qualitative measures can be used to supplement quantitative data as well as structure experimental methodology. Per Ravasio et al. (2004), the qualitative measures aim to discover structures, circumstances, relations, connections, and dependencies. These discoveries can identify factors of influence as well as aid in the construction of quantitative studies.

This exploration is found in the model development and validation. At the same time, the model is also being used for quantitative measures of performance. Both qualitative and quantitative measures were conducted in parallel, although their results are clearly distinguished. The qualitative measures focus on model development and validation along with ease of use feedback gathered through the questionnaire. The quantitative measures are errors and time on task along with comparisons of effectiveness and efficiency between levels of experience.

Turning qualitative data into numeric results can lose the depth and richness of some qualitative analysis techniques (Adams, Lunt, & Cairns, 2008). Consequently, the study also gathered qualitative information in an open questionnaire format. Qualitative methods are exploratory and allow researchers to assume active roles in identifying unexpected phenomena

(Bim, Leitão, & de Souza, 2007). Consequently, both qualitative and quantitative methods are valuable in identifying usability issues (Sauro, 2004).

A primary tool used in this study included Likert scales for gathering qualitative data. Likert scales provide a method for measuring a users' qualitative assessment of the system. They provide a relative judgment (Nicholls et al., 2006) of the item in question, usually using a seven or nine point scale (Beal, Dawson, 2007). While there may be some bias (Barnum, 2011), Likert scales provide an effective means of gathering qualitative data. An example question using the Likert scale is "It was easy to recover when making an error using GOST." Cicchetti et al. (1985) have shown that a 7 point scale is optimal. A 7 point rating scale was used; with semantic anchors at 1 (Strongly Disagree), 4 (Neutral), and 7 (Strongly Agree).

2.5.3 Report Scoring

An important element of the intelligence analyst task is the end product report. While this report is focused on responding to the scenario task, the structure is determined by the participant. This allows for a wide variation of report formats which must be scored by the researcher. Consequently, a scoring methodology and corresponding rubric must be developed to handle reports spanning a wide variety of structures and reflecting various levels of expertise.

As Lane (2010) and Messick (1994) contend, a scoring rubric should be domain specific, hence develop metrics relevant to OSINT, such as outlined in Lieberthal (2009) and the NATO OSINT handbook (NATO, 2001). The issue is to evaluate a task-driven performance assessment, composed of open-ended and semi-structured response formats. This allows the participant to tap domain knowledge relevant to the task. Scoring must address the analytic aspect of the report, such as content, organization, mechanics, and focus, assigning a score to each one (Lane, 2010).

The NATO Open Source Intelligence Handbook (Steele, 2007) identifies content that should be present and identified by the scoring system. This includes references to source material, an analytical summary, and Internet link tables. The report should be clear and concise and follow a logical structure (McDowell, 1997), as well as use plain and unambiguous language (McDowell, 2009). McDowell (2009) states that "the report should be used to display key points, conclusions, suggestions, and a synopsis of the supporting rationale." In addition, it should describe the quality and reliability of sources along with uncertainty associated with analytic judgments, and include alternative analyses where applicable (Lieberthal, 2009).

III. RESEARCH COMPONENTS

3.1 Overview

As indicated in the prior literature, intelligence analysts work under time pressure to generate products relevant to tasks. Tools are being developed to aid in the process of searching for and processing information that can be transformed into relevant knowledge. System analyses can inform the software development process and provide the basis for more detailed research. This research effort posits that model development can be used to investigate the effects of toolsets and expertise on analyst performance.

3.1 Research Framework

A research framework, shown in Figure 3, was developed to investigate the research questions and associated hypotheses listed in Table 6. This framework consists of four phases: System Analysis, Modeling, Validation, and Evaluation. The system analysis phase was performed as part of the background research. As part of this phase, semi-structured interviews with intelligence analysts were conducted to elicit information about tools and processes used in work tasks. Background research was performed to aid in domain understanding and a function analysis (Appendix D) was conducted to better understand the system being studied. A heuristic analysis was conducted by two human factors engineers in order to gauge potential strengths and weaknesses. A task scenario was developed in order to create a structured system walkthrough as part of the usability analysis. A follow-up questionnaire queried the participants on their use of the system.

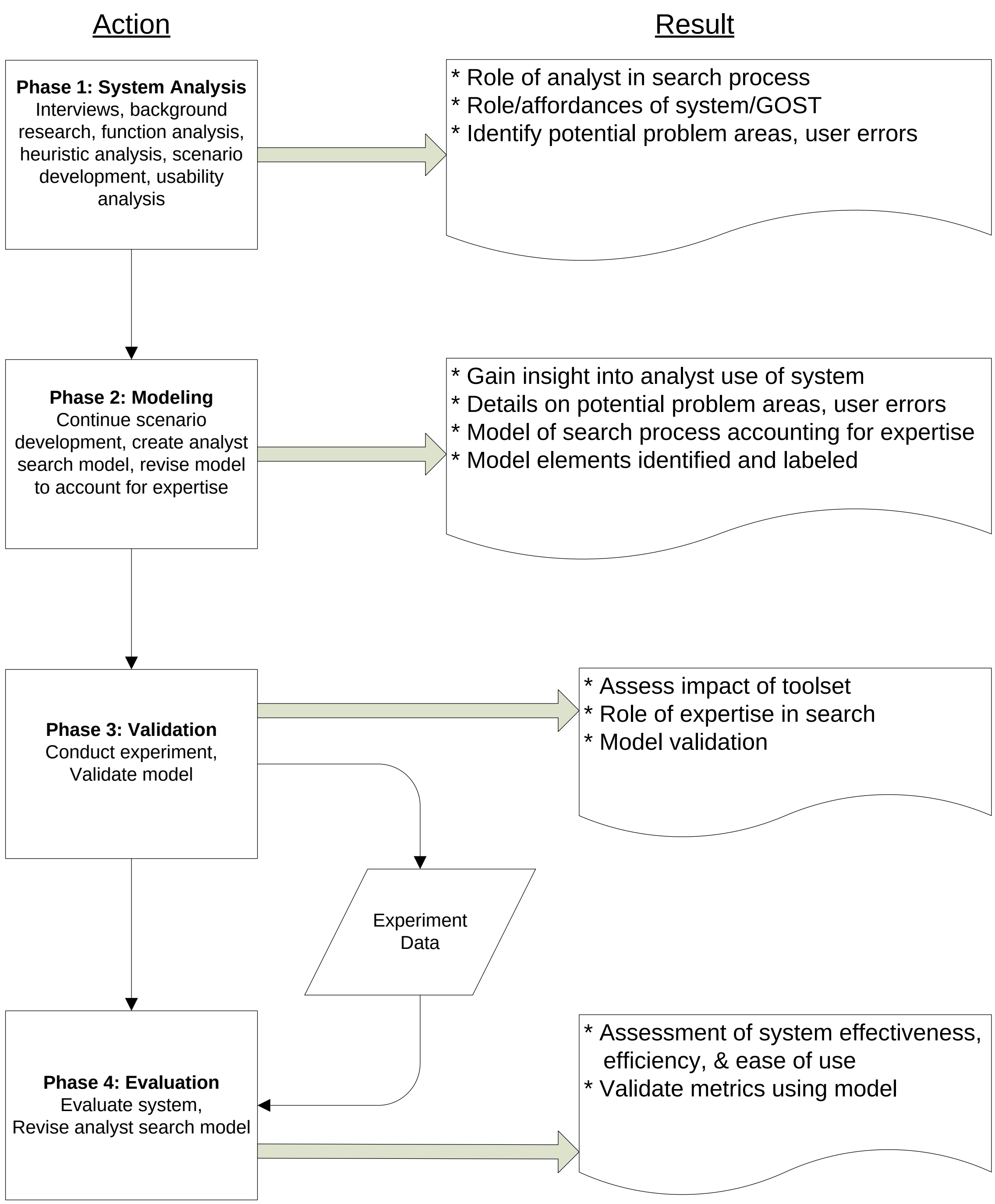


Figure 3: Research Framework

The modeling phase consisted of developing the process model, continuing development of scenario tasks in an autonomous task format. The task process model was developed and then revised to account for expertise. Model elements were identified and labeled in order to facilitate participant tracking. This resulted in a more robust model that would accommodate the iterative task aspects as well as various participant preferences.

Table 6: Research Questions and Hypotheses

Research Question	**Associated Hypothesis**
What are performance differences between expert and novice?	H_0: Performance SME = performance novice H_1: Performance SME $\neq$ performance novice
What are performance differences between systems, i.e., baseline and GOST?	H_0: Performance GOST = performance baseline H_1: Performance GOST $\neq$ performance baseline
Can a model be developed and validated that reflects the analyst search process?	
Does the model provide an accurate description of the role of both human and system?	

The validation phase consisted of conducting the experiment and validating the model. Both novice and expert participants completed two scenario tasks, one with each toolset. System actions performed by the participants were tracked along with physiological measures. The system actions were then labeled to match the model in order to track if and how the participant followed the proposed model. Changes were made to the model to reflect variations in participant behavior.

The evaluation phase consisted of analyzing experiment data in order to validate the model and evaluate system performance. It also assessed the effect of expertise in the task performance. Participant action was analyzed with respect to the revised process model, including error rates and segment completion times. A NASA TLX cognitive workload measurement was performed after each session to gather workload information and a questionnaire was completed after using the GOST toolset to elicit qualitative feedback.

3.2 Initial model

Observation of four analysts led to the development of the process model illustrated in Figure 4. One of the primary affordances of the RPD model is its ability to account for a changing context. For this reason, the model shown in Figure 4 utilizes RPD as a component. The RPD sub-sections of the model are indicated by labels Simple RPD and Complex RPD. These indicate the form of the RPD model being used from Tables 4 and 5. In the case of the Complex RPD, there is a need for the more multifaceted RPD strategy because this section is focused on assessing the task and determining what existing information and mental model can be applied. In the case of conducting a search, reviewing results, and checking the task status, each of these constitute a simple match with existing information, so the Simple RPD form is used. Each of these sections is labeled Simple RPD and encompasses or overlaps the Data Gathering, Information Processing, and Knowledge & Understanding Transfer stages of the data transformation process.

Each RPD component begins with an "Experience the Situation" event and concludes with one of the following actions: Enter Search Terms & Execute Search, Assess/Categorize, Extract report components, or Submit Report. Each of these actions corresponds to the "Implement" step in the RPD model. As such, the RPD model can be integrated at any step where the analyst must make a decision and take action.

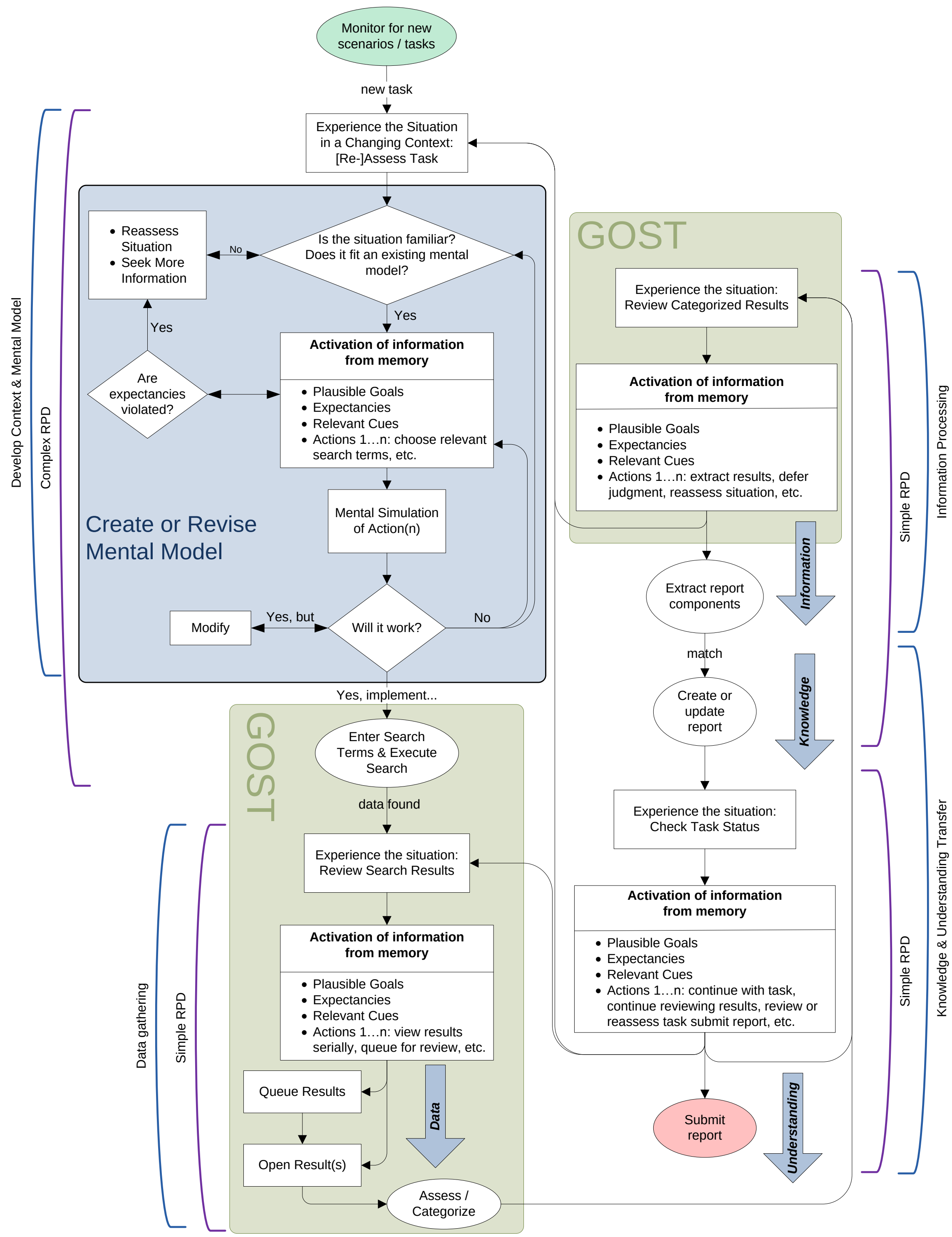


Figure 4: Analyst Process Model

In the analyst task, the function of performing the search and displaying the results is allocated to the system while the interpretation and evaluation aspects are allocated to the human. The user has three primary categories of decision making in the task. First, they need to decide which term(s) to include in the search for information. Second, they must decide which of the search results presented by the system are relevant to their task scenario. Lastly, they must decide what part of the relevant items selected will be included in their final report. As new content is displayed by the system, the user must be aware of the change in system state. Normally, because this is a user-initiated process, situation awareness is not an issue, although there may be cases where an unexpected change occurs of which the user is not immediately aware.

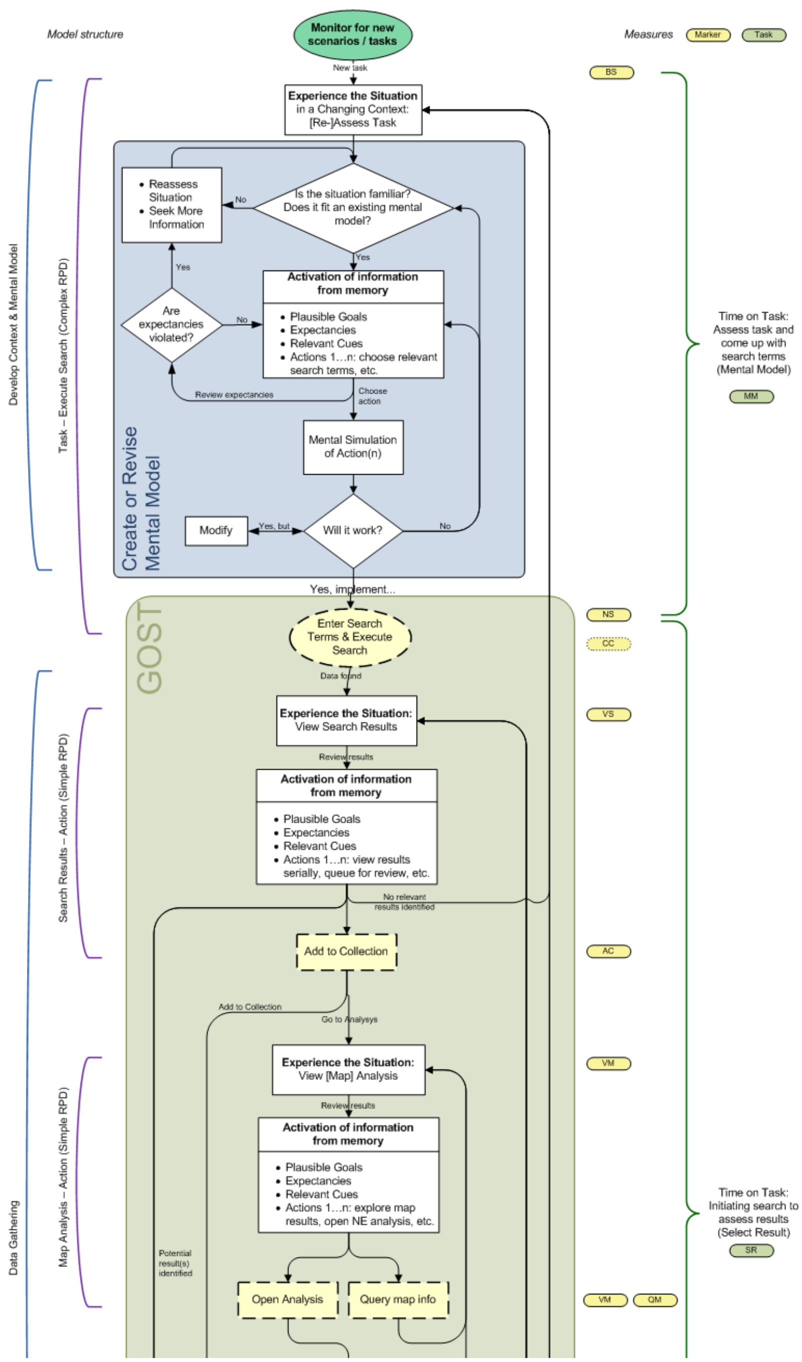
One of the challenges in building an accurate and useful model is the ability to account for flexibility in constraint parameters. Ideally, the model should account for varying levels of knowledge and expertise along with variable amounts of information, existing schemas, mental models, and task context. These present challenges in creating a flexible model which can take these variables into account while simultaneously presenting a succinct representation of the analyst search process.

3.3 Revised Model

One of the advantages of using the RPD model as a subsection of the overall process model is the ability to easily insert measures of effectiveness. Each subsection can be addressed separately to track errors in execution and tool use along with mental model formation and development. The RPD subsection also allows easy measurement of efficiency by tracking time on task. Workload can be assessed both through qualitative methods as well as comparing various iterations of time on task for a specific section.

3.3.1 Model Structure

As shown in Figure 5, there are three primary components to the model diagram: model structure, process detail, and measures. The model structure provides an overview of the model along with relevant sections which are indicated by brackets along the left side of the diagram. The model process detail provides the individual process steps along with major segments related to GOST system affordances and key processes. The measures listed along the right side indicate segments where experimental measurements can be taken to provide insight into the process.



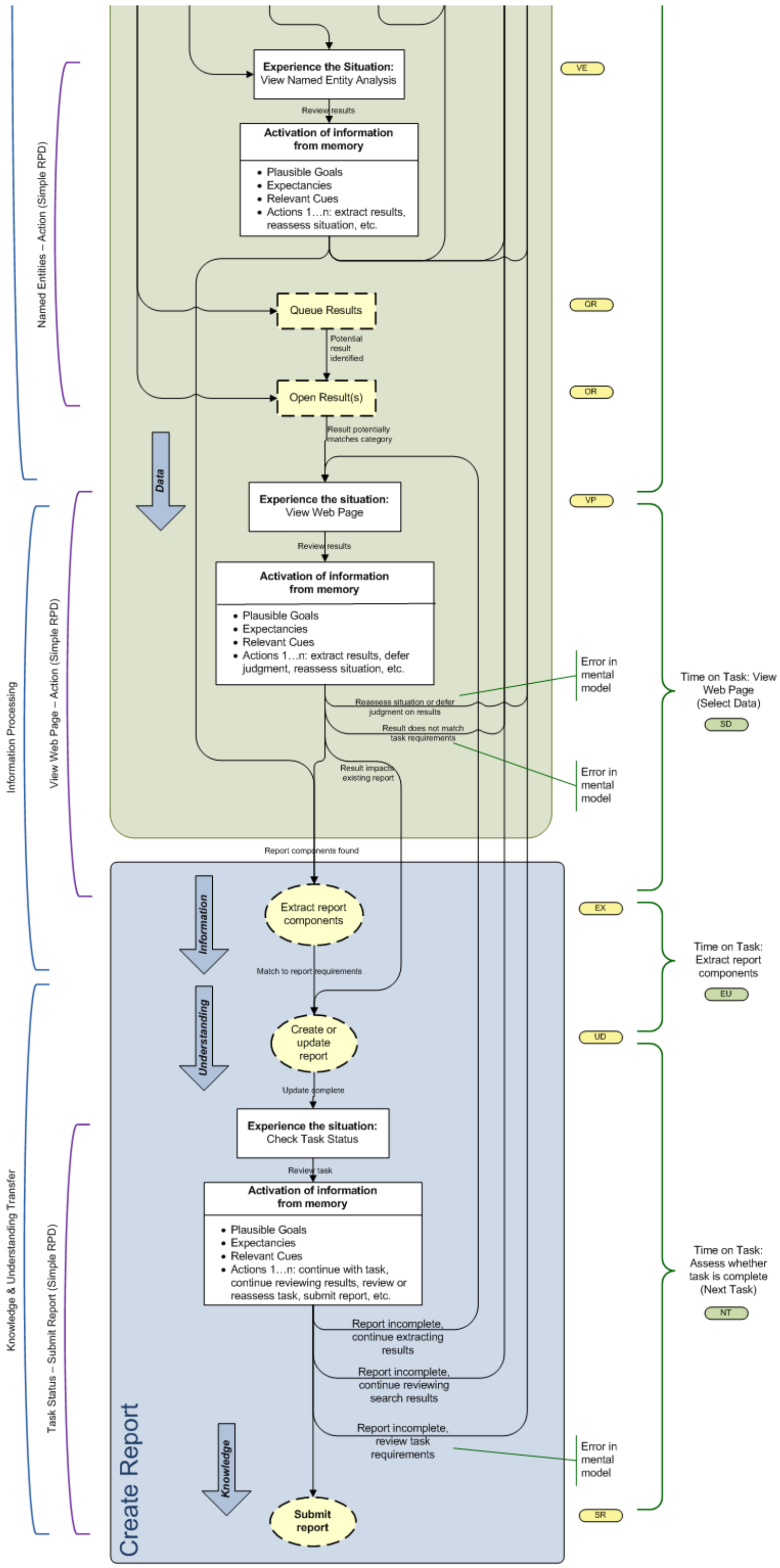


Figure 5: Revised Process Model

The model detail is divided into physical and cognitive actions. Physical actions taken by the analyst are highlighted in yellow with segmented borders. Sections labeled “Experience the Situation” indicate a visual action where the analyst is reading or otherwise assimilating relevant visual cues. All other components are cognitive actions.

The model is subdivided into four phases: developing context and mental model, data gathering, information processing, and knowledge and understanding transfer. The key elements of the data transformation are indicated, as well as the affordances of GOST. There are four action sections: Task – Execute Search, Search Results – Action, Map Analysis – Action, Named Entities – Action, View Web Page – Action, and Task Status – Submit Report. Each of these sections affords the researcher the ability to distinguish between cognitive and physical functions of the analyst. The four Time on Task sections correspond with actions taken at the conclusion of RPD sections. Also, two notes labeled “Error in mental model” indicate areas where mental model revision was deemed likely. Regarding the transformation of data into understanding, there are four arrows indicating the transformation of data into understanding. Note that each of these transformations occurs based on human action. Finally, there are three major sections that indicate distribution of tasks between human and computer: Create or revise mental model (human), GOST (computer/system), and Create report (human).

3.3.2 Analyst Process

The first step in the analyst process is to develop an appropriate context and mental model. After accepting the new task, the analyst assesses whether the task situation is familiar. This begins by assessing the scenario and the goals for the task and applying previous experience (knowledge and understanding) to create a mental model. The mental model is used to identify and develop associated questions. This provides a framework for the subsequent data gathering and information structure. If the task is not familiar, the analyst will reassess the task and seek more information until sufficient schema constructs are available to begin the task. The analyst then reviews relevant memory for plausible goals, expectancies, and cues. If no expectancies are violated, they will create a mental simulation of action, including identification of search terms and topics. If this is deemed feasible, they will begin the search process by entering search terms and executing a search. This action will occur in a web search engine (Google, Bing, etc.) or in GOST.

The next step is data gathering. The data gathering begins with the search execution and with viewing the search results. The analyst reviews the search results, looking for results that match the goals, expectancies, and cues established in the mental model. The results of the search are identified as potentially relevant or not, and the appropriate action is taken to either open the result for further review or to discard. The relevant results are assessed in detail and categorized within the task structure.

The analyst then enters the data gathering and information processing phase. During this phase, the analyst begins to search for data to enhance the mental model and answer outstanding questions. As this process progresses, data are categorized as applicable to the scenario and an overall information structure develops. Information that is relevant but not accounted for by the existing mental model or task structure may prompt the analyst to reassess the mental model or to develop or revise questions. As relevant information is extracted in this phase, components are added to the report and questions are answered. Both the mental model and information context become more robust.

The next stage is information processing. The analyst reviews the categorized data for relevant goals, expectancies and cues to determine which data components to extract as information. This extraction of information from data concludes the information processing stage.

The next stage is the knowledge and understanding transfer. The analyst then takes the extracted information in order to create or update the task report, using components that match the task requirements. The analyst checks the task status by reviewing the report, testing to see if the relevant goals, expectancies, and cues have been met. If the task requirements have been met, then the topical understanding is complete and the report will be submitted. If not, the analyst will continue to extract search results and review task requirements, adjusting their mental model as necessary. This phase is focused on answering the questions posed by the task and ensuring that the mental model is complete and that an understanding of the scenario has been attained. When finished, the knowledge and understanding transfer are complete.

3.3.3 Data Transformation

As indicated in Figure 5, there are points where data transformation occurs. The arrows for data, information, knowledge, and understanding correspond to the steps in Figure 1. When data is assessed and found to match the task context, it is retained by the analyst and changes to information. The information may be of a geospatial, temporal, or topical nature which fits the mental model being developed by the analyst. Likewise, when information is categorized and found to be relevant to the topic, it becomes knowledge. Finally, when knowledge is combined with the finished mental model, it represents an understanding of the topic which can then be conveyed. The topical and contextual understanding can then be applied in a predictive capacity.

An example of the data transformation process would be as follows. The analysts begin a search based on a set of relevant keywords that would be refined in order to produce results containing information that matches the task. Because they are working with an incomplete mental model, they will try to identify search terms that help to develop their mental model. This process can happen through trial and error or may be informed by their domain expertise.

As the search process continues, they may identify data that matches key terms such as people, places, or organizations. They may also identify matches based on temporal data. The analyst is continually questioning how the key search terms and data are correlated. The process of exploring the data and developing the corresponding mental model are key steps needed to successfully accomplish the task.

The analyst is then able to categorize the relevant data based on the matching points of identification while also reviewing the structure of the mental model. While the relevant matching points may be specific to the task, the categorization is more likely to take place at least one level higher in the topical taxonomy. This allows for a broader grouping of information which reflects the combined structure of the mental model and the task requirements. This step validates the information against the task which is fundamental in identifying which pieces of information are transformed into topical knowledge. Finally, as the various task requirements are completed, a general understanding is attained.

3.3.4 GOST

The affordances provided by GOST are highlighted in Figure 5. The GOST system is designed to provide the analyst with the ability to more effectively and efficiently find temporal, topical, and geospatial data and determine relevancy. Consequently, the model reflects the areas where GOST contributes in the process of searching for and assessing data. The system affords the user with the ability to find data that matches the task and to advance data transformation from data to information. GOST then aids in categorizing information so that it can be further utilized by the analyst. This is done through the use of collections which allow the analyst to identify, gather, and retain relevant information.

3.3.5 Model Affordances

The primary goal of model creation is to afford the researcher with a structure to facilitate experimental insight into a process or system. As shown in Table 7, the model structure presented here provides the following affordances for the researcher. While each of these have been discussed as part of the model construct, it is useful to reiterate that these are important aspects of creating a model.

Table 7: Model Affordances

Affordance	Implementation
Ability to distinguish between human cognition and system functions	Yellow boxes with dotted outlines indicate participant actions and interaction with system, all other are cognitive function
Allow implementation of performance measures	Markers and task blocks allow for Measures of Performance (MOPs)
Allow tracking of data transformation into knowledge	Data transformation arrows are overlaid on model
Identify GOST affordances	Highlighted in green box

3.4 Model & Measures

Through the utilization of these affordances, the researcher has the ability to measure time between physical actions which suggest an amount of cognitive action where the analyst is experiencing the situation.

The goal of using the Klein RPD is to create sections in the analyst process model that accurately reflect analyst work process and are easily quantifiable. As such, time on task measures can be applied to each RPD section. It is expected that errors and error recovery will be found within and between RPD sections. Each RPD section represents a starting state and ending state which are recognizable when using observation techniques to analyze task completion.

The measures afforded by the process model include the following: Time on Task, Scenario Task Completion, Errors, and Cognitive Workload. Time on Task is addressed through the use of the RPD and its ability to easily identify task actions. Scenario Task Completion is measured through the observation of actions that indicate completion of various aspects of the scenario task. Errors can be identified as deviations from the process model or by observations that indicated that the analyst is revising their mental model. Cognitive Workload can be measured using the NASA TLX after task completion (Hart, 2006).

Time on Task can be measured with the use of Morae which allows the researcher to annotate a recording of the participants' actions with the system. By marking these recordings with the measures labels from the process model (Figure 5), the researcher can trace the progress of the participant through the model. Morae can then export the measures labels with time stamps in order to facilitate inter-marker analysis. Task times and frequencies can then be calculated along with error rates.

IV. EVALUATION/METHODOLOGY

4.1 Experimental Design

The goal of this study was twofold. First, to evaluate cognitive workload of the participants while using GOST. Second, compare subject matter experts (intelligence analysts) with novice users utilizing participants who have not previously been exposed to GOST. The experiment utilized a mixed design model, with a between subjects design used to compare the two sample populations, experts and novices, and a within subjects design used to compare toolset use within each sample population. The two sample populations reflect the expertise levels being studied, expert and novice, which correspond to two distinct populations, expert analysts and ATIC analysts, respectively.

4.1.1 Participants

Expert participants were recruited with at least four years of experience working in the field. All had intelligence analyst skills, such as image analysis, geospatial and open-source knowledge. Novice users were recruited from the Advanced Technical Intelligence Center (ATIC) staff and student population, with less than one year of experience as an analyst. There were a total of 8 participants, including four experts and four novices.

4.1.2 Facilities / Equipment

Research was conducted at Advanced Technology Intelligence Center (ATIC) located at 2685 Hibiscus Way, Suite 110, Beavercreek, OH 45431. A desktop computer with the GOST application and supporting software was used in an air conditioned room. Each participant's interaction with the application was monitored by the facilitator seated in the same room. Note takers and data logger(s) monitored the sessions in the same room. The test sessions were recorded with Morae, SmartEye, and Equivital equipment. (Data collected with SmartEye and Equivital was not used in this study.)

4.1.3 Trial Procedure

Participants signed an informed consent (Appendix A) acknowledging that participation is voluntary, participation can cease at any time, and the session would be recorded, but their privacy would be safeguarded. The facilitator asked the participant if they had any questions. Participants completed a pretest demographic and background information questionnaire (Appendix B).

The investigator provided a training session to familiarize the participant with GOST. Then the participant was asked to complete tasks utilizing the system's affordances through the use of a representative scenario (Appendices E and F). After completing the task, the user completed a survey that included a description of the instantiated capabilities and several related questions that required utilizing a rating scale and answering open ended questions. A post-test questionnaire (Appendix C) with a Likert scale was administered to gather quantitative and qualitative feedback. NASA TLX was administered as well, with a final interview and debriefing.

4.1.4 Scenario

The tasks used for this evaluation were derived from test scenarios developed from use cases with the guidance provided by subject matter experts. Due to the number of functional capabilities, and the short time for which each participant would be available, the tasks selected were representative of real use and were used to substantially evaluate a subset of the capabilities of GOST.

4.1.5 Report scoring

End product reports generated by the participants were scored by a senior intelligence analyst. Relevancy and quality of content was scored along with an overall rating, which was used for comparison and analysis.

4.1.6 Treatment Order

Treatment order was randomized using a Latin square design. Due to the within subjects design used to test each toolset with each participant, there were two scenarios which were presented in alternating order as shown in Table 8.

Table 8: Design of Experiment

Expertise	Toolset	Scenario	Scenario Order
Novice	Baseline	Airlift	1
Novice	Baseline	Airlift	2
Novice	Baseline	Stealth	1
Novice	Baseline	Stealth	2
Novice	GOST	Airlift	1
Novice	GOST	Airlift	2
Novice	GOST	Stealth	1
Novice	GOST	Stealth	2
Expert	Baseline	Airlift	1
Expert	Baseline	Airlift	2
Expert	Baseline	Stealth	1
Expert	Baseline	Stealth	2
Expert	GOST	Airlift	1
Expert	GOST	Airlift	2
Expert	GOST	Stealth	1
Expert	GOST	Stealth	2

4.1.7 Independent Variables

As stated earlier, there are two groups of interest in this study, expert and novice analysts. These two levels of expertise constitute the two participant groups. The independent variable for tool use contains two levels, baseline and GOST. This reflects the two toolsets being tested.

4.1.8 Dependent Variables

Three dependent variables were analyzed: errors, cognitive workload, and report quality. Errors consisted of both critical and non-critical errors committed by the participant during the experiment. Cognitive workload was measured using NASA TLX. Report scoring was on a scale of 0-100, low to high.

Critical errors can also be assigned when the participant initiates (or attempts to initiate) an action that will result in the goal state becoming unobtainable. In general, critical errors are unresolved errors during the process of completing the task or errors that produce an incorrect outcome.

Non-critical errors are errors that are recovered from by the participant or, if not detected, do not result in processing problems or unexpected results. Although non-critical errors can be undetected by the participant, when they are detected they are generally frustrating to the participant. These errors may be procedural, in which the participant does not complete a scenario in the most optimal means (e.g., excessive steps and keystrokes). These errors may also be errors of confusion (ex., initially selecting the wrong function, using a user-interface control incorrectly such as attempting to edit an un-editable field). Noncritical errors can always be recovered from during the process of completing the scenario. Exploratory behaviors, such as opening the wrong menu while searching for a function, were coded as non-critical errors.

Cognitive workload is the amount of effort expended by the participant to complete a task. It is an indication of the difficulty of the task and/or the tool being used. Data gathered using Morae, NASA TLX, and post-test questionnaires were used to measure cognitive workload. Hart (2006) contends that the NASA-TLX is a benchmark tool in the measurement of cognitive workload. Burke et al. (2005) demonstrate the applicability to web-based systems.

4.1.9 Subjective Measures

Subjective evaluations regarding ease of use and satisfaction were collected via questionnaires, and during debriefing at the conclusion of the session. The questionnaires (Appendices B and C) utilized free-form responses and rating scales. Subjective opinions about specific tasks, time to perform each task, features, and functionality were surveyed. At the end of the test, participants rated their satisfaction with the overall system. Qualitative measures consisted of the measures listed in Table 9.

Table 9: Qualitative measures

Qualitative Measures
User satisfaction with task experience.
Aesthetic appeal of the user interface.
Level of frustration with using the system.
Level of motivation to continue using the system.
Ease of learning the system.
Satisfaction with search time and results.
Match of system to current mental model from past online experiences.
Amount that the system taxes user memory.
Efficiency gains as the system is learned.

Combined with the interview/debriefing session, these data were used to assess attitudes of the participants. Subjective and quantitative measures are presented in the next section.

V. RESULTS

The system was evaluated on three quantitative measures: report quality, errors, and cognitive workload. Due to the crossover design of the experiment, these were analyzed by group and within subjects. Results were evaluated for significance and tested for period and carryover effects. No interaction effects were found. Qualitative measures included a post-test questionnaire on qualities of the GOST system. Time on task measures were also evaluated.

5.1 Performance Metrics

A two-period crossover study analysis was performed (Fleiss, 1986), and, as shown in Table 10, no significant period or carryover effects were found. Applying the Bonferroni criterion ($\alpha = 0.05$, $\alpha_{test\text{-}wise} = 0.05/3 = 0.01667$) at the standard $\alpha = 0.05$ level, no significant results were found. If the overall level of significance was relaxed to the higher $\alpha = 0.10$ level, the data indicates that report quality for experts was significant and errors for novices could be considered to be marginally significant. This would support the alternate hypothesis that there is a significant performance difference between experts and novices.

Table 10: Treatment, Period & Carryover Effects

		Toolset		*Period*		*Carryover*	
		t stat	**p-value**	**t stat**	**p-value**	**t stat**	**p-value**
Novice	**Report Quality**	1.8733	0.2019	0.8631	0.4791	0.5610	0.6312
	Errors	-5.0000	0.0377	-1.0000	0.4226	1.0000	0.4226
	Cog Workload	0.7249	0.5438	0.1208	0.9149	0.1903	0.8667
Expert	**Report Quality**	5.6921	0.0295	3.7947	0.0630	-1.0738	0.3953
	Errors	-1.9426	0.1915	0.1943	0.8639	-0.4216	0.7143
	Cog Workload	0.7589	0.5271	-2.4033	0.1381	1.1487	0.3695

Table 11 shows the mean and standard deviation for each of the dependent variables. These will be discussed in the following sections.

Table 11: Mean & Standard Deviation for Dependent Variables

		Baseline		*GOST*	
		Mean	**S.D.**	**Mean**	**S.D.**
Novice	**Report Quality**	0.438	0.175	0.345	0.084
	Errors	0.500	0.577	3.000	0.816
	Cog Workload	62.083	6.255	61.042	18.601
Expert	**Report Quality**	0.588	0.165	0.363	0.214
	Errors	0.250	0.500	5.250	4.113
	Cog Workload	43.958	14.741	55.833	16.116

5.1.1 User Type

As shown in Table 11 and Figure 6, the mean report quality scores for experts were higher than novices while the cognitive workload was lower, but neither of these measures reached the standard ($\alpha = 0.05$) level of significance. This would support the null hypothesis that the performance of novices and experts is not significantly different.

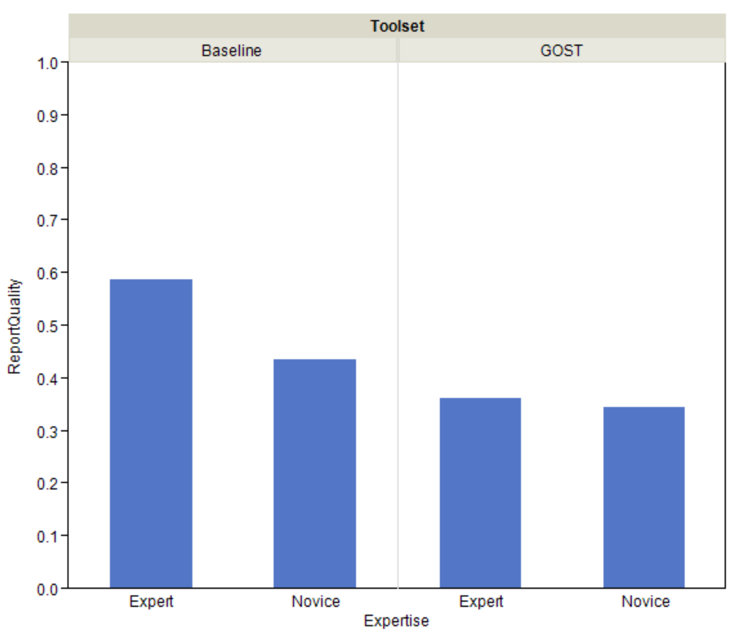


Figure 6: Report Quality scores

Figure 7 compares the dependent measures by level of expertise. This allows for a visual recognition of patterns and outliers.

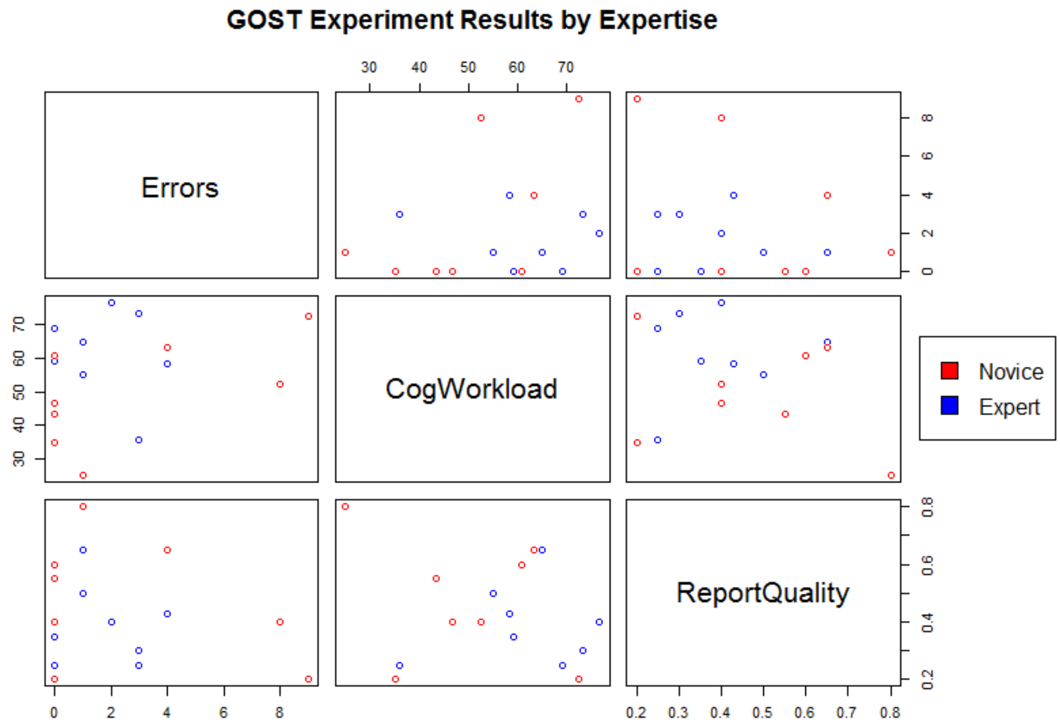


Figure 7: Comparison of measures by level of expertise

5.1.2 Tool Used

As shown in Table 11, errors for novices were significantly higher with GOST than with the baseline toolset. Experts showed a marginally significantly higher report quality score with baseline tools over GOST. This supports the alternate hypothesis that the performance of the baseline and GOST toolsets are not equivalent. Figure 8 compares the dependent measures by toolset, giving a visual representation of patterns and outliers.

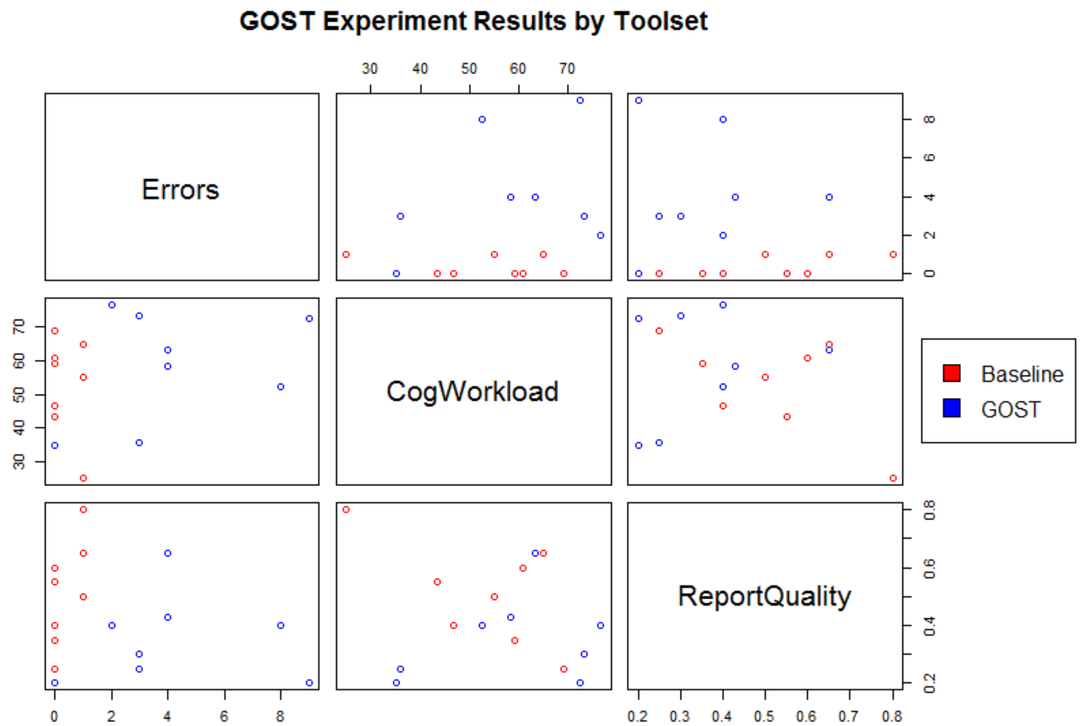


Figure 8: Comparison of measures by toolset

5.1.3 Errors

As shown in Table 11, errors for novices were significantly higher with GOST than with the baseline toolset. This supports the alternate hypothesis that the performance with baseline toolset and GOST are significantly different. The sample distribution for error data was found to have significant evidence to reject the normality assumption via a Shapiro-Wilk goodness of fit test ($W = 0.787766$, $prob<W = 0.0019$).

The Shapiro-Wilk Goodness-of-Fit test indicates that the error data does not fit a normal distribution. A closer look at the Goodness-of-Fit test, by expertise and toolset, as shown in Table 12, indicates that the baseline data fails the normality test. Due to the small data set and the variability between the participant groups, the data was treated as a normal distribution for the purposes of this study.

Table 12: Goodness-of-Fit Test (Shapiro-Wilk W Test)

	Novice		*Expert*	
	W	**Prob<W**	**W**	**Prob<W**
Baseline	0.7286	0.0239	0.6298	0.0012
GOST	0.9447	0.6830	0.9248	0.5641

The ANOVA F-test, as shown in Table 13, indicates evidence that the error distributions for each toolset are significantly different. This evidence agrees with the results found in the Goodness-of-Fit test in Table 12 above.

Table 13: F-test for results

	Errors	**CogWorkload**	**ReportQuality**
F =	0.0001	0.7292	0.6824
$P(F \leq x)$ =	0.0001	0.4217	0.4056

As shown in Table 14, Experts using GOST displayed the greatest variability in error rates, with a standard deviation of 4.11. Detailed error information shown in Figure 9 indicates that error rates for experts using GOST ranged from 0 to 9. This may indicate the need for more learning time with the toolset in order to become acclimated.

Table 14: Error Rate Means and Standard Deviations by Toolset and Expertise

	Mean	**Std Dev**
Baseline/Novice	0.50	0.58
Baseline/Expert	0.25	0.50
GOST/Novice	3.00	0.82
GOST/Expert	3.00	4.11

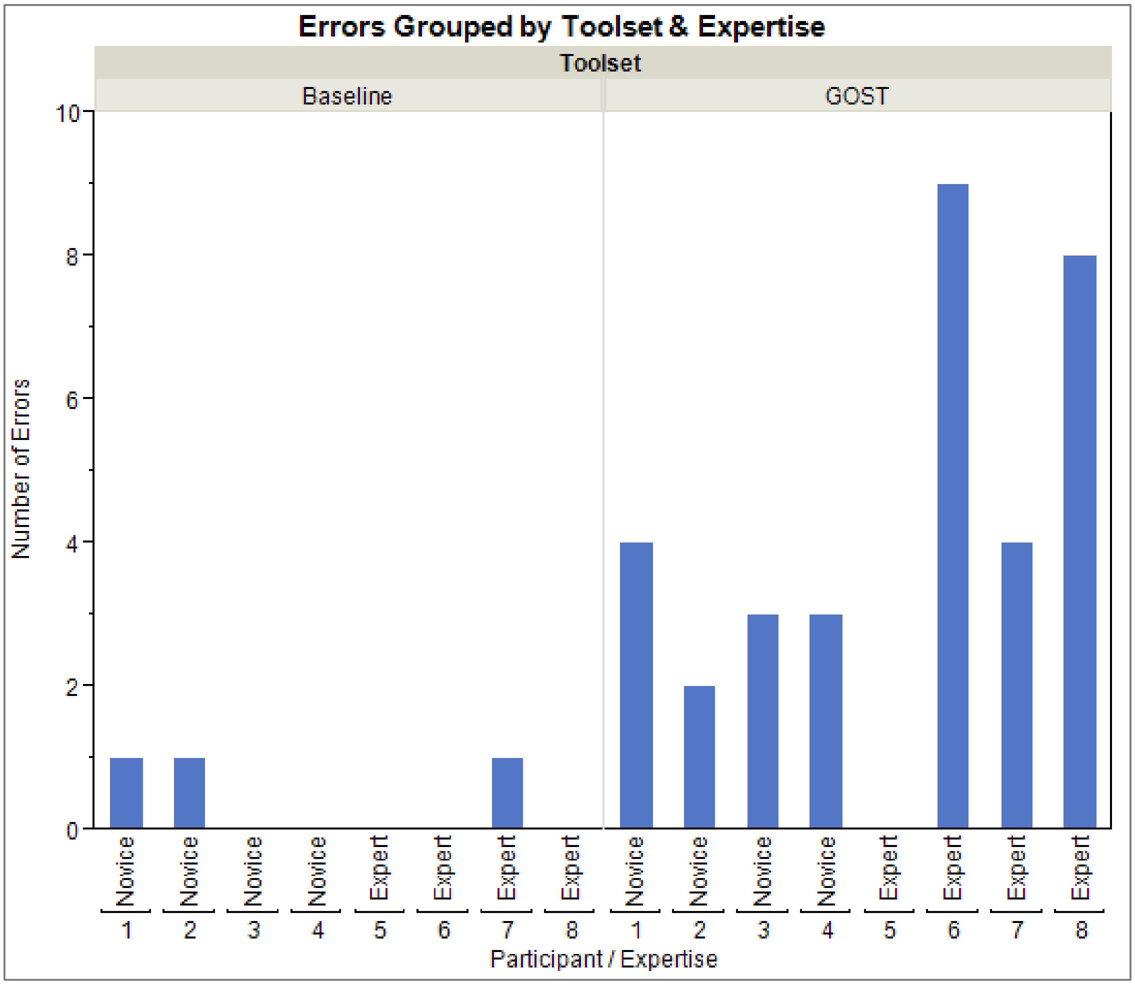


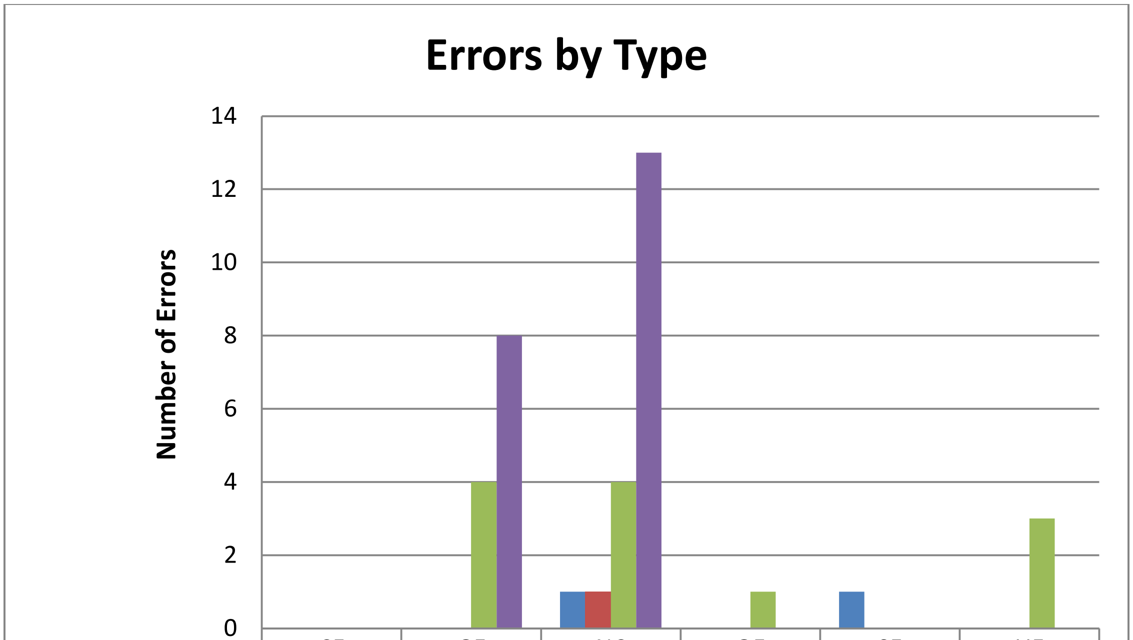
Figure 9: Participant Errors Grouped by Toolset and Expertise

Errors were classified as one of six types, as shown in Table 15. A Critical Error (CE) is one which the participant is unable to recover from without assistance. A GOST Error (GE) indicates a situation where the system was unable to accommodate the intentions of the user and displayed an error message. A Non-Critical Error (NC) is an error caused by a participant action which does not accomplish the desired task. An Other Error (OE) is an error that does not fall into one of the other five error categories. A Search Error (SE) occurs when the system returns an error in response to a participant search request. Commonly, this results in a "Page not found" message. A User Error (UE) occurs when the participant attempts to utilize an affordance unsuccessfully.

Table 15: Error Type Marker Abbreviation and Description

Marker	Description
CE	Critical Error
GE	GOST Error
NC	Non-Critical Error
OE	Other Error (used for NOC system errors)
SE	Search Error
UE	User Error

As indicated in Figure 10, a further breakdown of errors by type shows that most of the GOST errors, both for novices and experts, fell into the GOST Error (GE) or Non-Critical Error (NC) categories. The GOST Errors indicate that the participant was not fully acclimatized to the system or that the system did not respond as expected. The NC error indicates that the participant had difficulty accessing the appropriate system features but was able to complete the task through a course of "trial and error." NC errors indicate that the participant has not fully internalized the available system affordances.

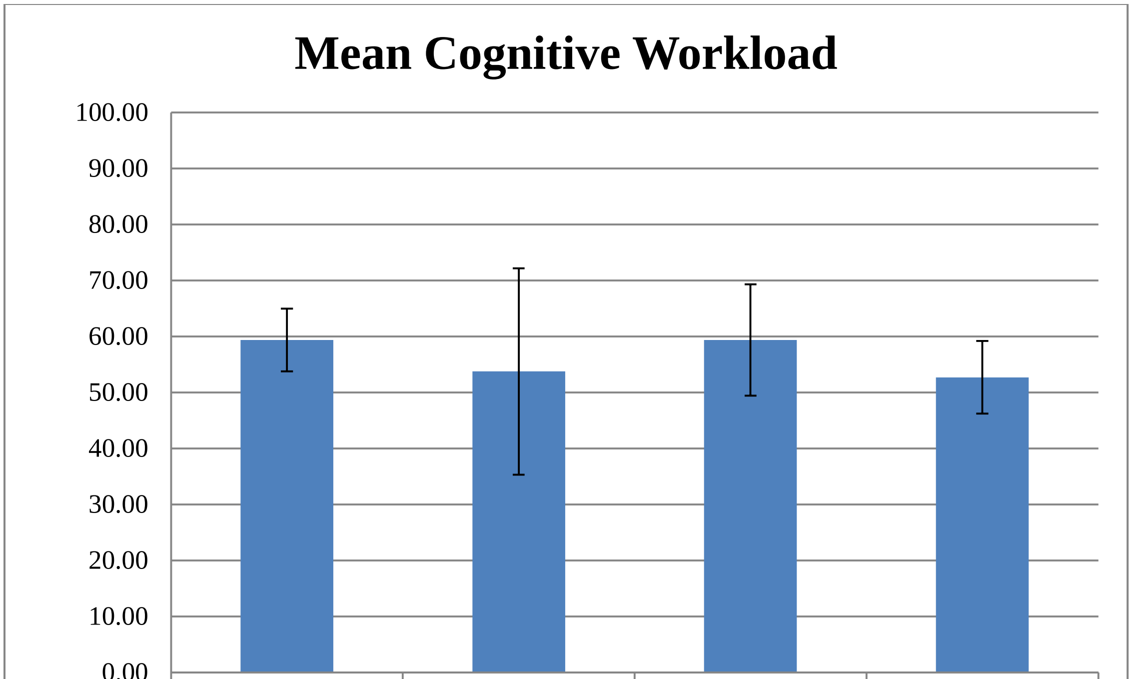


	CE	GE	NC	OE	SE	UE
Baseline/Novice	0	0	1	0	1	0
Baseline/Expert	0	0	1	0	0	0
GOST/Novice	0	4	4	1	0	3
GOST/Expert	0	8	13	0	0	0

Figure 10: Number of Errors by Error Type, Toolset, and Expertise

5.1.4 Cognitive Workload

Cognitive workload was measured using NASA TLX and scored on a scale of 0-100, low to high. As shown in Figure 11, cognitive workload was not significantly impacted by toolset.



	Baseline / Novice	Baseline / Expert	GOST / Novice	GOST / Expert
Mean	59.38	53.75	59.38	52.71
Std Dev	5.58	18.44	9.94	6.50

Figure 11: Mean Cognitive Workload (NASA-TLX)

5.1.5 Report

The task reports generated by the participants were scored on a scale of 0 to 1 by an experienced analyst. As shown in Figure 12, mean report scores for experts were significantly higher with the baseline toolset. This supports the alternate hypothesis that the performance of experts is significantly different from novices. This may indicate that experts need more time learning a new tool whereas novices need more attention to leaning a new process.

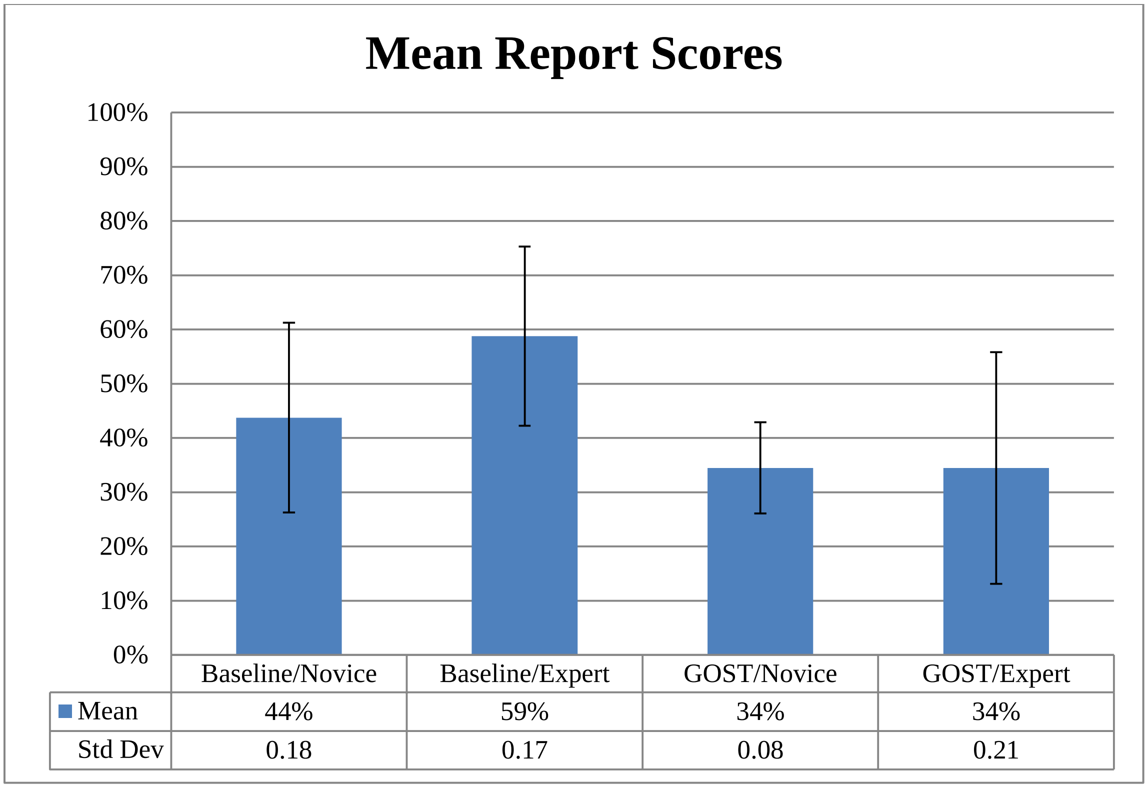


Figure 12: Mean Report Scores

5.1.6 Questionnaire

As shown in Table 16, results from post-test questionnaire about GOST system indicated significant differences between novice and expert participants on the highlighted questions. Comments related to these questions indicated some of the weaknesses. On question 5, expert participants cited a steep learning curve while novices cited lack of familiarity with subject matter. Regarding question 7, both groups cited the need for additional training and time to become more familiar with GOST. General comments reiterated the need for more time and training to become familiar with GOST.

Table 16: Post-test questionnaire results

#	Question	Novice	Expert
1	GOST is easy to learn.	5.88	5.25
2	GOST is intuitive to use.	5.00	5.75
3	It was easy to recover when making an error using GOST.	5.75	4.50
4	GOST aided in the ability to assess uncertainty inherent in final product.	4.13	3.00
5	GOST aided in the ability to meet tasking requirements.	5.88	3.25
6	GOST increased the speed with which products are created.	5.00	3.25
7	GOST will help a less experienced analyst understand the workflow.	5.75	3.75
8	GOST reduced overall workload.	6.13	6.00
9	GOST could be effective in analyst training.	5.75	4.50
10	GOST provides capabilities that are currently unavailable to me.	5.75	6.50
11	GOST would quickly allow me to determine the relevancy of source material.	6.75	6.50
12	I can see the applicability of GOST capabilities to my work flow.	5.00	6.00
13	GOST will be accepted by analysts.	3.63	5.50
14	How motivated are you to continue to learn and use the system?	6.00	5.25
15	Overall, how does using GOST compare to current methods for the tasks completed today?	5.75	3.75
16	What functions does GOST provide that are helpful?	5.50	5.25
17	The system taxed my memory during use.	3.13	3.50
18	The system matched my mental model of online experiences.	5.88	3.50
19	I was satisfied with the overall task experience.	5.38	4.00
20	GOST will help a less experienced analyst understand the workflow.	6.33	5.00
21	GOST could be effective in analyst training.	6.67	6.00

1----------2----------3----------4----------5----------6----------7

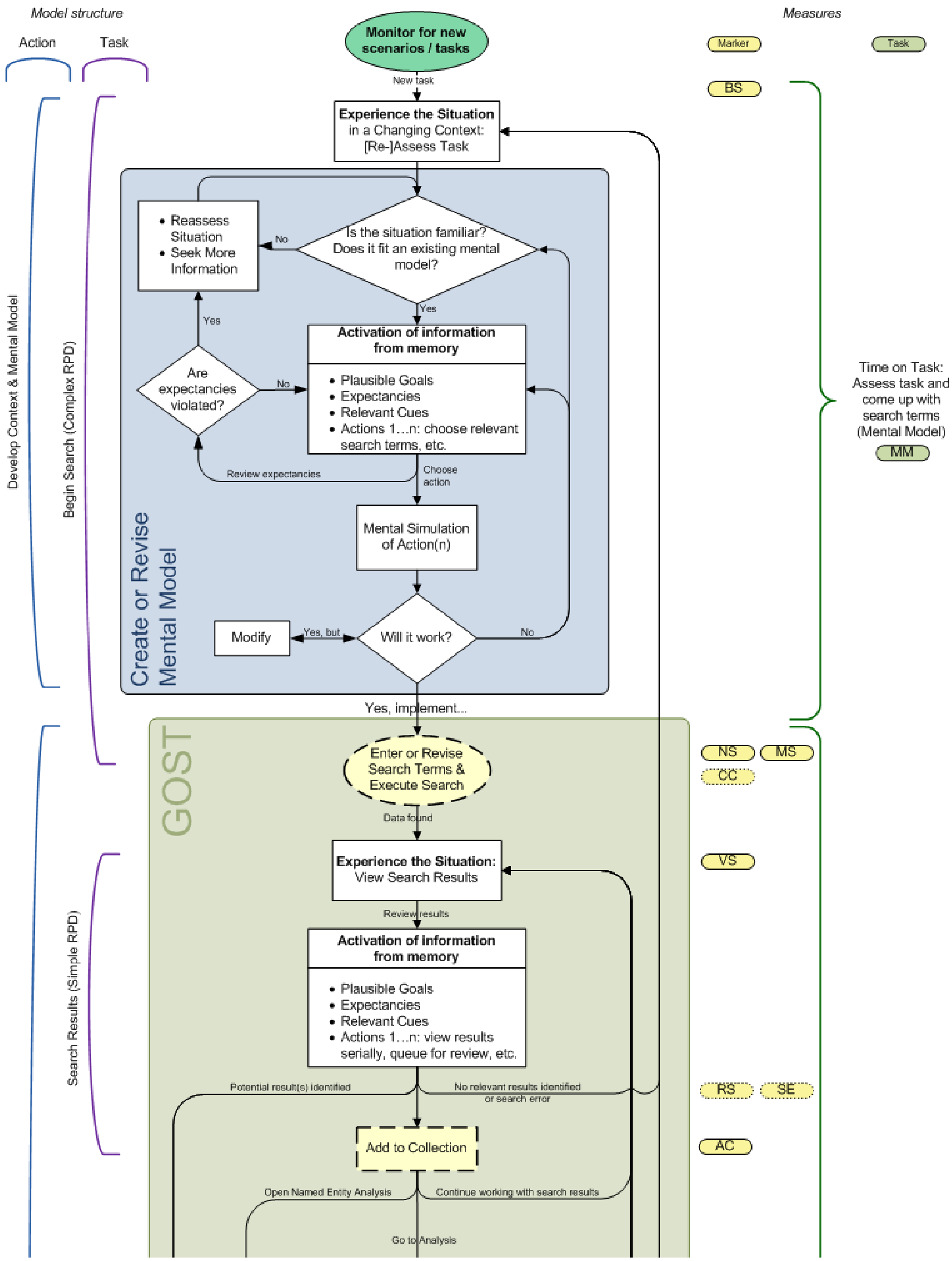
Strongly Disagree   Neutral   Strongly Agree

5.2 Model

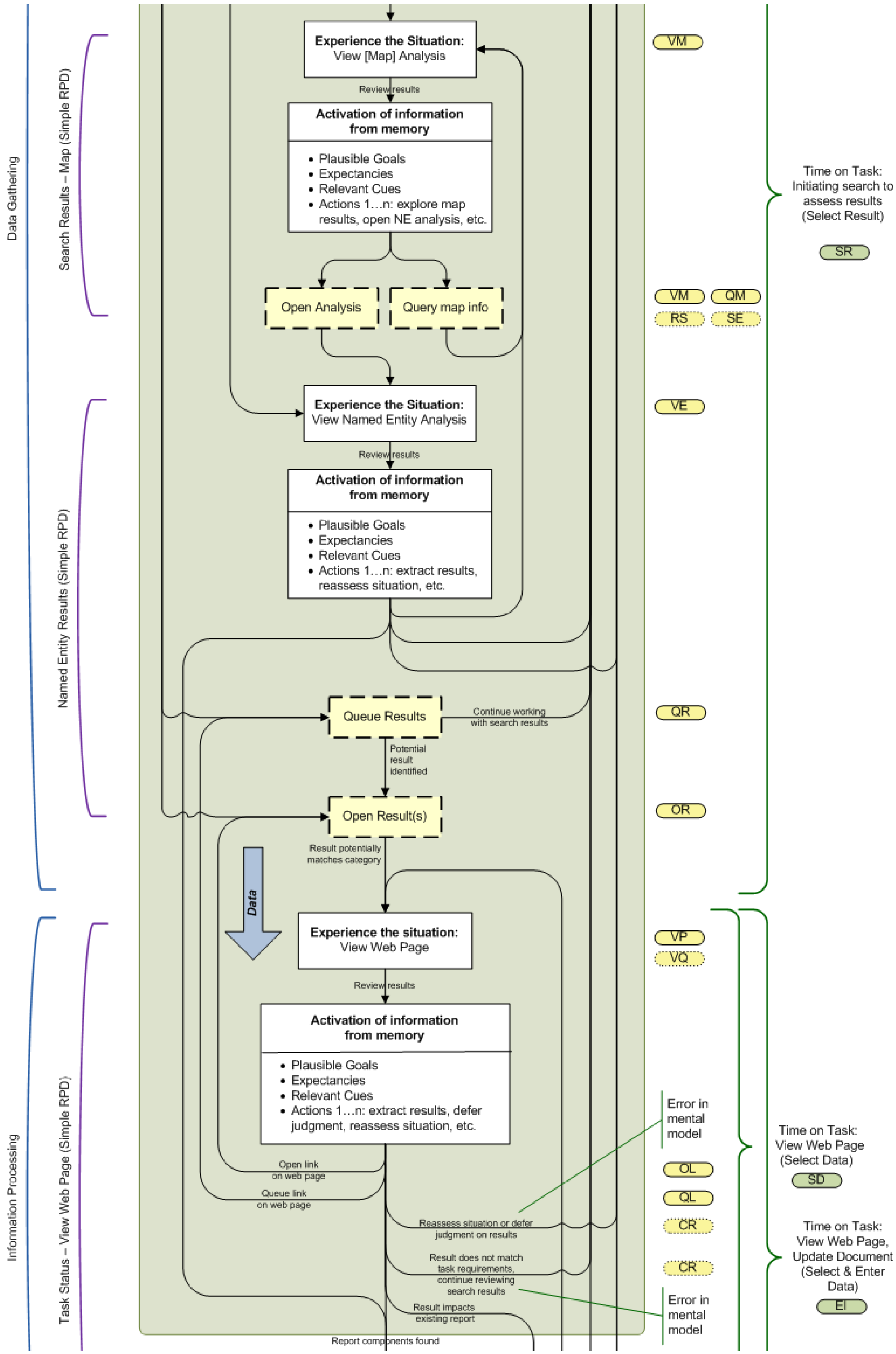
Analysis of Morae data provided for final revision and validation of the process model. The Morae session data for each participant was reviewed and annotated with the model markers. The time stamp information from Morae was combined with the model markers to allow the researcher to correlate participant actions with the process model. The final model provides additional affordances and metrics that allow for supplementary insight into the analyst process as well as more detailed analysis for the researcher.

5.2.1 Final Model

Figure 13 shows the final analyst process model. The model structure indicates the actions being taken by the analyst as well as the type of RPD model being employed. The measures indicate the MOPs and task information gathered to aid in research analysis. In addition to the affordances provided by the revised process model shown in Figure 5, the final model allows for tracking unconstrained participant actions, additional tracking measures, and insight into how analysts move between task sections.



Model structure
Action
Task
Measures
Marker
Task
Monitor for new scenarios / tasks
New task
Experience the Situation in a Changing Context: [Re-]Assess Task
BS
Develop Context & Mental Model
Begin Search (Complex RPD)
Reassess Situation
Seek More Information
No
Is the situation familiar? Does it fit an existing mental model?
Yes
Yes
Activation of information from memory
Are expectancies violated?
No
Plausible Goals
Expectancies
Relevant Cues
Actions 1…n: choose relevant search terms, etc.
Review expectancies
Choose action
Mental Simulation of Action(n)
Create or Revise Mental Model
Modify
Yes, but
Will it work?
No
Time on Task: Assess task and come up with search terms (Mental Model)
MM
Yes, implement...
GOST
Enter or Revise Search Terms & Execute Search
NS
MS
CC
Data found
Experience the Situation: View Search Results
VS
Review results
Activation of information from memory
Plausible Goals
Expectancies
Relevant Cues
Actions 1…n: view results serially, queue for review, etc.
Search Results (Simple RPD)
Potential result(s) identified
No relevant results identified or search error
RS
SE
Add to Collection
AC
Open Named Entity Analysis
Continue working with search results
Go to Analysis


Data Gathering
Search Results – Map (Simple RPD)
Named Entity Results (Simple RPD)
Information Processing
Task Status – View Web Page (Simple RPD)
Experience the Situation: View [Map] Analysis
Review results
Activation of information from memory
• Plausible Goals
• Expectancies
• Relevant Cues
• Actions 1…n: explore map results, open NE analysis, etc.
Open Analysis
Query map info
Experience the Situation: View Named Entity Analysis
Review results
Activation of information from memory
• Plausible Goals
• Expectancies
• Relevant Cues
• Actions 1…n: extract results, reassess situation, etc.
Queue Results
Continue working with search results
Potential result identified
Open Result(s)
Result potentially matches category
Data
Experience the situation: View Web Page
Review results
Activation of information from memory
• Plausible Goals
• Expectancies
• Relevant Cues
• Actions 1…n: extract results, defer judgment, reassess situation, etc.
Open link on web page
Queue link on web page
Reassess situation or defer judgment on results
Result does not match task requirements, continue reviewing search results
Result impacts existing report
Report components found
VM
VM
QM
RS
SE
VE
QR
OR
VP
VQ
Error in mental model
OL
QL
CR
CR
Error in mental model
Time on Task: Initiating search to assess results (Select Result)
SR
Time on Task: View Web Page (Select Data)
SD
Time on Task: View Web Page, Update Document (Select & Enter Data)
EI

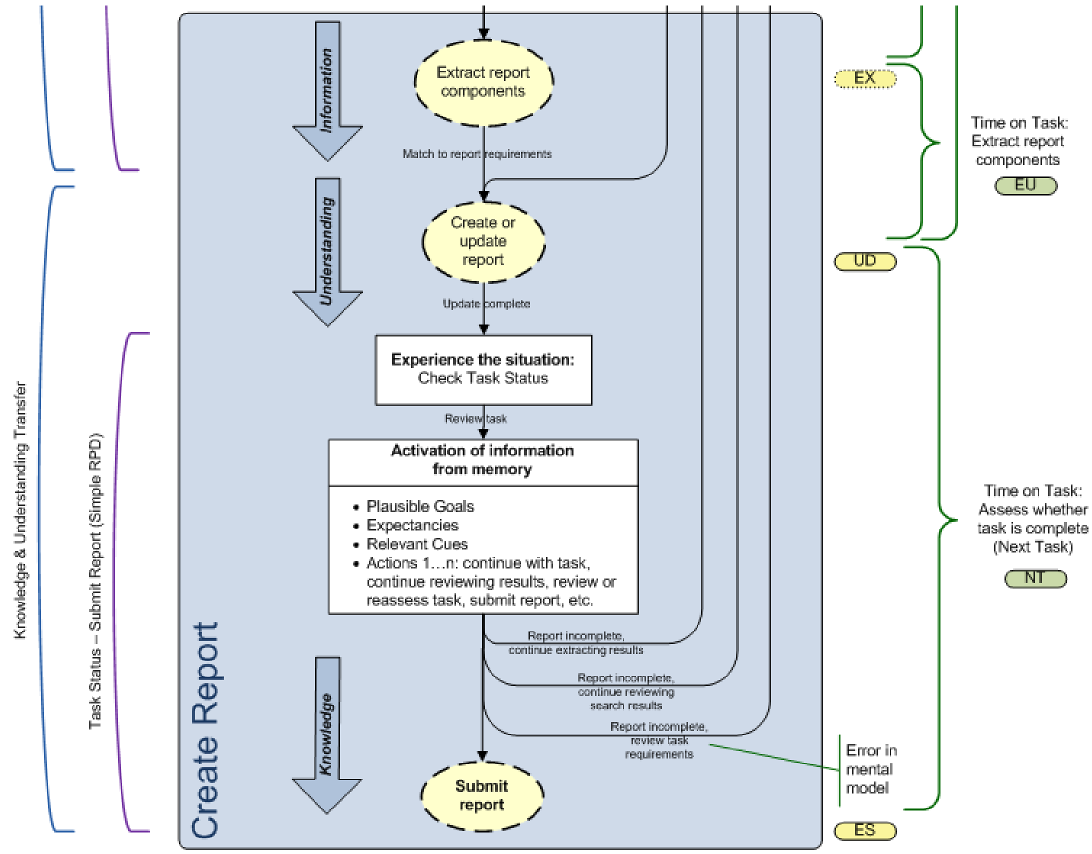


Figure 13: Final Analyst Process Model

Figure 14 shows actions and corresponding measures which are not constrained in regard to when they happen during the task. Actions are performed as needed during the task and are grouped by type. Yellow items indicate participant actions while the grey items are researcher actions.

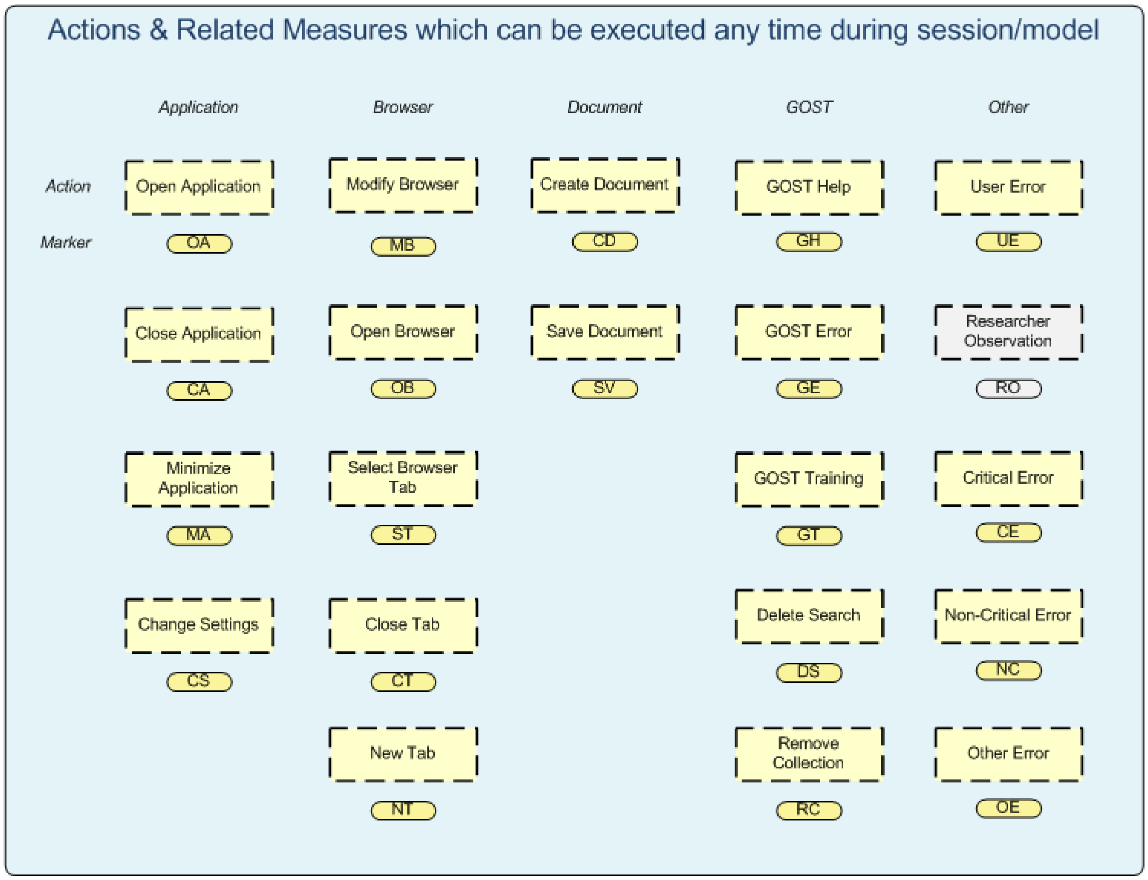


Figure 14: Unconstrained Actions & Related Measures

5.2.2 Time on Task

The time to complete a task element is referred to as "time on task." It is measured from the time the person begins the scenario task element to the time he/she completes or abandons the task. This data was derived from applying model labels to Morae data. Table 17 lists the task labels used in the model along with a description of each. As shown in the process model (Figure 13), the Extract Information (EI) task is a combination of Extract data and Update document (EU) and Select Data (SD). This combined task labeling resulted from the analysis of participant behavior, but the EI task category is not necessary for subsequent data analysis.

Table 17: Model Task Labels & Descriptions

Task	Task Description
EI	Extract Information / update document
EU	Extract data and update document
MM	Mental Model
NC	Not Classified
NT	Next Task
SD	Select Data
SR	Select Result

Each scenario task required that the participant search for relevant data using the given toolset. Scenario Task Completion measures the ability of the participant to complete the given task elements and was analyzed forensically. As part of the model markers, two types of errors were tracked, critical and non-critical. A critical error prevents the user from completing the task and a non-critical error causes user difficulty, but the task can be completed.

Table 18 summarizes the task breakdown comparison between the baseline and GOST toolsets. Figures 15 and 16 give visual representations of the data in Table 18. An ANOVA (α = 0.05, F Ratio = 38.2804, Prob > F = <0.0001) indicates that there is evidence to support the conclusion that SD and SR task types are significantly different. An ANOVA (α = 0.05, F Ratio = 3.7787, Prob > F = 0.0723) indicates that the difference for EU is marginally significant.

Table 18: Task breakdown by toolset

	Baseline		*GOST*	
Task	**Time**	**%**	**Time**	**%**
EU	2:20:57.8	38.5%	1:30:49.3	24.8%
MM	0:01:10.4	0.3%	0:02:13.2	0.6%
NC	0:11:25.3	3.1%	0:30:21.8	8.3%
NT	0:01:14.5	0.3%	0:01:34.9	0.4%
SD	2:18:03.8	37.7%	0:57:13.9	15.6%
SR	1:13:23.7	20.0%	3:08:17.5	51.4%

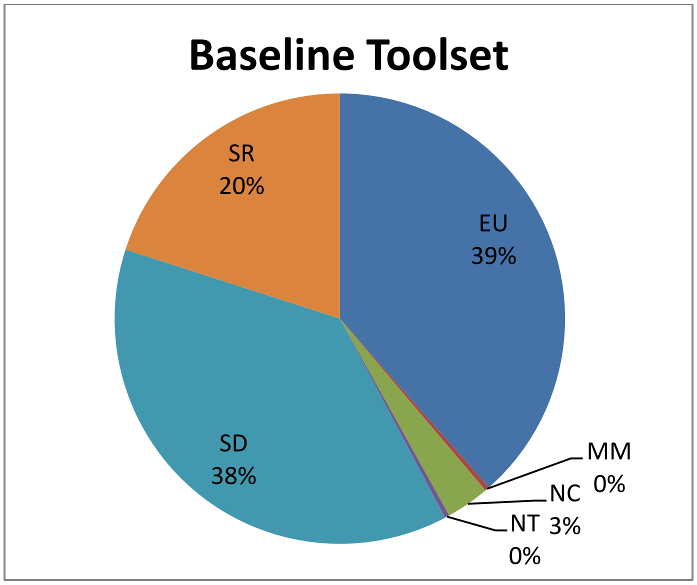


Figure 15: Task breakdown for baseline toolset

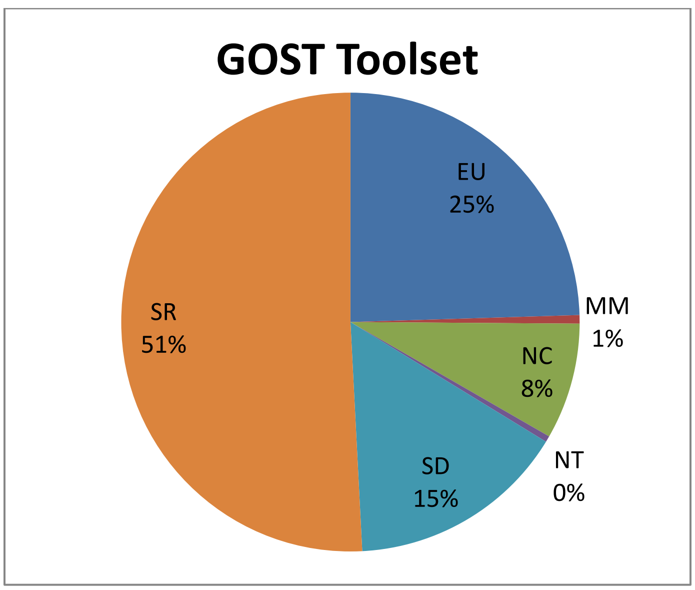


Figure 16: Task breakdown for GOST toolset

Table 19 summarizes the task breakdown comparison between the novice and expert levels of expertise. Figures 17 and 18 give visual representations of the data in Table 19. There were no significant differences due to level of expertise.

Table 19: Task breakdown by expertise level

	Novice		***Expert***	
Task	**Time**	**%**	**Time**	**%**
EU	2:04:49.8	34.0%	1:46:57.3	29.1%
MM	0:01:09.3	0.3%	0:02:14.3	0.6%
NC	0:18:03.5	4.9%	0:23:43.7	6.5%
NT	0:01:25.3	0.4%	0:01:24.1	0.4%
SD	1:28:48.7	24.2%	1:46:28.9	29.0%
SR	2:12:57.3	36.2%	2:08:43.9	35.1%

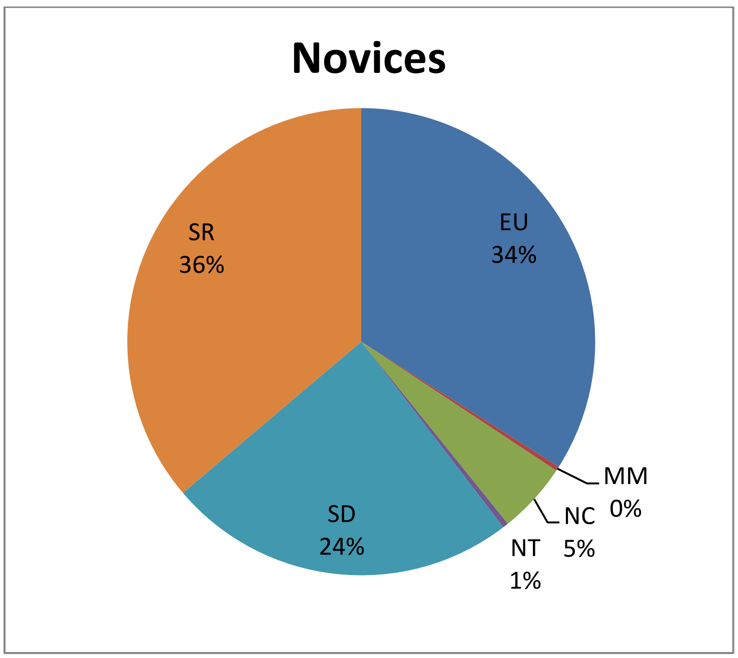


Figure 17: Task breakdown for Novices

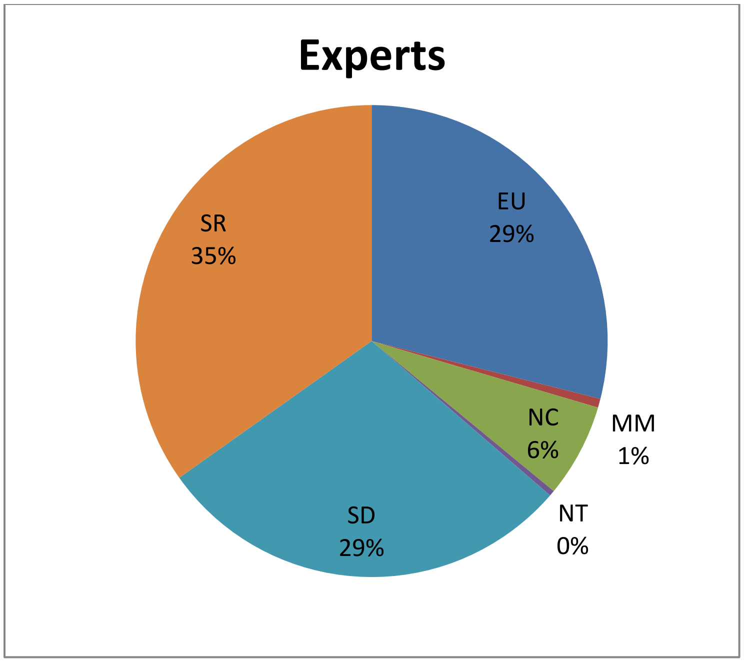
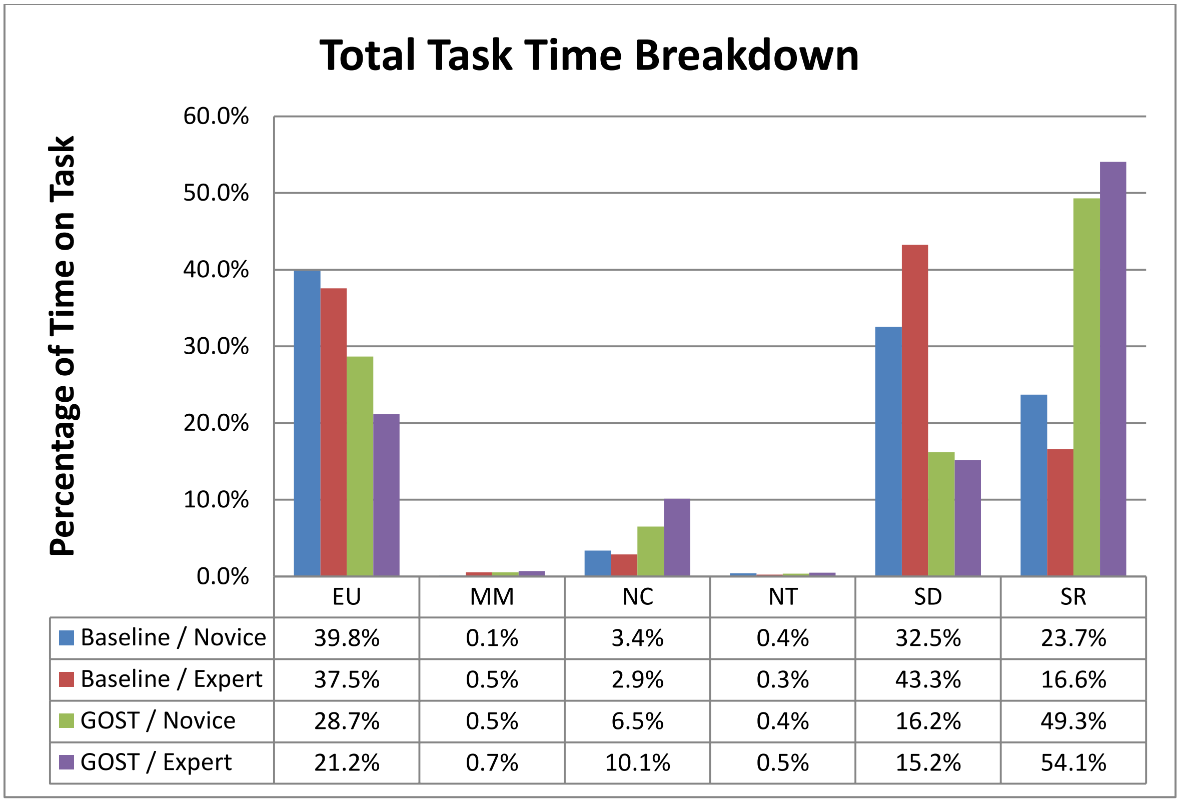


Figure 18: Task breakdown for Experts

Figure 19 summarizes the task on task information. Use of toolset had a notable impact on the amount of time participants spent on each task type. As shown in Figure 19, expertise does not have a meaningful impact on task time but toolset use has a substantial impact on the Select Data (SD) and Select Result (SR) task categories and a marginal impact on the Extract data and Update document (EU) category. Reviewing the location of these task types in the model indicates that the greater time spent on Select Result (SR) tasks may be due to using GOST. As with Mean Error Rates, the unfamiliar system may cause the participant to require more time to complete this task. Additional analysis or follow-on studies could provide more insight into the reason for this result.



	EU	MM	NC	NT	SD	SR
Baseline / Novice	39.8%	0.1%	3.4%	0.4%	32.5%	23.7%
Baseline / Expert	37.5%	0.5%	2.9%	0.3%	43.3%	16.6%
GOST / Novice	28.7%	0.5%	6.5%	0.4%	16.2%	49.3%
GOST / Expert	21.2%	0.7%	10.1%	0.5%	15.2%	54.1%

Figure 19: Task Time Breakdown by Toolset and Expertise

VI. DISCUSSION

6.1 Conclusions and Recommendations

The analyst process model offers visibility into the decision making process that analysts follow as they execute a search. The model indicates how the analyst is able to create and revise their mental model while tracking how they filter, categorize, and extract data, transforming it into information. It shows how the analysts process this information to update their knowledge base, and then integrate and transfer the knowledge to become understanding. This research attempts to aid in the ISR community's understanding of analyst decision making and how to measure and validate performance.

While developers may prefer to have new systems perform better and with fewer errors than existing tools, this is unrealistic while the system is in development and participants are unfamiliar with the toolset. Participants are generally more effective and efficient in producing results using familiar tools in scenarios with which they have experience. While participants were given a training session prior to completing the task, it is likely that lack of familiarity with the new system was the cause for many of the errors. This study found that while the toolset had a significant effect on the report quality of experts, it did not have the same effect on novices. The higher errors rates with GOST may have been due to the lack of participant familiarity with the system as indicated by the post-test questionnaire comments.

For novices, the smaller standard deviation in error rates using GOST along with the smaller difference in report quality and cognitive workload between toolsets may be an indication that novices more readily adapt to new toolsets and may be willing to leverage them as a way to learn a new process or task. With less prior experience as well as fewer heuristics and biases, novices may adapt more readily to new toolsets. This ability of novices to more easily learn and adapt may provide an opportunity for leveraging the process model as a tool for training new analysts.

As far as task breakdown is concerned, the greater amount of time spent in the Select Result (SR) section relative to Select Data (SD) with GOST may indicate a lack of familiarity with the toolset. This, in combination with the lack of significant difference of the report quality based on toolset indicates the potential for increased scores with additional toolset training and acclimation.

With regard to testing new toolsets and software development, the experimental methodology used in this study appears to weigh against new toolsets scoring well in this context, especially with experts who are familiar with the current toolset and search process. A revised methodology may benefit from providing more training on new toolsets prior to testing.

It should be noted that the toolset developers did not benefit from the process model during the software development process. Doing initial research and developing a model to gain more insight into the analyst process could allow developers to better tailor their toolset to the process. Tools such as Google and Bing are generic in the sense that they are not tailored to the analyst process, and because of this, the analyst chooses how to use them within the search process. In contrast, toolsets such as GOST attempt to support the analyst through a broader portion of the search process. While this may provide more affordances to the analyst, it also requires adjustment on the part of the analyst to fully realize the benefits of the enhanced tool. In this respect, toolset development may benefit from an understanding of the analyst process earlier in the development process.

With regard to the methodology, four areas of improvement bear mention. Increasing the number of participants would increase the statistical validity of the results. Also, focusing participant selection to accurately represent the target population would increase the ability to validate the process model as well as gain more accurate feedback related to toolset development. Doing this in combination with conducting studies at various points during software development would more effectively leverage the use of the process model. Finally, providing better toolset training during the experiment would benefit the participants, as well as provide a better understanding of how the toolset can be integrated into the overall training of new analysts. Because of the limited scope of this study, it is unknown whether the results presented here are typical of all new toolsets or specific to GOST. Conducting follow-on studies, along with a meta-analysis, with the same structure using both GOST and other search toolsets would lend greater statistical validity to the results.

Previous research has shown that utilizing system analysis and evaluation during the software development process can result in improved performance. The goal of this research was to study the performance of experts and novices along with the impact of toolsets in completing representative search tasks. The contributions of this research include (1) providing feedback to software development regarding toolset performance, (2) providing insight into the analyst search process through the development of a process model, (3) establishing a model framework for adding performance metrics, (4) providing insight into the differences between experts and novices in conducting a search task, and (5) providing a basis for developing analyst training related to search tasks and toolset use. The results of this study provide a better understanding of the impact of expertise and toolsets on analyst performance and may provide the basis for future research in the geospatial and open source domains. This could also be useful in extending the research into other analyst processes to aid in developing and integrating new toolsets to improve analyst performance.

In conclusion, analyst performance in the context of searching for relevant information in the data transformation process with new toolsets lends itself to study using cognitive design principles along with usability tools and metrics. These principles and tools can aid in toolset development and implementation by identifying inefficient actions and providing insight into current analyst processes and behaviors. Combining this information as part of the software development process can ultimately foster timely integration of new toolsets and improve analyst performance.

VII. APPENDIX A: Informed Consent

Attachment A: Information Protected By The Privacy Act of 1974

Informed Consent Document
For
Investigation of Potential Capability Improvements for Intelligence Analysts

IRB Director: William Butler, Col, 711 HPW/IR, Commercial 937-656-5436, William.Butler2@wpafb.af.mil

IRB Deputy Director: Kim London, Civ, 711 HPW/IR, Commercial 937-656-5688, Kim.London@wpafb.af.mil

Principal Investigator: Dr. Lisa Tripp, DR-II, 711 HPW/RHAS, Commercial 937-938-4032, Lisa.Tripp@wpafb.af.mil

Associate Investigators: Dr. Geoffrey Barbier, DR-III, 711 HPW/RHAS, Commercial 937-938-3562, Geoffrey.Barbier@wpafb.af.mil

Dr. Paul Havig, DR-III, 711 HPW/RHCV, Commercial 937-255-3951

Dr. Ben Knott, DR-III, 711 HPW/RHCP, Commercial 937-938-3599, Benjamin.Knott@wpafb.af.mil

Vic Finomore, DR-II, 711 HPW/RHCB, Commercial 937-904-7123 Victor.Finomore@wpafb.af.mil

Dr. Matthew Valenti, DR-II, 711 HPW/RHXM, Commercial 937-798-4391

Jennifer Lopez, DR-I, 711 HPW/RHXM, Commercial 937-255-9972, Jennifer.Lopez@wpafb.af.mil

Ashley Alexander, Lt, 711 HPW/RHAS, Commercial 937-938-2843, Ashley.Alexander@wpafb.af.mil

Robert Nelson, Lt, 711 HPW/RHAS, Commercial 937-938-4037, Robert.Nelson@wpafb.af.mil

Elliot Humphrey, Lt, 711HPW/RHAS, Commercial 937-938-4021 Elliot.Humphrey@wpafb.af.mil

Kevin Durkee, Ctr, Aptima Inc., Commercial 937-490-8010, kdurkee@aptima.com

Mary Fendley, Ctr, Wright State University, Commercial 937-781-2444, mary.fendley@wright.edu

Ali Reiter, Ctr, SAIC, Commercial 937-241-0351, ali.k.reiter@saic.com

Anna Maresca, Ctr, Wright State Research Institute, Commercial 937-705-1021, anna.maresca@wright.edu

Ositadimma Eziolisa, Ctr, Wright State University, Cell 937-231-3423, eziolisa.2@wright.edu

Jennifer Winner, Ctr, Lumir Research Institute, Commercial 937-938-4016, Jennifer.winner.ctr@wpafb.af.mil

George Reis, DR-II, 711 HPW/RHCV, Commercial 937-255-8863, George.reis@wpafb.af.mil

Sharon Ulring, Ctr, SRA, Commercial 937-910-6484, Shari_Ulring@sra.com

Karl Hendrickson, Wright State University, Commerical (937) 425-0745, karl.hendrickson@wright.edu

Adam Hoenle, Wright State University, Commerical (937) 320-0966, karl.hendrickson@wright.edu

1. **Nature and purpose**: You have been offered the opportunity to participate in the "Investigation of potential capability improvements for intelligence analysts" study. The purpose of this research is to evaluate new training techniques and technologies that may result in capability improvements for intelligence analysts, and additionally identify problem areas and potential solution paths for developers, the acquisition community, and end users.

The time requirement for each volunteer participant is anticipated to be a total of 1 to 10 visits of approximately 1 hour to 12 hours, with a maximum of participation time of three consecutive 12 hour days per 7 days work week for up to three weeks. A total of approximately 600 participants may be enrolled in this experiment. In order to participate, you must have normal or corrected to normal vision. At the beginning of the study, a number of eye and/or hearing tests may be administered. You may be excluded from the study if your vision and/or hearing do not test as normal (or corrected to normal). Subjects may be unpaid volunteers that are Department of Defense employees, active duty personnel, or contractors, as well as students attending the Advanced Technical Intelligence Center (ATIC) or Wright State University. Although there are no stated requirements regarding gender, we anticipate an approximately equal ratio of male to female subjects. Subjects will be adults 18 and older.

2. **Experimental procedures**: If you decide to participate, you will be asked to participate in a number of scenarios which are designed to simulate typical tasking of intelligence analysts. Tasks may include active tasks such as tracking of high value targets, performing a visual search of a road for cues associated with IED detection, and performing threat detection such as in a Blue Force overwatch, and forensic tasks such as aggregation of information from multiple intelligence sources for report generation and prediction of future events based on multiple missions. While performing these tasks your reaction time, mission completion time, report generation time, accuracy, number of errors, number of mission objects met, chat session, direction of gaze, electrocardiography (ECG), and respiration rate may be recorded. To record your responses you will be asked to provide input via a mouse, joystick, or keyboard. Prior to performing the task, or immediately following the task, the experimenter may also ask you a series of questions and/or ask you to fill out questionnaires to assess the task workload, fatigue, trust in the computer system, situational awareness, or usability. These questions are designed to elicit information to inform the development of training, procedures, technologies to decrease workload and fatigue associated with the tasks while increasing trust in the system, situational awareness, and system usability. The information collected will not be used as a personal reflection on you or your performance of the task. The types of questions you may be asked involve the degree of difficulty, frustration, and fatigue associated with the task and the degree to which you found the system easy to use and reliable. No personal data will be requested of you. Prior to beginning the experiment, the experimenter will provide you with a document detailing your task for this experiment (e.g., which buttons to press on the input device, etc.). The experimenter will also verbally describe the task. If you have any questions regarding the procedure please feel free ask the experimenter at any time.

   You will be seated in a chair in an air conditioned room. Your participation may be a maximum of tweleve hours per day for no more than ten days and no more than three consecutive days per 7 day work week.

   Opportunities for rest breaks will be given at the end of each set of scenarios. Should you require additional rest breaks at any time, please inform the experimenter and he or she will pause the experiment. Restrooms, water, and vending machines are available. Should you feel uncomfortable at any time or wish to discontinue the experiment for any reason, please inform the experimenter and he or she will end the experiment.

3. **Discomfort and risks**: There are minimal risks in participating in this study including eye strain, headache, and exhaustion. Risk and discomfort levels should be comparable to work tasks at a computer. Some of these symptoms may be in result of sitting there too long, but breaks will be offered. Preventative measures you may take include proper posture while sitting/standing, frequent breaks, and wearing proper corrective lenses, if applicable. If at any time you feel uncomfortable please let the experimenter know and he/she will stop the experiment.

4. **Precautions for female subjects, or subjects who are or may become pregnant during the course of this study**: There are no known additional precautions required for female participants.

5. **Benefits**: The benefits of participating in this study are contribution to the intelligence community and knowing that you are making a difference in the futuring training of Air Force military and civilians. Other personal gains may result from the physiological measures that are conducted.

6. **Compensation**: Participation in this experiment is entirely voluntary. Choosing not to participate is your alternative to participating. There are no penalties for withdrawing for any reason. Participants who are active duty, USAF contract support and USAF government employees will not be compensated for participation. Local community volunteers and ATIC students will receive $15 per hour. Wright State University students will receive either course credit or compensation at the aforementioned rate.

7. **Entitlements and confidentiality**:

   a. Records of your participation in this study may only be disclosed according to federal law, including the Federal Privacy Act, 5 U.S.C. 552a, and its implementing regulations. Your personal information will be stored in a locked cabinet in an office that is locked when not occupied. Electronic files containing your personal information will be password protected and stored only on a DoD server. It is intended that the only people having access to your information will be the researchers named above and the AFRL Wright Site IRB or any other IRB involved in the review and approval of this protocol. When no longer needed for research purposes your information will be destroyed in a secure manner (shredding). Complete confidentiality for military personnel cannot be promised because information bearing on your health may be required to be reported to appropriate medical or command authorities.

   Your entitlements to medical and dental care and/or compensation in the event of injury are governed by federal laws and regulations, and that if you desire further information you may contact the base legal office (ASC/JA, 257-6142 for Wright-Patterson AFB).

   b. The decision to participate in this research is completely voluntary on your part. No one may coerce or intimidate you into participating in this program. You are participating because you want to. Dr. Lisa Tripp, or an associate, has adequately answered any and all questions you have about this study, your participation, and the procedures involved. Dr. Lisa Tripp can be reached at (937) 938-4030. Dr. Lisa Tripp or an associate will be available to answer any questions concerning procedures throughout this study. If significant new findings develop during the course of this research, which may relate to your decision to continue participation, you will be informed. You may withdraw this consent at any time and discontinue further participation in this study without prejudice to your entitlements. The investigator or medical monitor of this study may terminate your participation in this study if she or he feels this to be in your best interest. If you have any questions or concerns about your participation in this study or your rights as a research subject, please contact Col Butler at william.butler2@wpafb.af.mil, (937) 656-5436 or Ms. London at kim.london@wpafb.af.mil, (937) 656-5688.

c. Limited personal information will be collected. This may include your age, gender, and visual screening results. This information will be kept in a password protected electronic database and will remain there for approximately five (5) years. No personal information will be stored on removable storage devices, laptops, or personal computers. Data collected from you will not be stored with identifying information but will be coded by the experimenter. Subject number will be generated using a hash code method. This is the same method that is used to encrypt passwords on many websites. Participants will be asked to answer five questions. An algorithm will take those responses and output a code. All data will be stored using this code. The answers to the questions will be deleted. This minimizes the risk that the data would be traced back to a specific individual and facilitates tracking correlated pieces of data This data will also be stored in a password protected electronic database and will remain there indefinitely.

d. Your participation may be audio/video-taped during segments of this study which require you to interact with computers and/or other experimental apparatus. The audio/video recordings will be used as a part of the data collection and may be included in the final data analysis. There will be no final identifying features to link you back to the audio recording as your audio recording will be coded such that your identity will be known only to the experimenter. The audio/video recordings and the identifying coding will be stored on a password protected computer and transcribed into text files within two months of data collection. As soon as these files are transcribed the audio/video recordings will be deleted. You consent to the use of these media for training and data collection purposes. Any release of records of your participation in this study may only be disclosed according to federal law, including the Federal Privacy Act, 55 U.S.C. 552a, and its implementing regulations.

This means personal information will not be released to unauthorized source without your permission. These recording may be used for presentation or publication. They will be stored in a locked cabinet in a room that is locked when not occupied. Only the investigators of this study will have access to these media. They will be maintained for 5 years.

YOU ARE MAKING A DECISION WHETHER OR NOT TO PARTICIPATE. YOUR SIGNATURE INDICATES THAT YOU HAVE DECIDED TO PARTICIPATE HAVING READ THE INFORMATION PROVIDED ABOVE.

Volunteer Signature____Date_______________

Volunteer Name (printed)____

Advising Investigator Signature ______________________ Date _________________

Investigator Name (printed)____

Witness Signature __________________________________Date _________________

Witness Name (printed)____

Privacy Act Statement

Authority: We are requesting disclosure of personal information. Researchers are authorized to collect personal information on research subjects under The Privacy Act-5 USC 552a, 10 USC 55, 10 USC 8013, 32 CFR 219, 45 CFR Part 46, and EO 9397, November 1943.

Purpose: It is possible that latent risks or injuries inherent in this experiment will not be discovered until sometime in the future. The purpose of collecting this information is to aid researchers in locating you at a future date if further disclosures are appropriate.

Routine Uses: Information may be furnished to Federal, State and local agencies for any uses published by the Air Force in the Federal Register, 52 FR 16431, to include, furtherance of the research involved with this study and to provide medical care.

Disclosure: Disclosure of the requested information is voluntary. No adverse action whatsoever will be taken against you, and no privilege will be denied you based on the fact you do not disclose this information. However, your participation in this study may be impacted by a refusal to provide this information.

VIII. APPENDIX B: Pre-Test Questionnaire

Analyst initials: ____

Background / experience:

List the tools (if any) you have used in the following areas:

(Circle or underline the tool you most commonly use.)

1. Geospatial:

2. Entity extraction:

3. Gazetteer:

4. Content management:

5. Temporal / Timeline:

IX. APPENDIX C: Post-Test Questionnaire

Please answer the following regarding the GOST system that you used. Please provide comments whenever possible. When making comparisons, please compare to current practices or methods.

1. GOST is easy to learn.

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Comments:

2. GOST is intuitive to use.

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Comments:

3. It was easy to recover when making an error using GOST.

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Comments:

4. GOST aided in the ability to assess uncertainty inherent in final product.

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Comments:

5. GOST aided in the ability to meet tasking requirements.

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Comments:

6. GOST increased the speed with which products are created.

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Comments:

7. GOST reduced overall workload.

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Comments:

8. GOST provides capabilities that are currently unavailable to me.

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Comments:

9. GOST would quickly allow me to determine the relevancy of source material.

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Comments:

10. I can see the applicability of GOST capabilities to my work flow.

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Comments:

11. I was motivated to learn and use the system.

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Comments:

12. The user interface has aesthetic appeal.

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Comments:

13. I was frustrated using the system.

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Comments:

14. GOST completed searches quickly.

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Comments:

15. I was satisfied with my results.

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Comments:

16. The system became easier to use over the course of the session.

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Comments:

17. The system taxed my memory during use.

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Comments:

18. The system matched my mental model of online experiences.

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Comments:

19. I was satisfied with the overall task experience.

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Comments:

Questions for intelligence analysts only:

20. GOST will help a less experienced analyst understand the workflow.

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

Comments:

21. GOST could be effective in analyst training.

1----------2----------3----------4----------5----------6----------7

Strongly Disagree Neutral Strongly Agree

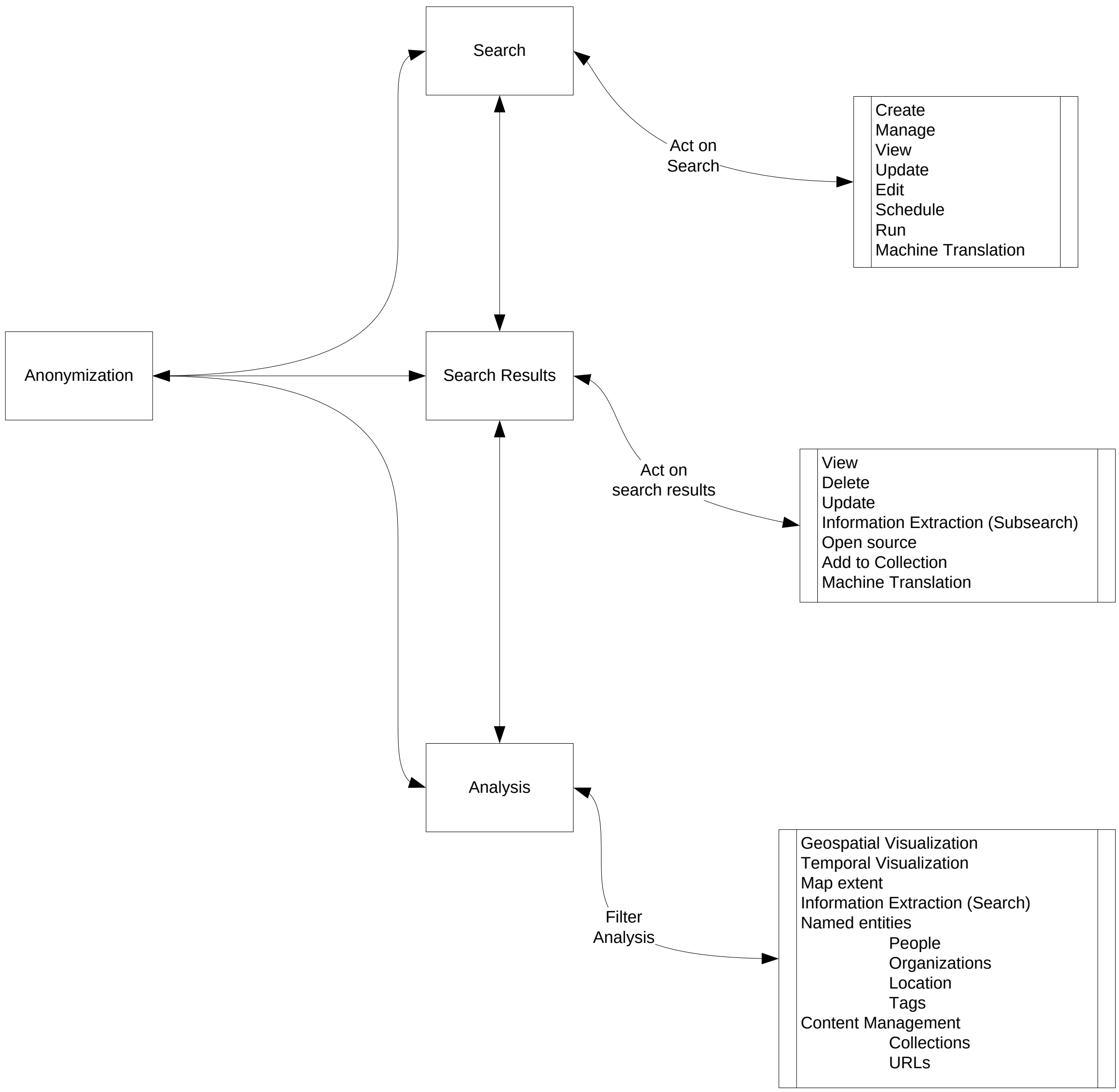
Comments:

22. What functions does GOST provide that are helpful?

23. Overall, how does using GOST compare to current methods for the tasks completed today?

(For example, how do users prioritize their actions? What design features and functions served as barriers to task completion? Which tool functions were most difficult to use? What tools would be useful to incorporate? Which functions are time sensitive?)

X. APPENDIX D: Function Analysis



XI. APPENDIX E: Stealth Task Scenario

Next Generation Stealth Aircraft Scenario

1. **Introduction and Scenario Background:** With the introduction of a new generation of fighter aircraft, 21st century air operations are transforming well beyond their traditional roles of air superiority, air defense, air dominance, strike, and support. The next generation (also referred by many as fourth or fifth-generation) aircraft incorporating advanced airframe design, stealthy technologies, advanced avionics, thrust vectoring, supercruise, and the like is having a significant impact in the role of air operations in support of air, ground and maritime operations. In fact, the current fourth and fifth-generation aircraft being developed and tested have already forced many services to face the challenge of transforming classic or formulating new roles, missions, and countermeasures. As next generation aircraft enter service in larger numbers, they will generate not only greater firepower (both kinetic and non-kinetic), but enable greater interoperability through enhanced connectivity, intelligence, surveillance, and reconnaissance (ISR), communications, and computational capabilities. These enhanced capabilities afforded the air assets to connect air, ground, and maritime forces throughout the battlespace will dramatically improved the decision-makers ability to make informed decision, distribute information, and shape the fighting force to meet combat objectives.

2. **Scenario:** Since the 2000, the web has seen a significant increase in posted articles, journals, magazines, videos, sketches, and photographs describing the development of next generation fighters employing stealthy technologies, high performance engines, advanced avionics, etc. These postings are no longer the exclusively associated with the United States. The employment of the next generation aircraft could be used to suppress our ability to use regional bases, airspace, or seas; level the playing field of competitors employing stealth and advanced avionics; force changes in combat strategies and force employment; and impact the use of beyond-visual-range (BVR) missiles. With the flights of these next generation aircraft, multiple countries, have demonstrated a national resolve to domestically and/or cooperatively, developed advance aerospace technologies, and the intent to deploy world-class stealthy aircraft.

3. **Scenario Details:** Post 2005, several countries have designed and flown next generation stealthy prototypes. These fights were an important strategic milestone in their country's next generation development programs. The flights were the culmination of a long list of technology developmental accomplishments. The flights demonstrated that they have a level of competency to design, construct and demonstrate a state-of-the-art combat aircraft. If these aircraft eventfully achieve deployed status, they could represent a change in the balance in airpower in multiple geographic regions throughout the world. To

achieve deployed status, each of these countries will face a long list of R&D challenges which could manifest themselves as entity relationships, and events with geotemporal considerations. Given the potential challenges ahead of them, the Geospatial Open Source Toolkit (GOST) could better enable the analyst to query, organize and navigate the large data landscape surrounding them, and investigate individual/groupings of documents by entity, events, locations, time, etc. As new content are encountered, these items are digested and merged into the knowledge representation, the situation is monitored for change in the status of their actors, relationships, events, timelines, concepts, etc. The list below provides an overview of the core milestones (not exhaustive) of a development program which could enable a country achieve production of a next generation fighter.

- **a.** Concept Exploration and Solution Analysis
- **b.** Requirements Specifications
- **c.** Design and Performance Data Evaluation
- **d.** Concept Development
- **e.** Concept Evaluation
- **f.** Concept Demonstration
- **g.** Flying Demonstrations
- **h.** Engineering and Manufacturing Development
- **i.** Capability Development and Integrated Flight Test
- **j.** Initial Production
- **k.** Production
- **l.** Deployment

4. **Scenario Objective:** To conduct threat analysis, in particular, that associated with weapon systems and technologies, the analysts must keep abreast of a wide variety of information objects (i.e. entities, concepts, etc) which is a critical adjunct to performing their S&TI analysis on specific weapon systems, technologies, and/or process. These information objects assist the analyst in understanding the content within the context of time and space, monitor situations, and possibly predict events. The objective is to assess the technical feasibility of achieving stealth and speed performance improvement to improve the survivability of air vehicles now through 2025 for next generation and subsequent generation aircraft. Investigate what technologies a foreign power may be developing or deploying that involves the use of speed and stealth to achieve a higher survivability against air defense systems (both air and ground-based systems) from the present time to 2025. The missions to be considered include but not limited to close air support (CAS), air interdiction, and long-range strike. Through analysis, we want to achieve a better understanding of this evolving threat, in particular:

   a. Design methodology - the underlying engineering methods and design philosophy utilized;

   b. Engineering analysis - analytical methods and tools used to design or evaluate a systems performance against operational requirements; and

c. Manufacturing know-how - information that provides detailed manufacturing processes and techniques needed to translate a detailed design into a finished system.

5. **Items of Interest:** This analysis requires a considerable amount of information to understand of the R&D process, technology capabilities/limitations/vulnerabilities, the intent, and the potential threat. This is not an exhaustive list however, the following list provides many of the areas of interest:

   a. What countries are involved in forth/fifth/next generation stealthy aircraft development?
      i. When was the first observance of this interest?
   b. Identify partnerships/collaborations between the various countries involved in the development.
      i. When was the first observance of this interest to collaborate?
   c. Identify entities (i.e. people, organizations, etc) involved in the research and development (R&D) and test and evaluation (T&E) programs.
      i. Where are these entities located?
      ii. When were these parties involved?
   d. Identify when and where did transition occurred between the various states from R&D, T&E, and deployment
   e. Identify flight test information to include:
      i. Individual(s) and organizations involved
      ii. Event dates/times/locations, etc
      iii. Describe the timeline progression
   f. Identify how many prototypes have been developed and:
      i. When and where each were identified (i.e. air show, R&D facility, on a broadcast, flight tests, etc)
      ii. Timeline and map events (i.e. dates/times/locations, etc)
   g. If static display or flight capable
      i. Individual(s) and organizations involved
   h. Identify any reported status changes, delays, technical issues, etc?
   i. Identify the projected number of aircraft to be built.
   j. Identify market countries for projected sales.
   k. Identify projected deployment locations and associated dates

   Create list of bullets to answer the above. Include images, maps, links, video. Create a timeline of events and document the R&D status. Please save all files under my documents.

6. **Additional Background/Guidance Information:**
   a. Investment strategies and plans
      i. Who is developing them (i.e. person, organization, location of person or organization?
      ii. When were they first observed and what was the temporal progression?
   b. Financial responsibilities

        i. Where is the funding coming from (i.e. person, organization, location of person or organization?
        ii. When was funding approved, allocated, transmitted, received
        iii. Estimated values for completed system and subsystems
c. Investors, partnerships and technology transfer
        i. By entity, location(s) and relationship(s) including those cooperating, and other stakeholders
d. Technologies being developed (including specific aircraft subcomponent technologies (i.e. airframe, surface materials, paints, mission sensors, avionics; propulsion, etc.)
        i. Individual(s) and organizations involved
        ii. Event dates/times/locations, etc
e. Near-term R&D needs and priorities
f. Far-term R&D opportunities identified
g. Missions expected to be undertaken by the platforms being developed
h. Capabilities required to complete these missions
i. Projected role(s) and the threat(s) the platform(s) are likely to face.
j. Airframe design and shaping (i.e. stealth shaping, angular, rounded, chin, nose, canopy, etc)
        i. Wing and tailboom configuration (i.e. canted, delta, sweep angle, canards, etc)
        ii. Wing fuselage joining
        iii. Radar-cross sections
        iv. Construction materials and finishes
        v. Engine configuration (i.e. single or multiple)
        vi. Avionics fit
        vii. Weapons fit
        viii. Engine inlets and exhaust outlets configuration
        ix. Engine characteristics (i.e. thrust, fuel, etc)
        x. Landing gear and undercarriage door(s) locations and configuration
k. Remain conscious of evolving nomenclature or concepts
l. New or unique terms, concepts associated with the program(s)
m. Performance and flight characteristics (i.e. combat radius, flight profiles, supersonic, dash, etc)
n. Reporting of status changes, delays, technical issues, etc
o. Identification of entities associated with the program(s) (i.e. researchers, developers, test pilot(s), universities, etc.)
p. Pilot training requirements
q. Transitions between the various states from R&D, T&E, and deployment
r. Preparedness during each the R&D, T&E, and deployment states
s. Foreign sales
        i. To whom (i.e. country(s), organization(s) and individual(s)
        ii. Event dates/times/locations, etc

XII. APPENDIX F: Airlift Task Scenario

Airlift Aircraft Scenario

7. **Introduction and Scenario Background:** Airlift aircraft (aka freight aircraft, freighter, airlifter air freighter, air transport, air cargo, or cargo jet) is a ***fixed-wing*** aircraft (***helicopters will not be addressed***) designed or converted for the carriage of goods, supplies, and personnel. In the case of military operations, airlift aircraft operate across a range of six broad tasks: deployment, employment, redeployment, sustainment, aeromedical evacuation (AE), and military operations other than war, such as foreign humanitarian assistance and noncombatant evacuation operations. Military ***strategic airlift*** (inter-theater), perform a long-haul capability, whereas tactical airlift (intra-theater) provides direct airlift support to ground forces. ***Tactical airlift*** aircraft are designed to be more maneuverable, providing improved low-altitude flight to avoid radar detection for the airdropping of supplies. Within the civilian sector, air cargo or air transport, is a vital component of many international logistic networks, essential to managing and controlling the flow of goods, energy, information and other resources like products, services, and people, from the source of production to the marketplace.

8. **Scenario Details:** The United States has by far the greatest military strategic airlift capacity of any nation in the world. Many countries' armed forces possess little or no strategic airlift capacity, preferring to lease from private-sector firms as needed. Alternatively, groups of nations - especially within formal alliances may choose to pool (i.e. airlift capability consortium) their strategic airlift resources rather than individually duplicating the substantial investment required to purchase and maintain such costly and, in many cases, seldom-used assets. As world politics and economics evolve, and emerging regional power status changes.

9. Since 2005, several countries have designed and flown new or next generation strategic long-range strategic transport aircraft. These fights were an important milestone in their country's next generation transport aviation development programs. The flights demonstrated that they have a level of competency to design, construct and demonstrate a state-of-the-art strategic transport aircraft. If these aircraft eventfully achieve deployed status, they will enhance their strategic lift capabilities, establishing additional capability to intervene in regions to preserve peace, deploy rapid reaction forces, and provide full-spectrum logistics. To achieve deployed status, each of these countries have or will face a long list of R&D and T&E challenges which could be partially observed as entity relationships, and events with geotemporal considerations. Geospatial Open Source Toolkit (GOST) could better enable the analyst to query, organize and navigate the large data landscape surrounding them, and investigate individual/groupings of documents by entity, events, locations, time, etc. As new content are encountered, these items are

digested and merged into the knowledge representation, the situation is monitored for change in the status of their actors, relationships, events, timelines, concepts, etc. The list below provides an overview of the core milestones (not exhaustive) of a development program which could enable a country achieve production of a next generation strategic transport aircraft.

- **a.** Concept Exploration and Solution Analysis
- **b.** Requirements Specifications
- **c.** Design and Performance Data Evaluation
- **d.** Concept Development
- **e.** Concept Evaluation
- **f.** Concept Demonstration
- **g.** Flying Demonstrations
- **h.** Engineering and Manufacturing Development
- **i.** Capability Development and Integrated Flight Test
- **j.** Initial Production
- **k.** Production
- **l.** Deployment

10. **Scenario Objective:** To conduct threat analysis, in particular, that associated with weapon systems and technologies, the analysts must keep abreast of a wide variety of information objects (i.e. entities, concepts, etc) which is a critical adjunct to performing their S&TI analysis on specific weapon systems, technologies, and/or process. These information objects assist the analyst in understanding the content within the context of time and space, monitor situations, and possibly predict events. The objective is to assess the technical feasibility of achieving/improving a strategic airlift capability across the six broad tasks introduced above (i.e. deployment, employment, redeployment, sustainment, aeromedical evacuation (AE), and military operations other than war, such as foreign humanitarian assistance and noncombatant evacuation operations). This direct military connection can also provide support to airborne assault, and provide airborne, airmobile, and conventional ground forces battlefield mobility and forward area resupply. Through analysis, we want to achieve a better understanding of these evolving development, in particular:

    a. Design methodology - the underlying engineering methods and design philosophy utilized;

    b. Engineering analysis - analytical methods and tools used to design or evaluate a systems performance against operational requirements; and

    c. Manufacturing know-how - information that provides detailed manufacturing processes and techniques needed to translate a detailed design into a finished system.

11. **Items of Interest:** This analysis requires a considerable amount of information to understand of the R&D and the T&E process, technology capabilities/limitations/vulnerabilities, the intent, and the potential capability. This is not an exhaustive list however, the following list provides many of the areas of interest:

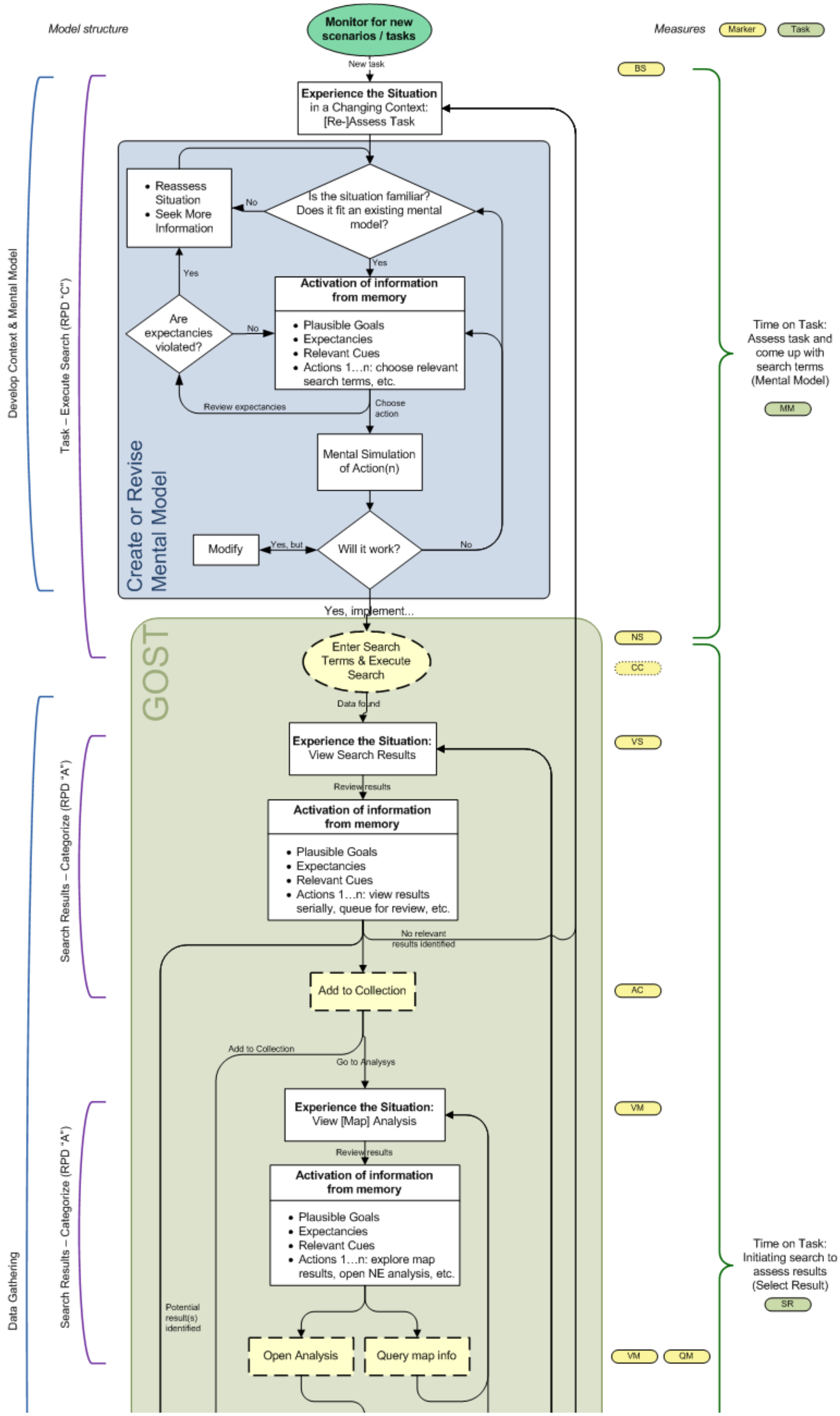
   a. What countries are involved in next generation strategic transport aircraft development?
      i. When was the first observance of the first interest in obtaining the capability?
   b. Identify partnerships/collaborations between the various countries involved in the development.
      i. When was the first observance of this interest to collaborate?
   c. Identify entities (i.e. people, organizations, etc) involved in the research and development (R&D) and test and evaluation (T&E) programs.
      i. Where are these entities located?
      ii. When were these parties involved?
   d. Identify when and where did transition occurred between the various states from R&D, T&E, and deployment
   e. Identify flight test information to include:
      i. Individual(s) and organizations involved
      ii. Event dates/times/locations, etc
      iii. Describe the timeline progression
   f. Identify how many prototypes have been developed and:
      i. When and where each were identified (i.e. air show, R&D facility, on a broadcast, flight tests, etc)
      ii. Timeline and map events (i.e. dates/times/locations, etc)
   g. If static display or flight capable
      i. Individual(s) and organizations involved
   h. Identify any reported status changes, delays, technical issues, etc?
   i. Identify the projected number of aircraft to be built.
   j. Identify market countries for projected sales.
   k. Identify projected deployment locations and associated dates

Create list of bullets to answer the above. Include images, maps, links, video. Create a timeline of events and document the R&D status. Please save all files under my documents.

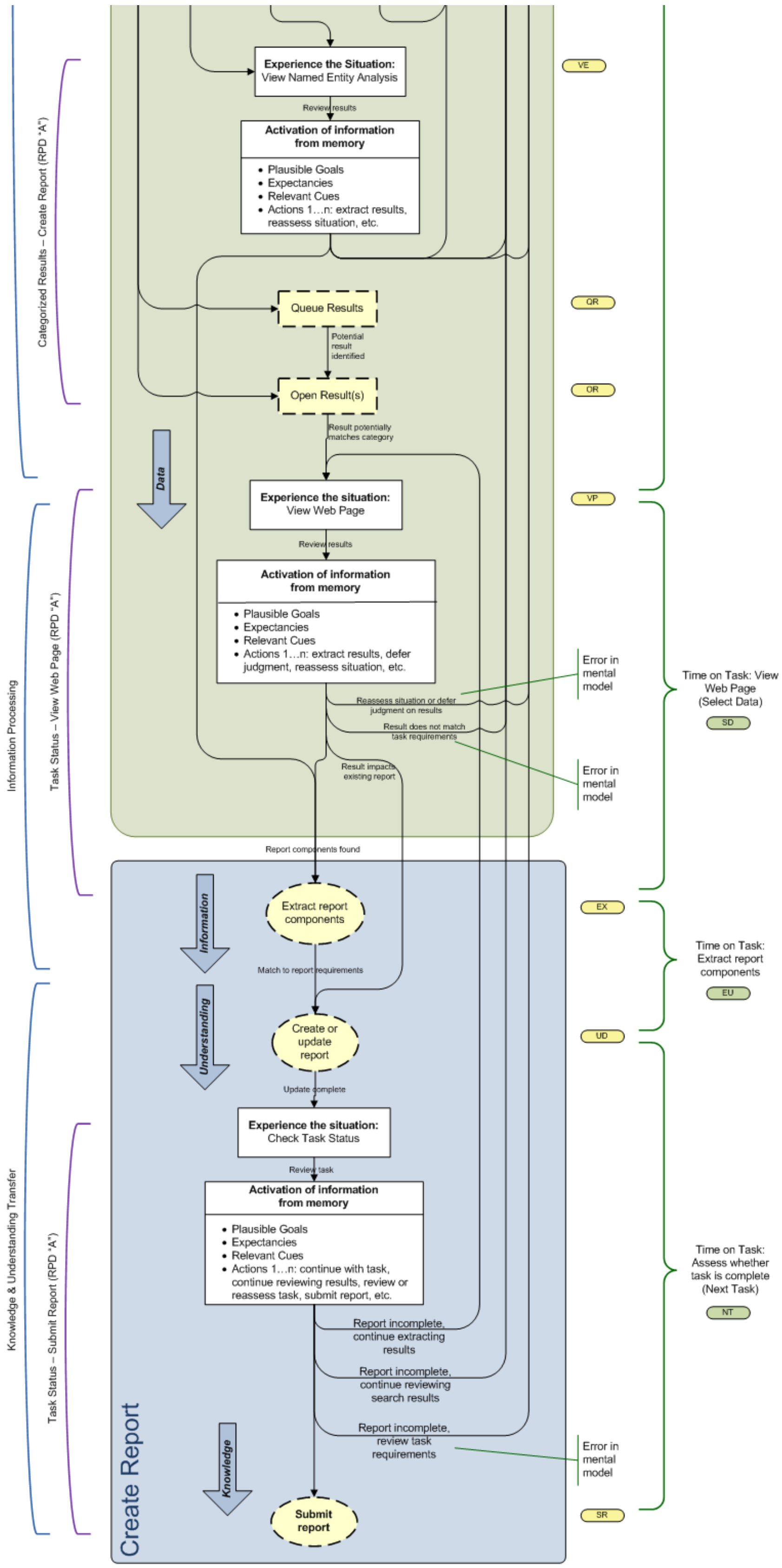
12. **Additional Background/Guidance Information:**
   a. Investment strategies and plans
      i. Who is developing them (i.e. person, organization, location of person or organization?
      ii. When were they first observed and what was the temporal progression?
   b. Financial responsibilities
      i. Where is the funding coming from (i.e. person, organization, location of person or organization?
      ii. When was funding approved, allocated, transmitted, received
      iii. Estimated values for completed system and subsystems

c. Investors, partnerships and technology transfer
    i. By entity, location(s) and relationship(s) including those cooperating, and other stakeholders
d. Technologies being developed (including specific aircraft subcomponent technologies (i.e. airframe, surface materials, paints, mission sensors, avionics; propulsion, etc.)
    i. Individual(s) and organizations involved
    ii. Event dates/times/locations, etc
e. Near-term R&D needs and priorities
f. Far-term R&D opportunities identified
g. Missions expected to be undertaken by the platforms being developed
h. Capabilities required to complete these missions
i. Projected role(s) and the threat(s) the platform(s) are likely to face.
j. Airframe design and shaping (i.e. angular, rounded, chin, nose, canopy, etc)
    i. Wing and tailboom configuration (i.e. canted, sweep angle, canards, etc)
    ii. Wing fuselage joining
    iii. Radar-cross sections
    iv. Construction materials and finishes
    v. Engine configuration (i.e. single or multiple)
    vi. Avionics fit
    vii. Self-protection fit
    viii. Lift Capacity
    ix. Range
    x. Engine inlets and exhaust outlets configuration
    xi. Engine characteristics (i.e. thrust, fuel, etc)
    xii. Landing gear and undercarriage door(s) locations and configuration
k. Remain conscious of evolving nomenclature or concepts
l. New or unique terms, concepts associated with the program(s)
m. Performance and flight characteristics (i.e. flight radius, flight profiles, etc)
n. Reporting of status changes, delays, technical issues, etc
o. Identification of entities associated with the program(s) (i.e. researchers, developers, test pilot(s), universities, etc.)
p. Pilot training requirements
q. Transitions between the various states from R&D, T&E, and deployment
r. Preparedness during each the R&D, T&E, and deployment states
s. Foreign sales
    i. To whom (i.e. country(s), organization(s) and individual(s)
    ii. Event dates/times/locations, etc

XIII. APPENDIX G: Interim Process Model



Distribution A: Approved for public release, distribution is unlimited.


Experience the Situation: View Named Entity Analysis
Review results
Activation of information from memory
• Plausible Goals
• Expectancies
• Relevant Cues
• Actions 1…n: extract results, reassess situation, etc.
Queue Results
Potential result identified
Open Result(s)
Result potentially matches category
Data
Experience the situation: View Web Page
Review results
Activation of information from memory
• Plausible Goals
• Expectancies
• Relevant Cues
• Actions 1…n: extract results, defer judgment, reassess situation, etc.
Reassess situation or defer judgment on results
Result does not match task requirements
Result impacts existing report
Report components found
Extract report components
Information
Match to report requirements
Create or update report
Understanding
Update complete
Experience the situation: Check Task Status
Review task
Activation of information from memory
• Plausible Goals
• Expectancies
• Relevant Cues
• Actions 1…n: continue with task, continue reviewing results, review or reassess task, submit report, etc.
Report incomplete, continue extracting results
Report incomplete, continue reviewing search results
Report incomplete, review task requirements
Knowledge
Submit report
Create Report
Categorized Results – Create Report (RPD "A")
Information Processing
Task Status – View Web Page (RPD "A")
Knowledge & Understanding Transfer
Task Status – Submit Report (RPD "A")
VE
QR
OR
VP
Error in mental model
Error in mental model
Time on Task: View Web Page (Select Data)
SD
EX
Time on Task: Extract report components
EU
UD
Time on Task: Assess whether task is complete (Next Task)
NT
Error in mental model
SR

XIV. APPENDIX H: Model Markers

This section is used to help read and understand the analyst process model. Tasks are green brackets and markers are yellow labels on right side of the model. During data analysis, Markers are applied to Morae data, then exported and grouped by task for further analysis.

Marker	Description	Task
AC	Add [search result(s)] to Collection (GOST)	SR
BS	Begin new Session	MM
CA	Close Application	NC
CC	Create new Collection (GOST)	SR
CD	Create new Document [Word, PowerPoint, etc.]	NC
CE	critical error	NC
CR	Close search Result / close browser tab [or equivalent]	SD
CS	Change Settings -- applies to any setting not already covered	NC
CT	Close Tab	NC
DS	Delete [GOST] Search	NC
EC	End SmartEye Calibration	NC
ES	End Session / Submit Report	NT
EX	Extract [Copy] content from web page	EU
GE	GOST error	NC
GH	access GOST Help document	NC
GT	GOST training	NC
MA	Minimize Application	NC
MB	Modify Browser config, settings, add-ons, etc.	NC
MS	Modify Search	SR
NC	non-critical error	NC
NS	New Search	SR
NT	New [browser] tab	NC
OA	Open Application	NC
OB	Open Browser	NC
OE	Other Error (used for NOC system errors)	NC
OL	Open Link (from web page already open)	SD

OR	Open search Result	SR
QL	Queue Linked web page / open a web page link in a new queued tab	SD
QM	Query Map / Analysis (GOST)	SR
QR	Queue search Result / open web page in new tab, not visible	SR
RC	Remove Collection	NC
RO	Researcher Observation	NC
RS	Refine Search (search within results, GOST)	SR
SC	SmartEye Calibration	NC
SE	Search Error	SR
ST	Select browser Tab	NC
SV	SaVe document	NC
UD	Update Document / paste content extracted from web page	EU
UE	User Error	NC
VE	View named Entities / Analysis (GOST)	SR
VM	View Map / Analysis Map (GOST)	SR
VP	View web Page	SD
VQ	View Queued web page	SD
VS	View Search results	SR

BIBLIOGRAPHY

Hammond, K. R., & Stewart, T. R. (2001). *The essential Brunswik: Beginnings, explications, applications*. Oxford University Press.

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, 39(1), 101-117.

Kabacoff, R. (2011). *R in Action*. Manning Publications Co..

Lehman, A. (2005). *JMP for basic univariate and multivariate statistics: a step-by-step guide*. SAS Institute.

Montgomery, D. C., & Runger, G. C. (2007). *Applied Statistics and Probability for Engineers, (With CD)*. John Wiley & Sons.

Montgomery, D. C. (2013). *Design and analysis of experiments*. Wiley.

Salvendy, G. (2012). *Handbook of human factors and ergonomics*. Wiley.

REFERENCES

Adams, A., Lunt, P., & Cairns, P. (2008). A qualititative approach to HCI research.

Alsumait, A., & Al-Osaimi, A. (2010). Usability heuristics evaluation for child e-learning applications. Journal of Software, 5(6), 654-661.

Anselin, L. (2012). From SpaceStat to CyberGIS: Twenty years of spatial data analysis software. International Regional Science Review, 35(2), 131-157.

Beal, D. J., & Dawson, J. F. (2007). On the use of likert-type scales in multilevel data: Influence on aggregate variables. Organizational Research Methods, 10(4), 657-672.

Barnum, C. M. (2011). Usability testing essentials: Ready, set-- test!. Burlington, Mass: Morgan Kaufmann.

Best, R., & Cumming, A. Open Source Intelligence (OSINT): Issues for Congress. December 5, 2007.

Bim, S. A., Leitão, C. F., & de Souza, C. S. (2007). The Challenge of Teaching HCI Qualitative Evaluation Methods: a Case Study on the Communicability Evaluation Method.

Blaschke, T., Strobl, J., & Donert, K. (2011). Geographic information science: Building a doctoral programme integrating interdisciplinary concepts and methods. Paper presented at the Procedia - Social and Behavioral Sciences, 21 139-146.

Burke, M., Hornof, A., Nilsen, E., & Gorman, N. (2005). High-cost banner blindness: Ads increase perceived workload, hinder visual search, and are forgotten. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *12*(4), 423-445.

Cicchetti, D. V., Showalter, D., & Tyrer, P. J. (1985). The effect of number of rating scale categories on levels of interrater reliability: A Monte Carlo investigation. Journal of Applied Psychology, 9, 31-36.

Clemen, R. T., & Reilly, T. (2001). Making hard decisions with DecisionTools Suite. Duxbury.

Crandall, B., Klein, G., & Hoffman, R. R. (2006). Working minds: A practitioner's guide to cognitive task analysis. MIT Press.

Czaja, S., & Nair, S. (2012). Human Factors Engineering and Systems Design. In Salvendy, G. (Ed.). *Handbook of human factors and ergonomics*. Wiley.

Dainoff, M., Maynard, W., & Robertson, M. (2012). Office Ergonomics. In Salvendy, G. (Ed.). *Handbook of human factors and ergonomics*. Wiley.

Dalinger, E., & Ley, D. (2011). A reference model for designing decision support systems in novel work domains. Paper presented at the Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics, 1615-1620.

De Figueiredo, A.D., (2010). A Concise Introduction to Qualitative Research Methods in Information Sciences & Technologies. University of Coimbra.

De Kock, E., Van Biljon, J., & Pretorius, M. (2009). Usability evaluation methods: Mind the gaps. Paper presented at the ACM International Conference Proceeding Series, 122-131.

Ericsson, K. A. (Ed.). (2006). *The Cambridge handbook of expertise and expert performance*. Cambridge University Press.

Feltovich, P. J., Ford, K. M., & Hoffman, R. R. (1997). Expertise in context: Human and machine. MIT Press.

Fendley, M., & Narayanan S. (2012). Decision Aiding to Overcome Biases in Object Identification, Advances in Human-Computer Interaction, vol. 2012, Article ID 790304, 7 pages. doi:10.1155/2012/790304.

Fleiss, J. L. (1986). The design and analysis of clinical experiments. *New York, John Wiley& Sons*.

Gerhardt-Powals, J. (1996). Cognitive engineering principles for enhancing human-computer performance. Plastics, *Rubber and Composites Processing and Applications*, 8(2), 189-211.

Gladwell, M. (2009). *Outliers: The story of success*. Penguin UK.

Graettinger, C., Garcia-Miller, S., Siviy, J., Van Syckle, P., & Schenk, R. (2002). Using the Technology Readiness Levels Scale to Support Technology Management in the DoD's ATD/STO Environments (A Findings and Recommendations Report Conducted for Army CECOM) (CMU/SEI-2002-SR-027). Retrieved December 30, 2012, from the Software Engineering Institute, Carnegie Mellon University website: http://www.sei.cmu.edu/library/abstracts/reports/02sr027.cfm

Greitzer, F.L. (2005). Methodology, Metrics and Measures for Testing and Evaluation of Intelligence Analysis Tools. Report by Battelle Memorial Institute, Pacific Northwest Division.

Hammond, K. R., & Hammond, K. R. (Eds.). (1966). *The psychology of Egon Brunswik*. Oxford: Holt, Rinehart and Winston.

Hart, S. G. (2006, October). NASA-task load index (NASA-TLX); 20 years later. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 50, No. 9, pp. 904-908). Sage Publications.

Hollands, J. G., & Wickens, C. D. (1999). Engineering psychology and human performance. New Jersey: Prentice Hall.

Homeland Security Institute, "Department of Homeland Security Science and Technology Readiness Level Calculator (ver 1.1), Final Report and User's Manual", 30 September 2009.

Horn, J., & Masunaga, H. (2006). A Merging Theory of Expertise and Intelligence. In Ericsson, K.A. (Ed.). *The Cambridge handbook of expertise and expert performance*. Cambridge University Press.

Horsky, J., McColgan, K., Pang, J. E., Melnikas, A. J., Linder, J. A., Schnipper, J. L., & Middleton, B. (2010). Complementary methods of system usability evaluation: Surveys and observations during software design and development cycles. Journal of Biomedical Informatics, 43(5), 782-790.

Hvannberg, E. T., Law, E. L., & Lárusdóttir, M. K. (2006). Heuristic evaluation: Comparing ways of finding and reporting usability problems. Interacting with Computers, 19(2), 225-240.

Jacko, J., Yi, J., Sainfort, F., & McClellen, M. (2012). Human Factors and Ergonomic Methods. In Salvendy, G. (Ed.). *Handbook of human factors and ergonomics*. Wiley.

Ji, Y., Massanari, R. M., Ager, J., Yen, J., Miller, R. E., & Ying, H. (2007). A fuzzy logic-based computational recognition-primed decision model. Information Sciences, 177(20), 4338-4353.

Karahoca, A., Bayraktar, E., Tatoglu, E., & Karahoca, D. (2010). Information system design for a hospital emergency department: A usability analysis of software prototypes. Journal of Biomedical Informatics, 43(2), 224-232.

Klein, G. A., & Klinger, D. (1991). Naturalistic decision making. In Human Systems IAC Gateway XI(3), 16-19.

Klein, G., Orasanu, J., Calderwood, R., & Zsambok, C. (1993). Decision making in action: Models and methods.

Kuperman, G. (1997). Human system interface (HSI) issues in assisted target recognition (ASTR), Aerospace and Electronics Conference, 1997. NAECON 1997, Proceedings of the IEEE 1997 National, vol.1, pp. 37-48.

Kurland, L., Gertner, A., Bartee, T., Chisholm, M., & McQuade, S. (2006). Using cognitive task analysis and eye tracking to understand imagery analysis. MITRE CORP BEDFORD MA.

Lane, S. (2010). *Performance assessment: The state of the art*. (SCOPE Student Performance Assessment Series). Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education.

Leiden, K., Laughery, K. R., Keller, J., French, J., Warwick, W., & Wood, S. D. (2001). A review of human performance models for the prediction of human error. Ann Arbor, 1001, 48105.

Lieberthal, K. (2009). *The US intelligence community and foreign policy: Getting analysis right*. Brookings Institution.

Libicki, M, & Johnson, S. (1995). Dominant Battlespace Knowledge. NATIONAL DEFENSE UNIV WASHINGTON DC INST FOR NATIONAL STRATEGIC STUDIES.

Mankins, J.C., (April 1995). Technology Readiness Levels: A White Paper. http://www.hq.nasa.gov/office/codeq/trl/trl.pdf retrieved on 12/29/2012.

Marchionini, G. 2006. Exploratory search: from finding to understanding. Communications of ACM 49, 4 (April 2006), 41-46.

McDowell, D. (2009). *Strategic intelligence: a handbook for practitioners, managers, and users*. Scarecrow Press.

McDowell, D. (1997). *Strategic Intelligence & Analysis*. The Intelligence Study Centre.

McNeese, M. D., Bautsch, H. S., & Narayanan, S. (1999). A framework for cognitive field studies. International Journal of Cognitive Ergonomics, 3(4), 307-331.

Meister, D. (2000). Cognitive processes in system design. Theoretical Issues in Ergonomics Science, 1(2), 113-138.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational researcher*, 23(2), 13-23.

Molich, Rolf, Nielsen, Jakob. Improving a human-computer dialogue. (1990) Communications of the ACM, 33 (3), pp. 338-348.

Moorhouse, D.J., (2002). Detailed Definitions and Guidance for Application of Technology Readiness Levels. Journal of Aircraft Volume 39, Issue 1, January 2002, pp. 190-192.

National Geospatial-Intelligence Agency (2009). GEOINT Analyst (Geospatial Analysis). Retrieved from https://www1.nga.mil/Careers/Occupations/04%20Geospatial%20Analyst.pdf on 8/12/2012.

National Geospatial-Intelligence Agency (2009). GEOINT Analyst (Open Source Research). Retrieved from https://www1.nga.mil/Careers/Occupations/07%20Open%20Source.pdf on 8/12/2012.

NATO Open Source Intelligence Handbook, November 2001. Open Source Solutions Website. http://www.oss.net/dynamaster/file_archive/030201/ca5fb66734f540fbb4f8f6ef759b258c /NATO%20OSINT%20Handbook%20v1.2%20-%20Jan%202002.pdf. 2-3.

North, C. (2012). Information Visualization. In Salvendy, G. (Ed.). *Handbook of human factors and ergonomics*. Wiley.

Nicholls, M. R., Orr, C. A., Okubo, M., & Loftus, A. (2006). Satisfaction Guaranteed: The Effect of Spatial Biases on Responses to Likert Scales. Psychological Science (Wiley-Blackwell), 17(12), 1027-1028.

Norman, D. (2002). *The design of everyday things [electronic resource] / Donald A. Norman*. New York : Basic Books, 2002, c1988.

Oxford American Dictionary and Thesaurus. Oxford University Press, 2009.

Payne, S. J. (2009). Mental models in human computer interaction. *Human-Computer interaction: Fundamentals,* 39-52.

Pirolli, P., & Card, S. (2005, May). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. InProceedings of International Conference on Intelligence Analysis (Vol. 5).

Proctor, R., Vu, K-P., (2012). Human Information Processing: An Overview for Human-Computer Interaction in *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*. CRC Press.

Quesenbery, W. Using the 5Es to Understand Users. http://www.wqusability.com/articles/getting-started.html. Retrieved on 8/12/12.

Rajlich, V.T., Bennett, K.H., "A staged model for the software life cycle," Computer , vol.33, no.7, pp.66-71, Jul 2000.

Ravasio, P., Guttormsen-Schar, S., & Tscherte, V. (2004). The qualitative experiment in HCI: Definition, occurrences, value and use. Transactions on Computer-Human Interaction, 1-24.

Salas, E., Klein, G. A., (2001). Linking expertise and naturalistic decision making. Naturalistic Decision Making Conference. Mahwah, N.J: Lawrence Erlbaum Associates, Publishers.

Sauro, J. (2004). The risks of discounted qualitative studies.

Shneiderman, B. (2002). Inventing discovery tools: combining information visualization with data mining, *Information Visualization*, *1*(1), 5-12.

Serfaty, D., MacMillan, J., Entin, E. E., & Entin, E. B. (1997). The decision-making expertise of battle commanders. *Naturalistic decision making*, 233-246.

Shanteau, J., Weiss, D. J., Thomas, R. P., Pounds, J., & Hall, B. (2003). How can you tell if someone is an expert? Empirical assessment of expertise. Emerging perspectives on judgement and decision research, 620-639.

Sivaji, A., Abdullah, A., & Downe, A. G. (2011). Usability testing methodology: Effectiveness of heuristic evaluation in E-government website development. Paper presented at the

Proceedings - AMS 2011: Asia Modelling Symposium 2011 - 5th Asia International Conference on Mathematical Modelling and Computer Simulation, 68-72.

Spath, D., Hermann, F., Peissner, M. & Sproll, S. (2012). User Requirements Collection and Analysis. In Salvendy, G. (Ed.). *Handbook of human factors and ergonomics*. Wiley. com.

Spence, R. (2000). Information visualization. ACM Press. Essex, England.

Steele, R. D. (2007). Open source intelligence. *Handbook of intelligence studies*, 129.

Vu, K-P., Proctor, R., & Garcia (2012). Website Design and Evaluation. In Salvendy, G. (Ed.). *Handbook of human factors and ergonomics*. Wiley.

Weiss, D. J., & Shanteau, J. (2005). CWS: A user's guide.

Wickens, C. D., Gordon, S. E., & Liu, Y. (2004). An introduction to human factors engineering.

Wickens, C., & Carswell, C. (2012). Information Processing. In Salvendy, G. (Ed.). *Handbook of human factors and ergonomics*. Wiley.

Woods, D. (1995). *Toward a theoretical base for representation design in the computer medium: Ecological perception and aiding human cognition* (pp. 157-188). Lawrence Erlbaum Associates.

Woods, D. D., & Hollnagel, E. (2006). *Joint cognitive systems: Patterns in cognitive systems engineering*. CRC Press.