SPECIAL OPERATIONS FORCES DATA QUALITY ISSUES

(Practice-Oriented)

Donna M. Wood
The MITRE Corporation
Dwood@mitre.org

Eric R. Hughes
The MITRE Corporation
Hughes@mitre.org

Abstract: This paper describes the work to date on a study initially designed to identify and remedy automated system deficiencies. Specifically, the task was to identify key issues related to Intelligence support to Special Operations Forces (SOF) Mission Planning automated systems interfaces. Once the foundational research was completed, the preliminary analysis revealed significant underlying data quality problems. This paper describes the process used to reach these conclusions, and the plans being formulated to address these issues. A brief background is presented describing the Intelligence and Operations relationships within and among the SOF community. The methodology employed in the conduct of this study is then presented. Following that, a discussion of the identified issues is provided. The paper concludes with a discussion of next steps.

Key Words: Data Quality, Information Quality

BACKGROUND

The United States Special Operations Command (USSOCOM) is conducting a study to determine whether systems and processes are providing the intelligence data necessary to perform SOF-specific Mission Planning tasks. In addition, if shortfalls are identified, recommendations for near- and longer-term solutions to resolve these deficiencies must be formulated. Specifically, the task is designed to determine the unique SOF intelligence support requirements that are not being met by current staffing, policy, source data, and production/visualization systems. The main objective of the task is to recommend remedies to system-specific deficiencies that can be implemented by USSOCOM.

Figure 1 graphically depicts the problem. Many different sources (using many different file and data formats) are used as input to the Special Operations Mission Planning automated systems. Once these Mission Planning systems process the inputs, varied outputs (again, using many different file and data formats) are created and forwarded to the receiving platforms. Many of the Mission Planners, as well as the receiving platforms, expressed significant concerns about the ability of the external source and Mission Planning systems to provide the right data at the right time to the right place. The study is making an attempt to investigate these concerns and identify and address the valid issues.

PROCESS AND METHODOLOGY

The following subsections describe the activities that comprise the study. The first four activities have been completed. The remaining activities will be performed during the summer of 2002. Samples of supporting documentation and results generated from the first four activities have been placed in Annexes

A, B, and C. The contents of these annexes are described in the remaining sections of this paper.

The study consists of nine activities:

- Identify deficiencies
- Identify echelon level where problem exists
- Generate matrix of deficiencies
- Analyze deficiencies
- Determine priority of deficiency based on impact
- Investigate and identify potential solutions
- Establish level of effort required to address deficiency
- Make recommendations
- Publish an analysis report and briefing

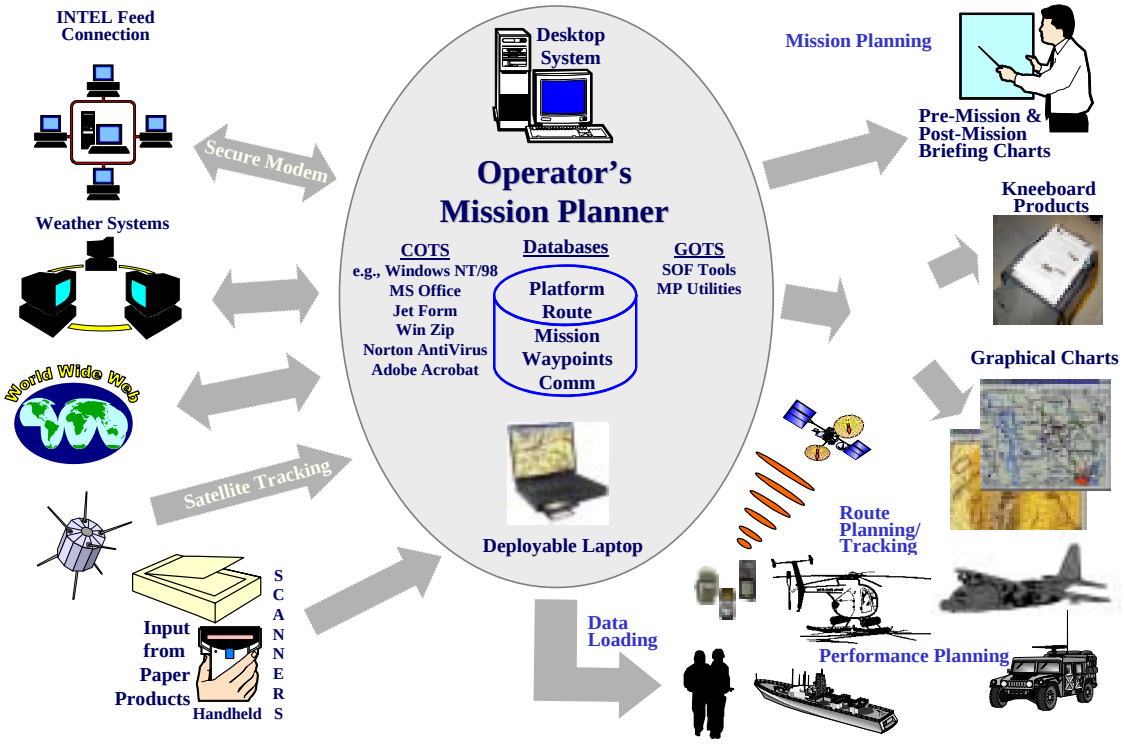

Figure 1. Task Overview – Intelligence Support to SOF Mission Planning

Identify Deficiencies

The methodology called for site visits to SOF components commands [3], [4], [5]. The component commands are Naval Special Warfare Command (NSWC), Air Force Special Operations Command (AFSOC), and United States Army Special Operations Command (USASOC). Each component command was asked to evaluate how well the validated, documented, Intelligence Support to Mission Planning requirements were being satisfied [1] [2]. Mission planners, intelligence analysts, and operations personnel were interviewed. The vehicle employed was a survey form, which was filled out

during the interviews. The requirements examined are listed in Annex A. The survey form is depicted in Annex B. The interviews emphasized SOF-specific deficiencies, but no deficiencies were precluded from the discussion. If the component command discussed an issue that was not SOF-specific, the issue was recorded. The focus of the interviews was also on the Intelligence Support to Mission Planning interface, at the automated system level. However, if the component command's issue was related to a different area (e.g., training), then that issue was also recorded. Special attention was paid to those areas that USSOCOM could address.

Each interview resulted in a component-specific matrix depicting how well that component felt that each validated requirement was being satisfied. There were three possible ratings for each requirement: Red, Yellow, or Green. Red is defined as the existence of a high or unacceptable risk in meeting the requirement. Yellow reflects a moderate or marginally acceptable risk in satisfying the requirement. Green depicts a low or acceptable risk. For each requirement that was evaluated as Red or Yellow, an explanation was recorded providing the rationale for that rating. A sample evaluation matrix with its respective explanatory notes is in Annex C.

Identify Echelon Where Problem Exists

All interviewees were asked to identify the echelon where the problem occurs. There were four echelons defined:

- Garrison – fixed location, non-mobile facilities
- En-route – during transport to the forward location (e.g., navy ship or aircraft platform)
- Forward-based (e.g., TSOC) – location near the target area that will be occupied temporarily; facilities and structures are mobile (e.g., tents), but do not move frequently
- Forward-deployed (e.g,. Forward Operating Base (FOB)/JSOTF) – at the target location

The results of this activity are reflected on the evaluation matrices as well (see Annex C for a sample).

Generate Matrix of Deficiencies

Once the interviews were completed and the collected data had been validated by the component commands, each of the "symptoms" was recorded in a matrix format. All concerns voiced by the component commands are referred to as a "symptoms" until analysis can take place and root causes can be identified. All entries in the evaluation matrices were depicted with the echelon and a Red, Green, or Yellow rating. Annex C contains an example evaluation matrix.

Analyze Deficiencies

As each deficiency/symptom was initially investigated, it was placed into one or more general categories. At the conclusion of this activity, the following general categories had been created:

1. Data content (e.g., missing attributes, complexity, resolution/fidelity)
2. Source data availability (e.g., existence of information, need to know)
3. Data access (e.g., permissions, connectivity)
4. Timeliness/reliability/consistency
5. Data parsing (e.g., ability to query and mine data)
6. Data visualization (e.g., ability to display thematic layers, flythrough in three dimensions)
7. Information overload (e.g., volume of information, data filtering)

8. Lack of authoritative data source (e.g., similar but conflicting information from multiple sources)
9. Hardware limitations
10. Software/Graphical User Interface (GUI) limitations
11. Manpower
12. Training
13. Data analysis/integration capability

All evaluation matrices were annotated with one or more of these categories for each non-Green entry. See Annex C for an example.

The results of all the individual evaluation matrices were then tabulated. A portion of the tabular results is included as Table 1. Note that of the 13 categories created from the analysis of the data, 9 are related to data quality – a significant proportion.

Conclusions and Next Steps

Since the initial analysis has revealed some unexpected results (i.e., data quality versus "automated systems" processing problems), the majority of the remaining study time will be spent addressing these symptoms. The component commands' concerns appear to stem from multiple causes, and at the highest level must therefore be addressed with integrated, coordinated, multi-faceted solutions. Specific automated interface and software application symptoms can be addressed with the current processes. However, the processes themselves need to be evaluated for possible enhancement and incorporation of data quality considerations. It is the authors' intent to formulate specific recommendations during the remaining portion of the study, and present these recommendations in a follow-on paper to be published in September 2002.

The results of this study can be considered in other domains. The four organizations surveyed can be considered "partners" of USSOCOM. By not constraining the breadth of issues raised, the study demonstrates the breadth of data quality problems encountered even when working with organizations of tens or hundreds of people. The results can be used as raw data to support enterprise-wide initiatives and investment prioritization, such as might be undertaken by a CIO.

REFERENCES

[1] United States Special Operations Command, Special Operations Mission Planning Environment (SOMPE) Operational and Tactical Levels Operational Requirements Document (ORD), 12 August 2001.

[2] United States Special Operations Command, Special Operations Mission Planning Environment (SOMPE) Program Master Plan, Draft Revision 6, undated.

[3] Wood, D. M., Trip Report – Intelligence Support to Mission Planning, Naval Special Warfare Command (NSWC), The MITRE Corporation, 14 February 2002.

[4] Wood, D. M., Trip Report – Intelligence Support to Mission Planning, Air Force Special Operations Command (AFSOC), The MITRE Corporation, 7 June 2002.

[5] Wood, D. M., Trip Report – Intelligence Support to Mission Planning, United States Army Special Operations Command (USASOC), The MITRE Corporation, 10 June 2002.

Category	NSWC	AFSOC	USASOC
Data Content (e.g., mission attributes, complexity, resolution/fidelity)	Information available for national feeds does not usually include complete pictures from USN SOF perspective	Resolution/fidelity insufficient for SOF-specific missions	USASOC requires better resolution imagery
Source data availability (e.g., existence of information)	Not available from traditional USN sources	No one-stop shopping for SOF-tailored products	Specific intelligence not available to CA/PSYOP
Data access (e.g., permissions, connectivity)	Information available on need-to-know basis only	File permissions; system hang-ups; LAN access; peripheral device access	- Required video data exists, but access to it is not available - Need automated interfaces between high and low sides (multi-level security)
Timeliness/reliability/ consistency		- Multiple versions of same report are similar but inconsistent - Must deconflict several sources - Manual entry also requires insertion of "dummy data" in some cases	- Timeliness, perishability/age of data, accuracy - Multiple copies of same data
Data visualization (e.g., ability to display thematic layers)	Files cannot be integrated into 3D virtual mission planning environment		Tool doesn't have right visual acuity for operations
Information overload (e.g., volume of information)	Considerable time spent culling pertinent information from message traffic		
Lack of authoritative data source (e.g., information from multiple sources)	Source data often not accompanied by confidence levels	Not able to derive confidence from the data	
Software/GUI interface		File formats not readable	Automated interface needed
Data analysis/integration capability	Personnel spend significant time converting data formats from/to external source systems, and within and between local systems	- Lat/lon coordinate conversions inconsistent - Some information virtually - impossible to fuse - Representations and semantics different between systems	- Lack of common mapping package across systems - Can process a given data type well, but cannot integrate with other data types

Table 1. Sample Consolidated Intelligence Support to Mission Planning SOF Symptoms

ANNEX A – SOME ORD REQUIREMENTS

Imagery Products
Interactive, 3D Scene Visualization Products
Target Intelligence Packages
IPB Graphic Overlays
Elevation (Topography) Digital Terrain Elevation Data (DTED)
Slope Polygons
Land Use/Vegetation
Soil Type
Soil Moisture Content
Terrain Feature Data
Navigation Aids/Hazards
Bathymetry (Water Depth)
Hydrography Feature Data
Meteorology
Oceanography
Force Locations/Activities
Location of Threat Systems/Entities
Detect and Engage (Threat Parametrics)
Impact of Environment on Threat
Systems/Entity Performance
Restricted Fire Areas
Fire Support Coordination Lines
Contamination Areas

ANNEX B – SURVEY TEMPLATE

Point of Contact (name, address, e-mail, phone, fax):

Requirement (e.g., Imagery Products):_____

Choose one of the following: Air, Maritime, Ground, Other. Explain as necessary.

Rating on Mission Planning Capabilities matrix and reason for rating:

What is the nature of the deficiency? Explain.

- Data content (attributes missing, complexity, poor resolution/fidelity, etc.)
- Data availability
- Data access (connectivity issues)
- Timeliness/reliability/consistency
- Data parsing/visualization capability (ability to display/mine source data)
- Information overload/lack of authoritative data source (conflicting input, information volume/ complexity)
- Hardware
- Communications
- Software
- Display/Graphical User Interface (GUI)
- Other

Priority/impact of deficiency (e.g., critical, mission essential, nice-to-have, etc.).

Echelon level where problem exists. Explain.

- Garrison
- En-route
- Forward-Based
- Forward-Deployed
- Other

Description of what is actually needed. Explain.

Documentation/references that we should look at:

Exercises that we can observe:

ANNEX C – SAMPLE INTELLIGENCE SUPPORT TO MISSION PLANNING EVALUATION MATRIX

	Source		Garrison		Forward-Based		Forward-Deployed	
	Rating	**Reason**	**Rating**	**Reason**	**Rating**	**Reason**	**Rating**	**Reason**
Imagery Products	Green		Green		Yellow	1, 6, 10		
Interactive, 3D Scene Vis. Products	Green		Yellow	11, 12, 13	Yellow	2		
Target Intelligence Packages	Yellow	1	Green		Yellow	6, 10		
IPB Graphic Overlays	Yellow	1	Green		Yellow	6, 10		
Elevation (Topography) DTED	Yellow	1	Yellow	4, 13	Yellow	1, 4		
Slope Polygons	Red	1, 2, 4	Yellow	1, 2, 4, 13	Yellow	1, 4, 6		
Land Use/Vegetation	Yellow	1, 2, 13	Green		Yellow	4, 6, 10		
Soil Type	Yellow	2, 4, 13	Yellow	2, 13	Yellow	6, 10		
Soil Moisture Content	Red	2	Red	2, 13	Red	2	Red	2
Terrain Feature Data	Green		Green		Yellow	6, 10	Yellow	6, 10
Navigation Aids/Hazards	Green		Green		Yellow	5, 6, 10	Yellow	5, 6, 10
Bathymetry (Water Depth)	Yellow	1,2, 4	Yellow	2, 4, 12, 13	Yellow	2, 6, 10	Yellow	2, 6, 10
Hydrography Feature Data	Green		Green		Yellow	5, 6, 10	Yellow	5, 6, 10
Meteorology	Green		Green		Yellow	6, 10		
Oceanography	Green		Green		Yellow	6, 10		
Force Locations/Activities: Blue	Green		Green		Green		Green	
Force Locations/Activities: Red	Yellow	1, 2	Yellow	2	Yellow	1, 6, 10	Yellow	1, 6, 10
Force Locations/Activities: White	Yellow	1, 2	Yellow	2	Yellow	1, 6, 10	Yellow	1, 6, 10
Location of Threat Systems/Entities	Yellow	1, 2	Yellow	2	Yellow	1	Yellow	1
Detect/Engage (Threat Parametrics)	Yellow	1, 2	Yellow	2, 7	Yellow	1, 6, 10	Yellow	1, 6, 10
Impact of Environment on Threat Systems/Entity Performance	Yellow	1	Green		Green		Green	
Restricted Fire Areas	Green		Green		Yellow	6, 10		
Fire Support Coordination Lines	Green		Green		Yellow	6, 10		
Contamination Areas	Green	2	Green	1, 2	Yellow	6, 10	Yellow	6, 10

Sample Evaluation Matrix Notes (partial)

Key:

1 Data content (e.g., missing attributes, complexity, resolution/fidelity)
2 Source data availability (e.g., existence of information, need to know)
3 Data access (e.g., permissions, connectivity)
4 Timeliness/reliability/consistency
5 Data parsing (e.g., ability to query and mine data)
6 Data visualization (e.g., ability to display thematic layers)
7 Information overload (e.g., volume of information, data filtering)
8 Lack of authoritative data source (e.g., similar but conflicting information from multiple sources)
9 Hardware limitation
10 Software/GUI limitations
11 Manpower
12 Training
13 Data analysis/integration capability

1. Imagery Products G – G – Y

   (1) data content – can't view attributes at far end
   (6) (10) Tool can't process vector formats

2. Interactive 3D Scene Visualization G – Y – Y

   Y – No ability to build the databases to pass on to consumer
   (2) FU - Can't get data tailored for their machines (laptops)

3. TIPs Y – G – Y

   Target Intelligence Package (TIP) is Y because traditional TIP doesn't have SOF-required fidelity (10) Can't be viewed at FU software application

4. Demographic Data G – G – G

5. IPB Graphic Overlays Y – G – Y

   Y same as TIP (1)
   Y at FU for (6) and (10) – visualization reasons

6. DTED Y – Y – Y

   Y (1) lots of source data to build DTED, but DTED not a product – imagery is the product
   Y at garrison
      Timeliness (4)
      No data integration capability (13) a la 3D scene visualization