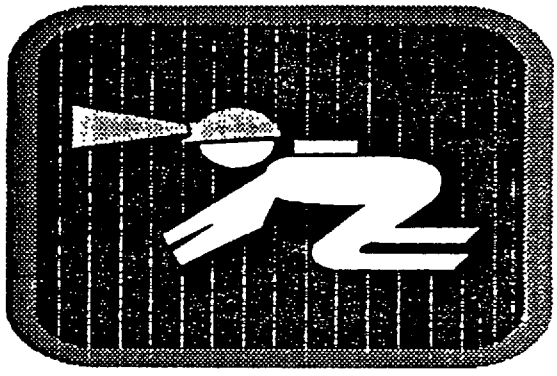


DATA MINING

Don't Buy or Build Your Shovel Until You Know What You're Digging Into

Mr. Robert D. Steele, President
OPEN SOURCE SOLUTIONS, Inc.
25 October 1994

Strategic Context of Data Mining

- War and peace have been transformed
- Operations other than war must be the foundation for planning
- Unconventional enemies and victims are a collection challenge
- Joint, coalition, and civil-military operations are a dissemination challenge

Transformation of War and Peace

- Information based warfare (and peacekeeping)
- Bulk of communications and computing now driven by intelligence
- Bulk of necessary information is from external sources
- For Third World, most of the needed information will come from open sources

Data Mining in Uncharted Terrain I

- Bad news: intelligence community is not trained, equipped, or organized to support the commander in operations other than war, and has not been collecting data on Third World operational areas

Data Mining in Uncharted Terrain II

- Good news: multi-media open sources, systems, and services in the private sector are robust, diverse, responsive, and relatively inexpensive.

Data Mining in Uncharted Terrain III

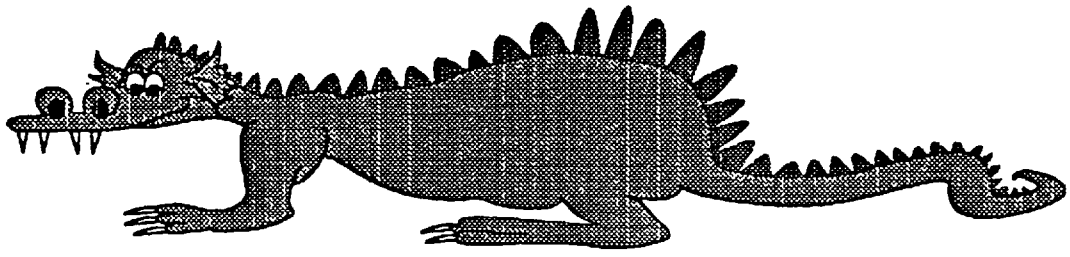
- Bad news: most of the data the commander really needs on Third World operational areas is not available in digital form, much less in packaged multi-media, and must be shared with coalition and civilian partners who are still using Royal typewriters and unable to receive, process, or exploit digital data, much less multi-media data.

Data Mining in Uncharted Terrain IV

- Good news: private sector is going to provide the tools for exploiting content; where investment is needed is in establishing a global capability to rapidly collect and digitize multi-media information essential to the commander, and in protecting multi-media data from destruction, distortion, or theft once it is acquired.

Information technology only makes bad management worse!

Mr. Paul Strassmann, (then) Director of Defense Information, speaking to OSS '92, 1 December 1992



Global Information Infrastructure

- The Global Information Infrastructure (GII) is a critical foundation for Army multi-media communications program
- Planning must focus on the "virtual intelligence community" and the varied open sources of information which the GII will bring online
- GII is a potential collection system

The International Information Commons

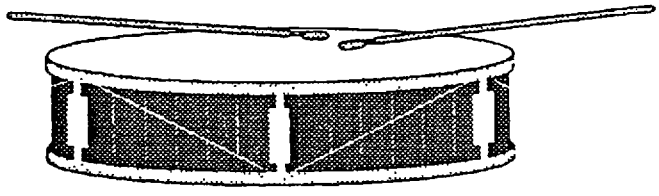
- The Internet is 10% of cyberspace, and cyberspace is 10% of knowledge
- Any global strategy for information-based operations *must* recognize that 90% of the needed information is not going to be digital!

Information Strategy

- Connectivity should not be the foundation for strategy or funding
- CONTENT should be the foundation
- Coordination of research & development can save dollars
- Communications & computing security seriously underfunded

Intelligence without communications is irrelevant; communications without intelligence is noise!

General Alfred M. Gray, (then) Commandant of the Marine Corps, Congressional Testimony 1991



Content and the Commander

- Roughly 80% of what the commander needs to know is not available within U.S. government (including intelligence) databases.
- Most of what the commander needs to know *is* available from the private sector.

USMC Intelligence Center

A Case Study

- Mapping, Charting, & Geodesy
  - ZERO digital data for 22 of 69 countries
  - OLD digital data for ports and capitals ONLY for 37 of 69 countries
  - VERY OLD digital data for remaining 10 of 69 countries

- Integrated Analysis Model
  - Civil Factors
  - Geographic Factors
  - Military Factors
  - Three Levels of Difficulty
  - Four Levels of Analysis

- Overview of Planning and Programming Factors for Expeditionary Operations in the Third World
  - Original study done by BDM for around $100K
  - Being updated today by PRC for around $250K
  - Used only open sources

Data Mining I

The Information Continuum

K-12 LIB PI/IB GOVT INTEL

UNIV BUS MEDIA DEFENSE

Iron Curtains Between Sectors
Bamboo Curtains Between Institutions
Plastic Curtains Between Individuals

Data Mining II

Functional Areas

EMPLOYEE OBSERVATION	DOCUMENT ACQUISITION	TELEPHONE SURVEYS
COMMERCIAL ONLINE SEARCHING	DOCUMENT TRANSLATION	MARKET RESEARCH
CURRENT AWARENESS	REVERSE ENGINEERING	RECRUITED AGENTS
EXPERT ON DEMAND	MULTI-EXPERT RESEARCH	INDUSTRIAL ESPIONAGE

Data Mining III

Status Report

TODAY:

Marginal in the U.S.
Fair in Netherlands, Singapore
Good in Israel, France, Sweden, Taiwan
Very Good in Germany (electronic only)

TOMORROW:

Marginal in Third World
Fair in U.S.
Good in Europe
Very Good in Asia

Data Mining IV

Vision Statement

In the age of information, national security and national competiveness require the full integration of and harnessing of what the Vice President has called "the distributed intelligence of the Nation".

The battlefield is now global, every human brain is a sensor, and every C4I architecture must provide for real time exploitation of all human, electronic, and hard-copy sources pertinent to battle planning and battle execution--the vast majority of those sources are civilian and foreign.

All these onramps, and all these drivers, are putting the superhighway into grid-lock--we won't be able to do super-computing.

Paraphrased gist of comments by Dr. Robert Kahn, talking to House Science, Space, and Technology staffer at MITRE, 29 June 1994

Defense Dollars Can make a Difference

- InfoTech R&D is *out of control*
  - Intelligence community
  - National laboratories
  - All DoD elements including services
  - Other major government departments
- Need national information strategy within which investments can be coordinated

Data Mining R&D I

- Tactical Document Acquisition & Digitization
  - Rapid and rugged scanning of rough documents that are crumpled, wet, and hard to read
  - Take on the non-trivial pattern recognition problems

Data Mining R&D II

- Automated time and space tags on all multi-media information
  - Work with intelligence community as well as private sector
  - Establish standards and methods which lead to all multi-media information having specific time & space tags

Data Mining R&D III

- Commercial Remote Sensing
  - Build the communications & computing enablers which will allow one meter synoptic resolution imagery from the private sector to:
    - Create 1:50,000 combat charts with contour lines on the fly
    - Provide guidance to precision munitions in real time

Data Mining R&D IV

- Digitize interactive speech
  - Enable coalition command & control
  - Military police and public affairs requirements
  - Improve prisoner interrogation process
- Automated translation in real time

Data Mining R&D V

- Communications & computing security
  - Totally vulnerable
  - Too reliant on commercial gauze tents, no standards, testing, certification, acknowledgement
  - Need rear area electronic security program ($1 billion a year)
  - Need electronic counterintelligence program ($500 million a year)

Multi-Media is the communicator's equivalent of the ultimate weapons system, such as a limited loiter fast-moving jet with precision munitions, and it will be just as useless to the Army as those jets are today, because neither the jets nor the multi-media communicators have the collection capabilities in place to find, fix, and feed "low slow singleton" target data to the shooter in real time.

Mr. Robert D. Steele, to the National Research Council, 25 October 1994

The Negroponte Factor

- He likes to say that computing capabiities have developed so fast, that it is as if we had gone from the Wright brothers to the 747 in two weeks
- Private sector will provide tools
- Private sector can acquire data

Sizing the Multimedia Data Bundle

- PC Magazine 25 October 1995 focuses on off-the-shelf multi-media kits.
  - Private sector will do fine with tools and storage
  - Private sector will do fine with massive multi-media pattern recognition
- How big a multi-media bundle does the soldier need to see?

Army intelligence is doing a great job with multi-media all-source fusion, and they are starting to do a better job of understanding open sources of intelligence in the private sector. They need to be a full partner in Army planning for multi-media communications <u>global sourcing of data</u>.

Mr. Robert D. Steele, to the National Research Council, 25 October 1994

First Catch Your Rabbit

- Connectivity without content is noise
- 90% of the data is not digital
- 80% of the consumers can't handle digital--may have to plan for liaison brigades
- Critical to define the target sources before building the multi-media munitions

THIRD INTERNATIONAL SYMPOSIUM: NATIONAL SECURITY & NATIONAL COMPETITIVENESS: OPEN SOURCE SOLUTIONS Proceedings, 1994 Volume I - Link Page

Previous	National Security Act of 1994
Next	Letter for the Open Source Lunch Club

Return to Electronic Index Page