

---

## Syndicate 4: Information Visualisation *Counterterror Intelligence*

Workshop IST-036/RWS-005  
“Massive Military Data Fusion and Visualisation: Users Talk with Developers”  
Halden, Norway  
10-13, September 2002

---

### Introduction

The participants in Syndicate 4 were Denis Gouin, Zachary Jacobson, “Kesh” Kesavadas, Hans Joachim Kolb, Vincent Taylor, Johan Carsten Thiis, and David Zeltzer.

By consensus, the members of the Syndicate selected the broad area of *information visualisation* as the topic of interest. In general, it is agreed that information visualisation refers to the presentation of “non-physical” data with no obvious 3D referents, as typified, for example, by multidimensional sonar or financial data; concepts embodied in documents and the relationships among them, or the morale and readiness of military units [Card, 1999 #1288].

It was felt that the mission of the Syndicate was to

- identify information visualisation issues in application domains of importance to NATO,
- identify and characterize the required capabilities and available technologies that address those domains, and
- recommend research and development priorities with respect to the technologies involved.

A technology/application matrix would be prepared, in which each technology would be mapped to application areas and required functionalities, and the matrix entries would quantify the relative maturity of the technology. R&D recommendations would then be derived from this matrix.

A number of application domains were considered, but due to the short time available to the Syndicate, this list was reduced to four, and ultimately only one application domain — counterterror intelligence — was addressed. This area is clearly of high priority to NATO, and at the same time, it is largely characterized by non-physical types of data that are problematic to present, and so counterterror intelligence is well-suited to consideration by Syndicate 4.

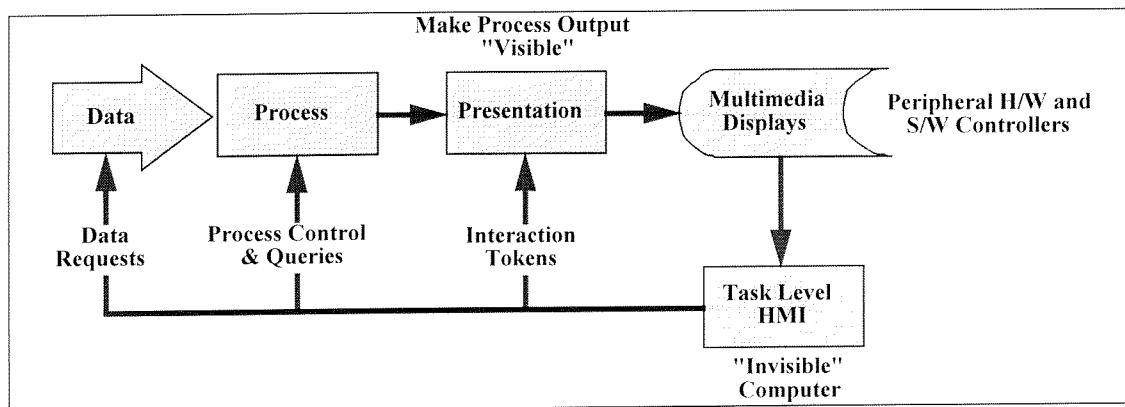
### Visualisation Reference Model

Once the topic of interest *information visualisation* and the application domain *counterterror intelligence* were selected, the next steps were to identify and characterize component technologies necessary for visualisation of counterterror intelligence data, and to estimate the level of technology maturity of these technologies.

It was thought that a *visualisation reference model* would be helpful in order to ensure that the Syndicate agree on the visualisation process under consideration, and to consider the technologies that comprise counterterror intelligence visualisation.

The model chosen is close to the VisTG model developed by Martin Taylor, but focuses primarily on the computational engines involved in data analysis and presentation (cf. *The VisTG Model for Visualisation*, these proceedings).

Figure 1 illustrates this visualisation model. The Syndicate specifically assumed that sensor technologies for gathering data was outside the scope of its considerations. It was also assumed that, in general, visualisation is a multimedia and multimodal activity. That is, data must be presented to analysts and decision makers visually, aurally and perhaps, haptically, as appropriate for the application in question. Likewise, they should be able to intuitively interact with presentations naturally with voice and gestures, again, according to the requirements of the application. Furthermore, a “task level” human-machine interface (HMI) enables decision makers to interact with a computer-mediated activity in terms of interest to the human, not the machine. Humans should not be burdened with extraneous cognitive tasks required to operate a computer system — a well-designed HMI should make the computer “invisible” to its users, in the words of Donald Norman [Norman, 1998 #1286].



**Figure 1.** A schematic view of the computational processes involved in the presentation of data to be visualised. The syndicate focused on characterizing the functionalities contained in the “Process” and “Presentation” modules, and identifying and rating the maturity of the technologies that address those functionalities.

## Counterterror Intel Requirements

The Syndicate considered three main requirements areas in this application domain. Data must be

- gathered from a variety of sources,
- analyzed with a range of tools, some automated and some human-in-the-loop, and the
- analyzed data would be presented to decision makers.

While this is not an exhaustive listing, the Syndicate identified four primary sources of data that would be of interest to the intelligence community:

- communications, such as
  - email, phone, FAX, radio, video, . . .;
- open sources, such as
  - newspapers, WWW, newsgroups, TV, . . .;
- commercial transactions; and
- behaviour of people and organizations.

For each of these data sources, functionalities and technologies required to analyze the data were characterized, and rated as to the respective technology maturity level —*high, medium, or low*. In addition, specific visualisation HMI issues were identified.

In addition, several main processing steps were identified. First, since it is practically impossible to attend to unfiltered streams of data of the magnitude represented by counterterror intelligence, the first step would be to rapidly analyze the incoming data streams for features of interest, which would be used to distinguish data to be analyzed further. Therefore, for each data source identified by the Syndicate, a first step would be to estimate the maturity of technologies available to recognize features in the various data streams. Once features can be identified, categorized and prioritized, the filtering of the data becomes straightforward.

Filtered data then becomes the input stream for further analysis. The Syndicate roughly characterized three further stages of data analysis:

- link analysis,
- data mining, and
- behaviour analysis.

In the view of the Syndicate, each of these processes may be implemented by arbitrarily complex algorithms and software systems, some of which might be completely automated, while others may be “human-in-the-loop”. Especially for human-in-the-loop processes, each of these analysis activities will require its own visualisation and HMI components.

In the final stages, the data output of the processing algorithms must be presented to decision makers for action. Visualisation and HMI issues were identified in each of the three analysis areas.

## **Feature Recognition and Communications**

Email, phone, FAX, radio, video

These represent essentially point-to-point communications in which content is arbitrary and unconstrained. This means that a robust natural language understanding (NL) capability is required to fully comprehend the content and intent of such messages, which is largely beyond current capabilities. Nonetheless, textual analysis technologies do exist to identify content features of such communication, so even though NL processing is not yet available, communications can still be categorized and related based on identified concepts contained therein.

In addition, many easily recognized parameters of communications can be derived, including, Source, destination(s), length, encrypted(?), language, subject field, attachments, routing, etc.

#### Content analysis

Textual concept recognition	
in some languages	<u>High</u>
for multilingual	<u>Low</u>
OCR	<u>High</u>
Speech recognition	<u>High</u>
Image and video feature recognition	<u>Low</u>
Intent recognition	<u>Low</u>

### Feature Recognition and Open Sources

Newspapers, WWW, newsgroups, TV, . . .

These are largely broadcast media, in which the domain of discourse is largely constrained by context. In such cases, for example, newspaper articles, NL technologies have been available for some time that can interpret such media and provide reliable paraphrased interpretations.

#### Content analysis

Textual concept recognition	
in some languages	<u>High</u>
for multilingual	<u>Low</u>
OCR	<u>High</u>
Speech recognition	<u>High</u>
Image and video feature recognition	<u>Low</u>

Intent recognition technologies                      Medium (NL paraphrasing technologies exist)

### Feature Recognition and Commercial Transactions

#### Transaction signatures

Customer ID  
 Credit card #  
 Product(s) purchased  
 Amount of product purchased  
 Purchasing frequency and history

. . .

All signature parameters maintained by merchants and subject to data mining

### Feature Recognition and Behaviours

#### Scope

Suspect entities

Behaviour signatures

- Phone calls
- Recipient and locations
- Travel
- Residence
- Biographical data
- Gait, action and mannerisms
- ...

Data sources

Current Law Enforcement Surveillance Methodologies

### Link Analysis

Link Analysis is a technique very useful to show relationships between people, organizations, events, incidents, behaviours and locations as shown on the left side of Figure 2 taken, from the U.S. company IntelCenter. Shown on the right side of Figure 2 is a subset of Mapping al-Qaeda v1.0, a product utilizing link analysis technology to provide visual maps of terrorist networks around the world and to help foster a better understanding of al-Qaeda's operational characteristics and organizational structure.

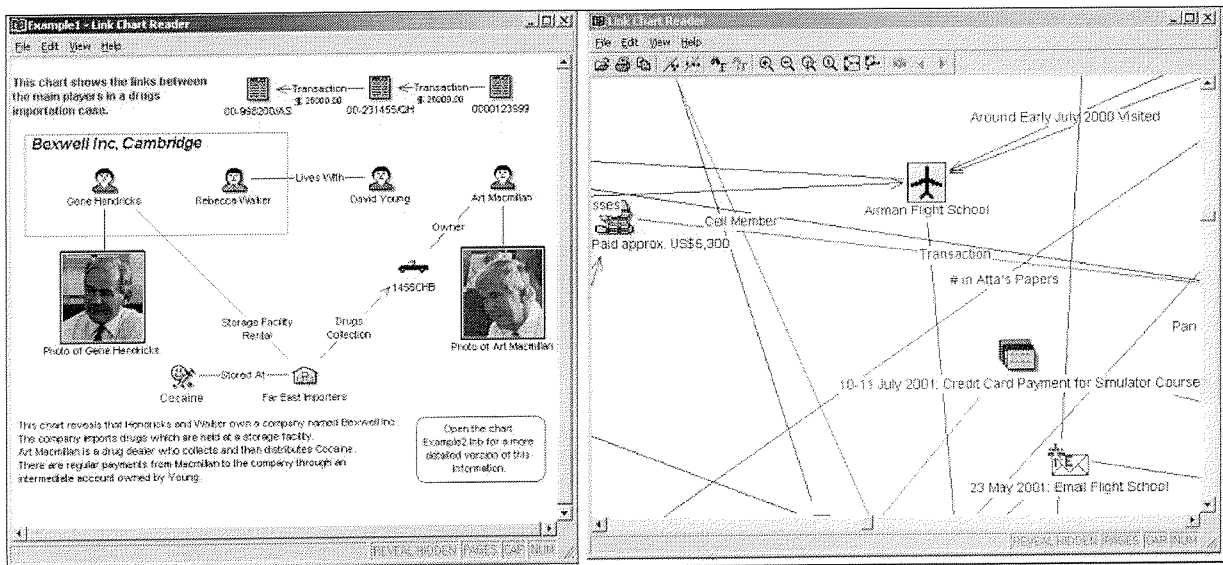


Figure 2. Examples of link analysis.

Mapping al-Qaeda v1.0 was produced by the private company, IntelCenter. The focus of IntelCenter is on studying terrorist groups and other threat actors and disseminating that information in a timely manner to those who can take action on it. Its primary client base is comprised of military, law enforcement and intelligence agencies in the US and other allied countries around the world (<http://www.intelcenter.com/linkanalysis.html>).

- Find patterns in recognized features
- Some tools available
  - Some automated

Some human-in-the-loop  $\Rightarrow$  visualisation

Medium technology maturity

Both Automated and Human-in-the-Loop Link Analysis Tools Require Further R&D Including Visualisation and HMI

## **Data Mining**

Search DBs

Recognized features

Others . .

Mining *structured* data

E.g., commercial transaction data

Off-the-shelf technologies available but difficult to use

High maturity but visualisation and HMI development required

Mining *unstructured* data

Low maturity

Data representation and association, automation tools, HMI and visualisation require major R&D

## **Behaviour Analysis**

Behaviour analysis is an emerging technique that allows investigators to identify suspect behaviours by comparing events with 'normal' information stored in a knowledge base. An example that could be drawn from a drug interdiction scenario is shown in Figure 3. In this case, the behaviour of a ship in terms of the itinerary, ports visited, time spent in each port is compared with information describing normal activities of the same category of vessel / ship [Kluchert, 1998 #1289].

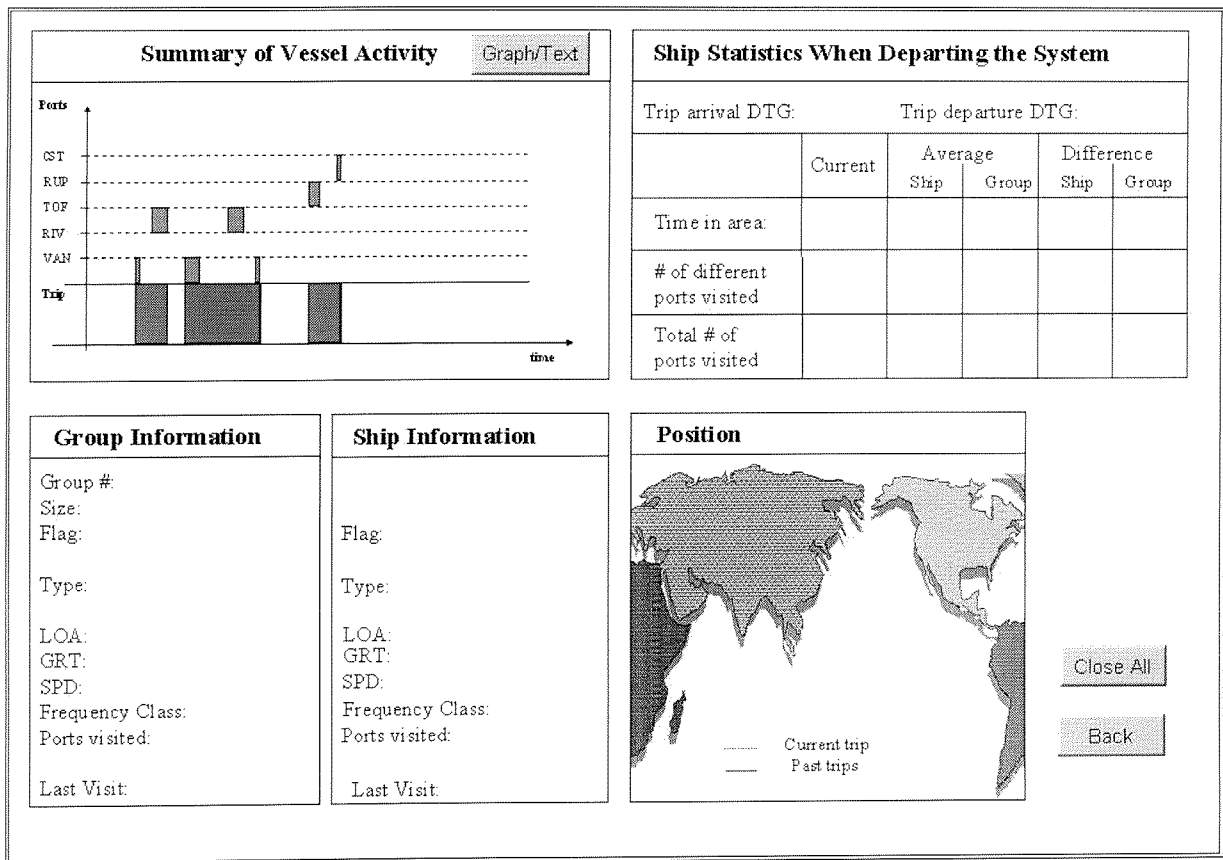


Figure 3. Examples of behavior analysis.

Scope

Suspect entities

Technology maturity Low

Many components available but major integration engineering required

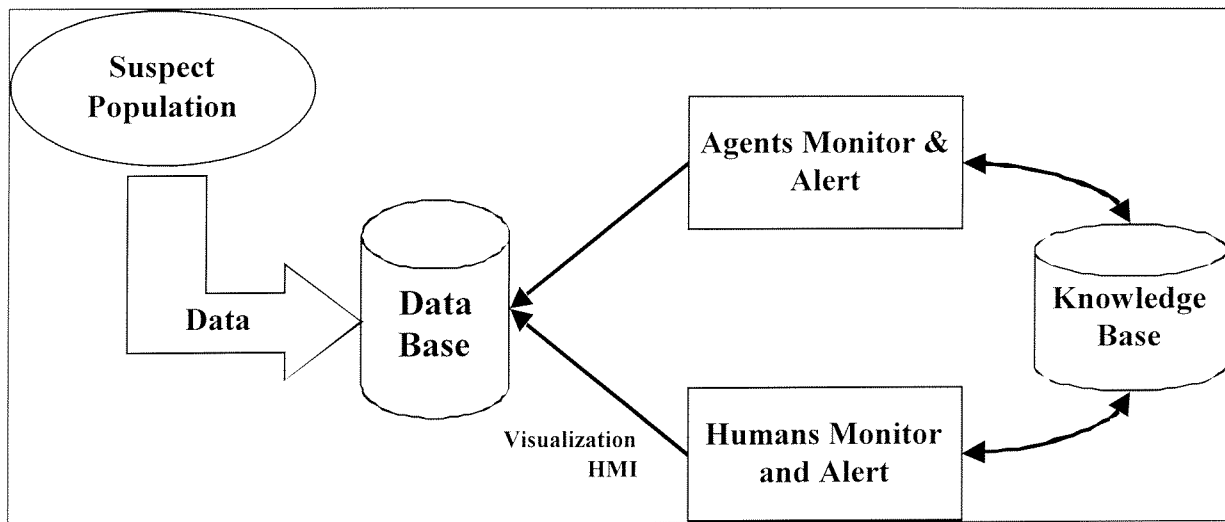
Robust and reliable monitoring technology not available

Prohibitively high false alarm rate

Human-in-the-loop signal detection requires visualisation and HMI R&D

Objective Distributed Technology

Regional, local, on-site, transportable. Figure 2.



**Figure 4.** A schematic view of a technology for monitoring a suspect population.

## Summary

Link analysis and data mining are “low hanging fruit”

Technologies “almost there” and potentially most productive in generating useful intelligence

Technology components exist but visualisation and HMI are poor

Most difficult challenge is algorithm “scaling”

Behaviour analysis is a promising application and is based on existing information technologies. Matching behavior to known parameters based on biometrics can also be a potential area of great promise.

Proof-of-concepts need to be developed.

Technologies are evolving and may be influenced by working groups

Matrix to be developed later