From Unstructured To Structured Information In Military Intelligence – Some Steps To Improve Information Fusion

Joachim Biermann, LtCol Louis de Chantal,

Reinert Korsnes, Jean Rohmer, Çagatay Ündeger

FGAN - FKIE
Neuenahrerstraße 20
53343 Wachtberg
Germany

EMAT/BCSF
14 rue St Dominique
00453 Armées – Paris,
France

FFI
PO Box 25
NO-2027 KJELLER
Norway

Thalès Communications /SAS/SEA
160 Bd de Valmy
92704 Colombes Cedex
France

Scientific Decision Support Centre
Turkish General Staff
Bakanliklar Ankara
Turkey

E-mail: j.biermann@fgan.de / louis.dechantal@fr.thalesgroup.com / Reinert.Korsnes@ffi.no / jean.rohmer@fr.thalesgroup.com / cundeger@ceng.metu.edu.tr

This paper is dedicated to the memory of our dear Danish friend and colleague
Dr. Per Husmann Rasmussen
He contributed so much to this project but was taken from his family and friends before he could see it brought to fruition. He will remain forever in our memories.

ABSTRACT

Military commanders and other decision takers require timely and accurate understanding of the situation in their respective area of responsibility as well as a prediction of the likely intentions and capabilities of supposed or potential adversaries. To achieve this intelligence cells have to process and evaluate information from all kinds of sources to deduce in a timely and most reliable manner an appropriate picture of the respective battlespace or evolving threat. Especially in the area of non-conventional conflicts, e.g. in the fight against terrorism, heterogeneous and complex non-military information factors are influencing the production of intelligence. Large volumes of information and data have to be processed but manual evaluation and the conventional presentation of results is much too time consuming. A first essential aspect in the automated support to the exploitation and fusion of information and data is to deal with very different natures of data: numerical data, usually in the form of simple tables, more complex structured data such as relational databases, semi structured messages, totally unstructured texts. Thus we have to deal with an impressive continuum of representations, from fully numeric and structured to totally textual and unstructured. A second challenging aspect in the automation of information fusion is provided by the heuristic nature of the real human processing of the imperfect available information. Fragments of information about the reality are compared with the current picture of the situation, related to the most likely aspects of estimated threats, aggregated to more complex and significant situation objects, and finally integrated into a new picture of the situation. Correlation and aggregation of information is typically driven by knowledge about the structure and behaviour of adversary factions. This paper will present some of the findings of the RTO Task Group on Information Fusion Demonstration (IST-038/RTG-016) concerning the before mentioned topical areas. Relevant open research aspects and an analysis and functional model approach of the human Intelligence processing is presented. All work was performed using a realistic operational scenario preparing a demonstration of the automated support of the Collation and Analysis steps in Intelligence processing.

1.0 INTRODUCTION

Within the Network Centric War and the global information environment of asymmetric threat, one challenge to military operations is to rapidly manage large volumes of data and information and to portray the results in a timely and adapted manner. In a Multinational Intelligence Cell (MIC) there is a danger that the processing capability will be overtaken by the volume of available information. In the current situational context, Operations Other Than War (OOTW) with Three Block War situations have become very important for NATO military organizations. This has brought an increased requirement for the processing of large volumes of HUMINT. A MIC needs the capability to collect, process, and disseminate a wide variety of data and information produced by the full spectrum of ELINT, IMINT, HUMINT, COMINT and socio-political sources.

This paper gives an overview on the work of the RTO Task Group on Information Fusion Demonstration (RTGonIFD IST-038/RTG-016) and the results achieved so far. The objective of the Task Group is to demonstrate that computer assisted information fusion is possible. The envisaged demonstration will be conducted using an operational scenario based on PSO and including references to counter terrorism.

Chapter 2 will presents a military view on the changed requirements in actual intelligence and argues for the necessity to have appropriate realistic scenarios for research and development in that area. The content and the character of the developed scenario will be described shortly.

In chapter 3 some of the relevant research aspects are presented which are inevitable to investigate when building an automated system for information fusion. A short introduction to the cognitive basis and framework of intelligence processing in reality illustrates some of the main questions to be studied to be able to automate this domain of human ability and professional skill. Some remarks are made on the aspects of “Knowledge Representation” and “Hypotheses Management” two topics which are correlated to the semantic nature and ambiguity of the problem space. Finally, the contribution about semantic networks presents a semantic tool approach to the question of information structuring .

Chapter 4 presents the analysis on the intelligence process and the results achieved so far concerning a functional model. Tools and software contributed by the participating nations will be used to proof the developed concept for automated support of information fusion. This will be done in the final demonstration mainly by supporting the Collation and the Analysis steps. The main objective of these two processing steps is to get an overview about the available intelligence information by putting every single piece of it into a predefined structure (set of classes or categories) and establishing correlations by categorizing, link analysis, and classification. The Conclusions in chapter 5 summarise the results .

2.0 MILITARY VIEW ON INFORMATION PROCESSING IN INTELLIGENCE

2.1 The operational structure of intelligence processing

Intelligence processing is an important part of Command and Control (C2) because a most accurate situational awareness of the battlespace is essential prior to all decisions and activities. It is a basic requirement which is independent from the ever changing and variety of potential conflicts. In order to fulfill the requirements from all various users and provide in the most timely and reliable fashion the appropriate pictures of the operations space, Intelligence cells have to process and evaluate incoming information. This task includes to determine the most actual location, strength, and activities of all engaged faction forces and to deduce their activities and likely intensions. A wide variety of information produced by the full spectrum of sensors and human sources has to be collected, filtered, processed and disseminated. This is done in a structured and systematic series of operations which is called the Intelligence Cycle. It includes four phases which are defined by the NATO Glossary of Terms and Definitions (AAP-6) [5]as follows:

i. *Direction:* “Determination of intelligence requirements, planning the collection effort, issuance of orders and requests to collection agencies and maintenance of a continuous check on the productivity of such agencies”

ii. *Collection:* “The exploitation of sources by collection agencies and the delivery of the information obtained to the appropriate processing unit for use in the production of intelligence”

iii. *Processing*: “The production of intelligence through collation, evaluation, analysis, integration and interpretation of information and/or other intelligence.”

iv. *Dissemination*: “The timely conveyance of intelligence, in an appropriate form and by any suitable means, to those who need it”.

These four discrete operations are conducted culminating in the distribution of the finished intelligence product. The sequence is cyclic in nature since intelligence requires constant reappraisal and updating if it is to remain current and relevant to users’ needs. The operations are discrete because, as information begins to flow, is processed and disseminated as intelligence, the operations will overlap and coincide so that they are being conducted concurrently and continuously rather than sequentially. The representation of the military intelligence function in Figure 1 shows the OODA Loop decision process interfacing with the direction phase of the Intelligence Cycle.

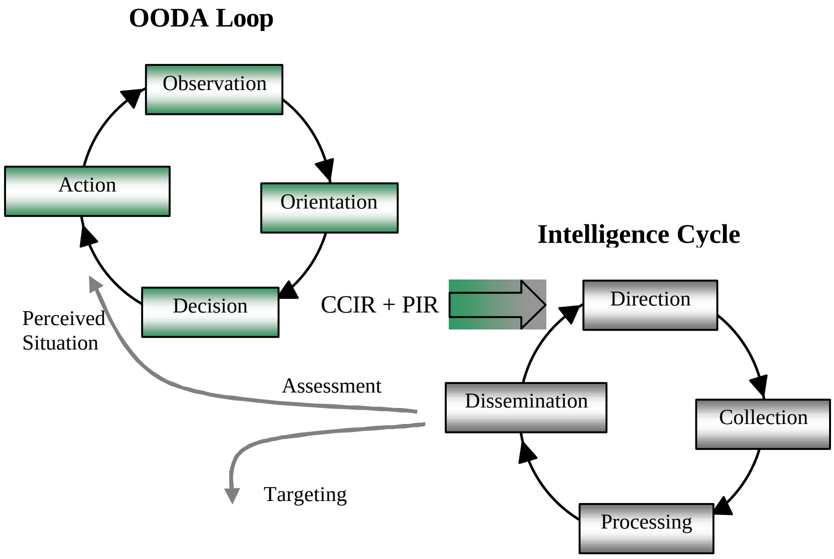

Figure 1 – Intelligence Cycle interfacing with OODA loop

2.2 The new essence of actual information processing

Intelligence doesn't work for its own purpose. It is no more than a service to all the force customers, ranks and file to commanders, who order products in the context of their own constraints. In that way, a single piece of data can participate in providing several different information each of which is strictly related to a different operational purpose. It is the nature of "data" to be part of all required common products (see Figure 2), usually reference images, heavily tinted with operational coloration. This aspect of a specific context highlights the essence of a delivered intelligence product: it is a set of raw data processed to information and subjectively engraved with the operational flavour of its addressee.

In that way, a huge difference shows up between the processing of data for high intensity conflict, symmetric operations, and crisis management or counter terrorism in asymmetric conflicts: on the one hand can we base our understanding and processing on stable, well known bases, while on the other hand we face unpredictable situations, with undisclosed threat and an enemy deeply integrated and sustained by local population and environment.

Due to modern operational doctrines, most of the future deployments will have to deal with urban areas and in total integration within the civilian population. Versatility of situation will be the standard and "Friendly-Foe-Neutral" discrimination in Three-Blocks-War is more than ever a reality, which implies to consider all human factors as major parameters. More than the knowledge of facts, their understanding and operational meaning is essential to adapt manoeuvre and reduce violence to its lowest possible level.

For this reason, the description and understanding of the requirements and the metrics of a system for the automated support of intelligence processing have to be based on a reliable scenario approaching very closely the environmental and informational context faced by soldiers on action. These aspects will become a major decisional precondition that cannot be avoided. It will be impossible to develop intelligence systems and getting high accuracy performances when using old Cold War considerations not taking into account the reality of urban operations, ranging from quiet discussion to armoured fight.

Thorough functional analysis of both operational and processing aspects are a necessary pre-requisite for the understanding of the deep and sophisticated nature of information and the different processing levels (see Figure 2) required for producing intelligence from data. The grown complexity is induced by the operational context with an ever increasing importance of not explicitly given contextual information.

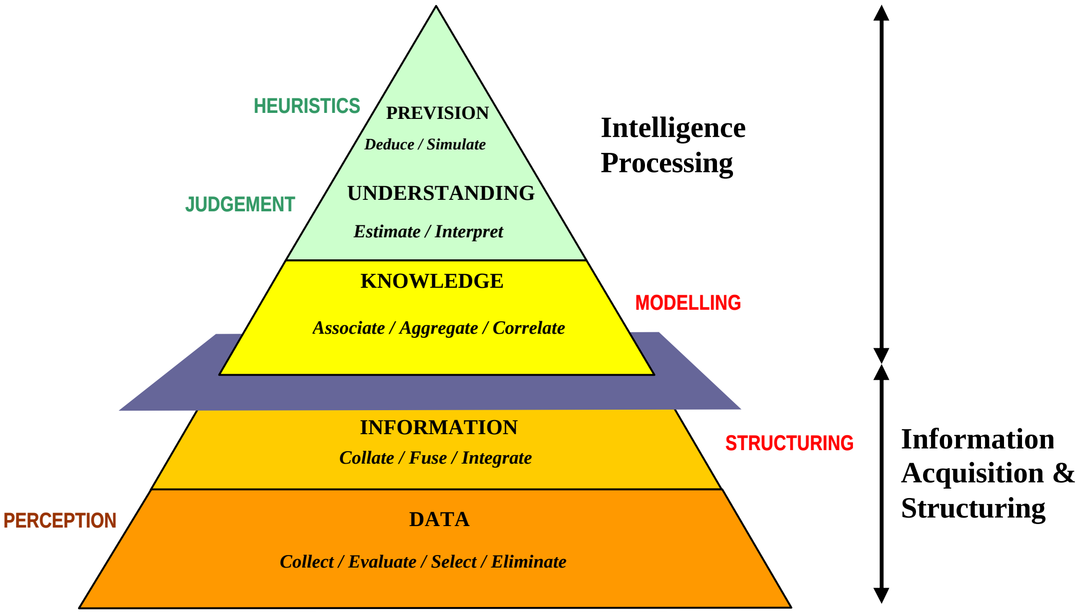

Figure 2 : The "Knowledge" Pyramid

Facts without problem context and without guiding question are without "meaning". An intelligence officer will never get all the relevant information, but will be impacted by partial, false, unreliable, irrelevant, redundant pieces of information which he will have to filter. The human brain permanently selects and inserts relevant information from its internal, mental understanding and model of the situation as it is perceived so far and the experience gained from "similar" problems in order to correlate and aggregate it all to a reasonable picture of the situation deciphering the meaning of all pieces of input data. To automate this intellectual process it has to be broken down into a series of functional steps as shown in Figure 3.

Collation

? System preparation	*Load background information, check ICP & DB*
? Get text file	*Get file in electronic format*
? Doc registration	*Check duplication of message; log document*
? Priority evaluation	*Filter urgency*
? Relation with ICP task number	*Check reference to collection plan*
?	Storage *Store, save doc in DB*
? Extraction of information	*Break into knowledge objects*
? Indexing	*Index full text word*
? Object identification	*Assign a category*
? Object creation	*Create a new object (name, log,…)*
?	Link building *Set relationships*
? Object merging	*Merge known concepts or objects*
? Link building	Discover new links from further inference

Analysis

? Knowledge extraction	*Query Knowledge DB for knowledge*
? Correlation	*Create links and objects of higher value*
? Integration *(aggregation)*	*Fuse knowledge for situation awareness*
? Interpretation	*Deductions, assessments, produce Intelligence*
? Dissemination	*Produce, disseminate, update Collection*

Figure 3 : Overview on Collation and Analysis steps

2.3 Necessity of a realistic research scenario

To be able to analyse the methods of intelligence processing it is important to have a scenario with all relevant aspects of real life processing. This means that information concerning the operational context as well as the expert knowledge about military behaviour and activities, which gives an understanding of the operational dynamics, has to be included.

A scenario required to run a manoeuvre is different from a scenario for elaborating a process. In the first case, the deliverable is a set of operational facts. In the second case, it is a set of information aiming at setting up an understandable background, in addition to the factual part of a pure manoeuvre scenario. Non factual data –immaterial field, environment, psychology and sociology,… – has to be provided.

To be able to build a module of such an automated system, it is necessary to carefully analyse and understand the human intelligence process. Knowing the required skills and having an extended military culture is not enough. The relationships between all pieces of information, the said –the facts- but the unsaid, implicitly presumed as well, have to be recognised and their respective position in the giant puzzle of the operational situation. Emerging information, pushed through by messages, is the source of this process analysis and knowledge construction.

This requires that, parallel to a central master event list, the scenario handles numerous elements of basic knowledge, that can be described in the context of the tasks of an Intelligence officer. The use of the "Pentagram" representation allows to deal with the 6 basic pillars of knowledge, the "said" world.

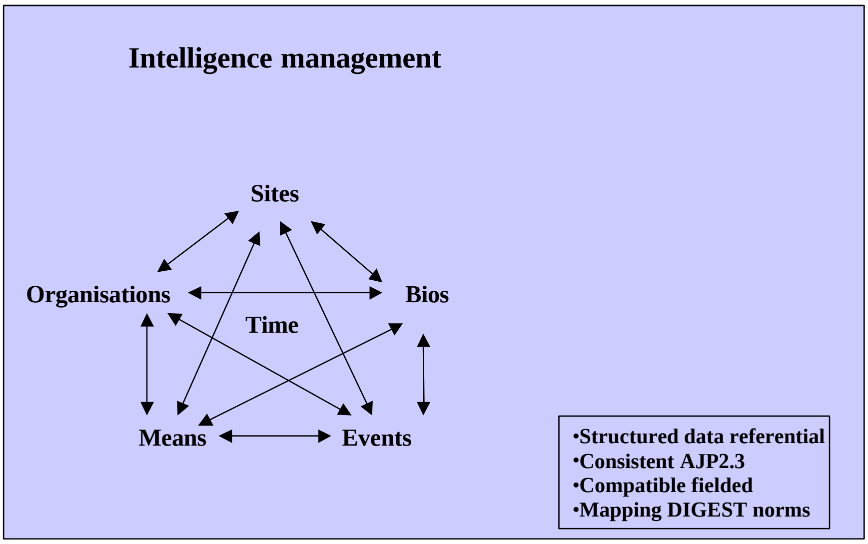

Figure 4: The "Pentagram" of knowledge categories

The "unsaid" world is extracted from a series of extracts of Real Ground Truth, declined into several granularity allowing to see an event:

- as a raw fact,
- as an interpreted fact by the Intelligence officer for a given environment
- within its own independent context (independent story) all along a given timeframe; this rationale can hugely differ from the operational explanation (field truth) of the Intelligence officer,
- within the global context, including the cross cueing of the merging stories and the real motivations of actors, independently of their action.

Such a scenario would not be performing if it did not include some aspects of real life, as irrelevant or confusing information messages, with the aim of voluntarily disturbing or bending the Intelligence deduction. It is obvious that not all parameters can be taken into a scenario. Nor may the scenario be too vast as the level of detail and coherence expends exponentially and would require a coherence management system. However, at R&D level, and in accordance with the terms of the Task Group TAP, "demonstrate through a successful proof of concept that computer assisted fusion is possible ", a limited scope is possible, but must encompass the most environmental details possible.

Description of the "RTGonIFD" scenario: Scenario micro - stories

- The scenario is composed of multiple stories related to 5 major themes:
  Politics, factions, Humanitarian, own forces, criminality.
- Within each theme, one story at least is complete and provides sufficient clues to make deductions and allow the G2 cell to embrace this micro situation, report to G3 and enable it to act if required.
- An excel spread sheet provides the necessary data to constitute an elementary background usable as a reference tool.

Example of theme story

"*Prijedor house blowing action*"

> "A group of 4 "Slav" activists in Prijedor harasses the Muslim population in order to make families move to another city. They organize night bombings of Muslim houses. The CZ detachment in Prijedor, helped by UK SAS, provide HUMINT, IMINT & SIGINT information on this action. It is firstly discovered that this group of 4 acts with a red Mercedes, then the name of 2 of the crew is advanced. Further on, the location of the driver' house is known, and more elements provide information about the relations between these men and their environment.
>
> After some additional gathering from diverse sources, the commando's course of action and its planned action can be understood, and its team members known as well. The analysis of these data can allow the G2 to trace this commando and the G3 to act."

Required and developed reference tables:

<u>mission dependant</u>

- City (gazetteer)
- geographic features
- organizations known in theatre (ORBAT) MO, members, location, related equipment
- important dates
- key infrastructures (incl. houses, street addresses, infrastructure description)
- known telephone number

<u>mission independent</u>

- organization leaders, personalities roles, equipment, holdings (quantity), weapons, locations
- persons key personalities (bio data), related organizations, weapons, vehicles, relationship, background info, equipment
- equipment
- locations places, building, bars/cafe, infrastructures, etc

- events critical event list
- blue location blue units / organization / locations
- red location red units / organization / locations
- white locations white units / organization / locations
- Services and Agencies (SandA) available sources & agencies

The final RTGonIFD scenario was elaborated from an initial theme selected out of real operational situations. This straw man was then transformed through personal experience and analysed to generate in a set of stories related under various forms, the official one being under the format of NATO messages including both structured and unstructured fields. Each message was linked to others, within or without its own story, and its contained pieces of information separately worked out under different forms as intelligence officer understanding, global or specific ground truth,…All pertaining information were merged into several folders according to the Pentagram.

As such, the scenario is divided in two portions: the factual one, including the messages, and the prospective one, for Scientists purpose, enabling to resituate and understand the mental process.

This cross referencing was based on real operational facts, extracted from the mastering of environment in the Bosnian conflict as part of the European Community Monitoring Mission (MCMM), a military diplomatic mission aiming at stopping local fights to reduce the level of violence and set the conditions for a return to normal life. As such, participants could enter all entities on both sides, consult archives and intelligence products, and analyse on the fields the various behaviours of any and each of the stakeholders. This knowledge was passed into this scenario to enhance its realism and credibility.

The change of operational focus impacts all related matters, including scenarios. Future operations will mostly deal with urban terrain and focus on reduction of level of violence. Thus, in the view of both a research scenario and an operational system, all environment components that interfere with military actions have to be taken into account to allow a deep situation analysis. The development of information fusion systems can no longer ignore non-military information and knowledge about the information context. It is obvious that ground truth restitution will highly facilitate the understanding of the process, but implies in turn to refer to real life. This will generate the need to implement most psychological and human parameters to enable comparison of human and automated process.

3.0 RESEARCH ASPECTS

3.1 Challenges for modelling the human intelligence process

In heuristic intelligence processing, rather than numerical data symbolic information, documents with text and pictures, has to be handled as input. The aim is not to discover and quantify single objects but to understand the behaviour and intentions of the adversary. A most reliable picture of the battlefield has to be integrated from imperfect information and the impact on the perceived development of the military situation has to be deduced. High level fusion answers the question "What does it mean?". In order to do this, intelligence officers practise by default a method of heuristic reasoning which relies on their knowledge about the standards and rules of the adversary and the assumption that the opposing forces will act according to these regulations and constraints for the benefit of their mission. It is common military experience and expert knowledge, that the production of intelligence can be done successfully by integrating current information based on the assumption of a default behaviour. This processing method is based on behaviour modelling, such as doctrinal templating or profiling. For this qualitative method a descriptive knowledge representation is used. Complex battlespace elements are described in a general manner using text and graphics. The descriptions include, if available, their relations and dynamic behaviour as well as their military intensions and concepts, ethnic and or religious position, their moral concepts and political involvements.

For high level fusion systems supporting intelligence two different kinds of models are needed[3]:

- behaviour models describing tactics of potential adversary factions and all necessary pre-conditions for their aggressive activities;
- models describing "normality", the common and unsuspicious behaviour of defined subgroups of the population.

In the first case high level fusion is the task of detecting indicators of a pattern of activity defined by the behaviour model. This approach is used e.g. in low and high intensity conflicts of military or paramilitary type. In the second case the task is to detect deviations from a pattern of "normal" behaviour. This method is used e.g. in combating terrorism, to be able to define indicators of suspicious activities.

From this the following requirements for the development of effective automated high level fusion systems can be deduced:

- to define the high level information fusion problem on a sufficient level of abstraction and to develop appropriate task models
- to develop according suitable fusion methods.

Besides that, for a fusion system it is necessary to automatically obtain access to the semantic content of the current input and the background information. As high level fusion is mainly based on document input, it is very important to have automated extraction of the content of the single input document. The semantic meaning of a document has to be accessible for a subsequent automated context based information analysis and integration. It is necessary to have methods and procedures for an automated text extraction and text analysis. Tools with efficient linguistic methods and adequate ontological models of each application domain are needed to understand the semantic meaning of the document input and to be able to cope with the information workload in network-centric intelligence and other C2 activities. Basically this is not the information fusion problem of intelligence processing in itself but as it is connected to the modelling problem, e.g. by its ontological approach to semantic text analysis, it should be regarded as part of the up to now unsolved problem of high level fusion systems.

Knowledge based information fusion is focussing on the heuristic human evaluation process[2]. Military information processing is modelled as a context dependent and template-based heuristic reasoning process and the real situation is modelled by a representation which is based on the above mentioned assumption that forces are organised in a structured manner and that they operate in a military reasonable and typical way. Respective rules, doctrines, or modus operandi can be used as a basis for matching templates for default reasoning. By using this approach, the analysis and fusion of incomplete and imperfect information of military reports and background knowledge can be supported substantially in an automated system. Such models can be developed by a thorough analysis of the real processes in intelligence and command and control. This has been shown to be true not only for classical war where doctrines are the basis for operations and training, but also for terrorist activities, like in Northern Ireland or in Spain. Even the preparing and planning activities of groups like al-Qa'ida exhibitsome systematic and structure which could have been recognised [9].

Describing the battlespace by default behaviour models defines a "Closed World" representation of the application area knowledge. But the information context which is relevant for a sufficient understanding of the situation is not fixed, it is open. Especially in new missions with low background intelligence and little knowledge about the behaviour of the adversary side the initially available knowledge will not be sufficient and the models will have to be modified or new ones will have to be created. Ontological models also define a "Closed World" representation of an application area[4]. Any concept which is not covered by an ontology cannot be taken into account during semantic analysis. As a consequence, a fusion system based on such a module will not be able to consider the missing information context, there is a blind spot for the fusion system. A solution to the obvious learning problem might be the concept of an interactive high level fusion system. It supports the human operator by providing formally correct fusion hypotheses using every information which is available to the system but leaves knowledge maintenance and final decisions to him. Such interactive systems will be adaptable to new and evolving situations.

According to the descriptive and qualitative nature of the behaviour models which are represented by non-numerical methods and the fact that high level fusion has to deduce from incomplete information, it is likely to use qualitative non-numeric methods for the matching of the imperfect information with the qualitative models. Among these are methods of non-monotonic logic like “Default Logic” and “Close World Assumption”, which both correspond very much with the human cognitive processing[2]. High level fusion will probably benefit from quantitative mathematical methods like fuzzy logic, evidential reasoning, and probability if a suitable abstraction of the problem is reached in order to be able to define detail problems which reasonably can be formulated as quantitative questions.

3.2 Knowledge representation

In the first phase of intelligence processing (collation), the intelligence system will get any type of structured or unstructured input documents such as texts, pictures and sounds, and store the raw data in the intelligence database. In order to understand the semantics of the stored data and continue further processing for classification, correlation, aggregation and interpretation, the data should be converted into a formal way, which is machine-understandable. Therefore any raw input data should be processed, transformed and represented as numerical and symbolic information. The critical point to be considered here is to have a formal representation, which would structurally represent any kind of real world information such as entities, concepts, relations, intentions, rules and doctrines, which may even have uncertainties and conflicts.

In addition to incoming intelligence information, the knowledge base should also be fed with background knowledge, which represents the conceptual model of the mission space. This ontological information will be the basis for the developed inference engines and fusion functions interpreting the available information to recognize untold significant facts, to produce hypothesis and to deduce possible future activities.

As clearly seen above, an intelligence knowledge base should deal with two kinds of knowledge: the conceptual model of the mission space, which mostly deals with abstract classes of the objects; and the collected intelligence information of the mission space, which mostly deals with real instances of the objects.

An intelligence knowledge base covering the pre-mentioned aspects could be handled using knowledge representation techniques deeply studied in artificial intelligence field. In general, knowledge representation can be described as an internal surrogate for external world [8]. Thus every intelligent entity doing reasoning needs to have a knowledge representation, because reasoning is a process that goes on internally, while most things it wishes to reason about exist only externally. An important question is “How close is the surrogate to the real thing?” What attributes of the original object does it capture and make explicit, and which does it omit? In fact one should be aware that perfect fidelity is in general impossible, both in practice and in principle. It is impossible in principle because the only completely accurate representation of an object is the object itself, and all other representations are inaccurate.

There are many ways of knowledge representation and reasoning methodologies such as First Order Logic, Fuzzy Logic, Production Systems, Neural Networks, Petri-Nets and Semantic Networks [7]. Some of these methodologies are highly structured and some are highly non-structured. Generally, the structured ones have immense significance over the non-structured models because the knowledge base can easily be represented by modular fashion, can easily be understood and can concurrently and efficiently be accessed.

The RTGonIFD mainly focused on Semantic Networks, a structured knowledge representation approach, which seems to be one of the best candidates to start with. A semantic net consists of two elementary tuples: nouns denoted by nodes and relationship between nouns denoted by links/arcs. For instance, the fact "cats hate dogs" can be defined in semantic networks as shown in Figure 5. Although the below figure illustrates a binary relation, it is also possible to define non-binary relations with semantic networks. However non-binary representation is not so common.

Figure 5: A semantic network saying that "cats hate dogs".

Another capability of semantic networks is to express absolute/monotonic and default inheritance and reasoning. For instance, the statement "birds can fly" means there should be no exception at all. Thus if a new statement is added to the knowledge base saying that a penguin is a bird and penguins cannot fly, then a contradiction will raise. Therefore the above statement should be relaxed as "birds can fly by default" to let the reasoner to find out that penguins cannot fly is more believable than penguins can fly if both are encountered. These absolute and default statements are illustrated in Figure 6. (dashed lines means default reasoning)

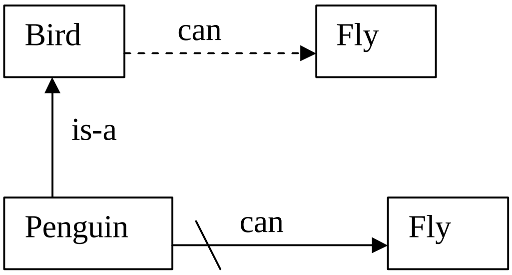

Figure 6: A semantic network representing a default assertion and an execption

Up to now, all the samples were covering classes of objects, which do not represent real objects but just the definitions. In semantic networks, the real instances can also be defined. In the following two sentences: "human is a living thing" and "John is a human", both classes and instances of objects can be seen. Nouns "human" and "living thing" are just classes, but "John" is an instance, which belongs to the human class. The above two sentences are illustrated in Figure 7.

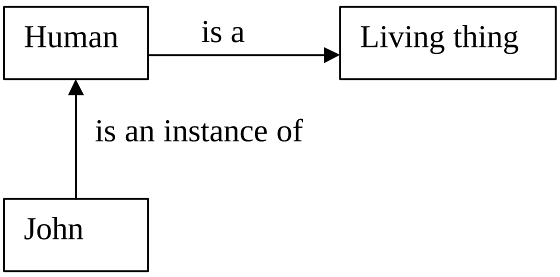

Figure 7: Classes and instances of objects

And finally, a good knowledge representation should also deal with uncertainty and time. For semantic networks, uncertainty and time information could be inserted into nouns and relations. For instance, existence of a noun can be limited with a specified time range (e.g. birth and death dates of a person), or start and end time of an event can be bounded (e.g. John was going home from 5to 6 o'clock), or strength of a relation could be defined with an attached probability value.

As mentioned above, there are several ways of representing knowledge. Although selecting one of these methods and applying on a domain is a difficult task to do, there is another important problem that remains unsolved, the difficulty of feeding the knowledge base with necessary background information. Without having the background information representing conceptual model of mission space, no fusion process could manage to extract significant facts and possible future activities within the incoming intelligence information. Therefore, the acquisition of expert knowledge to feed the knowledge bases is a major task in its own besides knowledge representation.

3.3 Hypotheses management

Uncertainty in observations and sparse data coverage can in principle give a large variety of possible estimates of a situation. An automated system supporting intelligence processing must have the capability to keep track of several hypotheses especially at an early stage of an operation. For this purpose a formalisation of hypothesis management can contribute to improve the overall intelligence processing and subsequently the management of an operation. Formal guidelines for activities can be considered as a restriction to flexibility. However, formalised guidelines are useful for a systematic evaluation of work and suggestions for improvements. Therefore in general formalisation has several advantages:

- It is the basis for automatisation and hence gives the opportunity to test a large number of possible hypotheses which cannot be processed manual by a human operator.
- Many hypotheses can in principle be treated fast, simultaneously and independently from human preference or black spots in the consciousness. There is no subjective assessment and prejudice which may affect the hypothesis treatment.
- Transfer and accumulation of experience will be more effective as compared to unsupported human knowledge management.
- Formal guidelines can in general support communication and to improve the transparency of processes and let others understand and monitor the activities.

The main task of an intelligence service should not be to only estimate the one state or situation considered to be the most likely explanation of the data. An effective situation awareness should identify all possible explanations of the available information which are of special interest within the actual operational context. Hence possible cost/risk estimates associated to decisions and associated actions should be central within situation awareness. This could lead to an automatic warning system to improve situation awareness.

Automated information processing can be used to enhance situation awareness for sets of hypotheses associated to threats. Assume, for example, a data treatment to search for indications of transports of explosives to a given location. There may be a large number of such hypotheses, and formalisation and systematic and automatic data treatment can therefore enhance situation awareness. Each actual observation, viewed isolated, may be uncertain and have little apparent meaning for a threat assessment. Observations can be car routes, correlations between movements of cars, people, transports, activities at possible places for storing, communications, observed suspected individuals etc. The overall correlation of the individual information, scattered in space and time and stored at very different places, may reveal a well known pattern, represented as a template for information matching, and thus indicate a threat (or possible cost) and initiate further investigations. Automatic formal procedures can help to discover such patterns, to let data look meaningful for a human operator.

3.4 Using semantic networks to represent and operate information fusion processes

3.4.1 How to represent textual Intelligence Information to facilitate fusion

Most Intelligence Information Systems manipulate documents written in natural language. These documents either come from open, public sources, or come from the inside of Intelligence Agencies. Texts are also associated with images. This association is performed by human operators, since no computer programs are able today to assign meaning to an image, even if they can help human operators by automatic detection of predefined patterns.

Technical data, -e;g; non textual data- are ultimately synthesized on maps, operations pictures, which, themselves, are ultimately described by words and sentences. This ubiquity of textual information finally turns into the ubiquity of document management systems, to assist Intelligence Agencies in information processing.

We advocate for a different approach: ***textual*** should **not** imply ***documents***, but ***textual*** should imply ***sentences.***

When reading a text, the brain does not register documents composed of hundreds or thousands of lines. Many neurobiological experiments indicate that associations are registered between concepts, which are nothing but sentences with words. Sentences share words, and they altogether build a network:

- *General Albert Bally commands the Third Army*
- *The Third Army fought in 2001 in South Mountains*
- *The Third Army was defeated by local troops in 2001*
- *South Mountains area suffered several hurricanes in 1999*

Our brain arranges information according to such a networked structure. We may have learnt each of these sentences by reading a document (a newspaper for instance), but our brain des not register the sequences of sentences found in each page of the newspaper. However, we in general keep the document – newspaper- form in our computers to store information. This discrepancy of information representation in brains and computers leads to a very poor optimisation of the man-machine cooperation in Intelligence. Choosing to represent information under the form of documents in computers originates in the choice to mimic the century-old books technology (using wood, paper, leather, ink, …). Books contain in general hundreds of pages because it happened to be a good economical compromise.

Electronic documents are a computerized simulation of antique production technologies and markets.

We can alternatively choose to use computers to electronically simulate the way brains represent information, i.e. as a network of sentences sharing words. These representation techniques have been thoroughly studied in the field of Artificial Intelligence since the early 50’s, under the name of “semantic networks”.

3.4.2 What to represent: beyond Pentagram

In everyday professional life, we are not used to carve textual information into small pieces (semantic networks). Our brain does it as a background, unconscious task, while we keep reading and writing pages and pages of text. When we want to turn to precise information, we consider *databases*, i.e. carefully structured pieces of information, in a freeze manner imposed by a computer application written by computer specialists using a programming language. This database technique implies that all the possible kinds of information, values and relationship are known by advance, and will seldom change, the etymology of the program is “written in advance”. This is clearly not acceptable for Intelligence missions where the nature of phenomenons to investigate is by definition unknown and unpredictable.

This database technique is well illustrated by the notion of "Pentagram" used in some Intelligence Systems (see Figure 4, page 6). The pentagram approach consists of modelling the universe through a fixed number of concepts. We consider "Events". An "Event" is characterised by five dimensions:

- People
- Organisations
- Place
- Date
- Means of Operation

This pentagram approach is a good departure from the pure document approach and a first step towards a network of sentences. However, it exhibits all the stiffness of the programmed database approach (besides its origin). We should better consider pentagrams as a small subset of semantic networks, and design systems where officers in charge can forge their own categories, dimensions and links to represent Intelligence Information the way they wish. For instance, operational experiments with the Thales Ideliance? Semantic Network Editor were conducted by French Army in 2003 and 2004, and led to semantic models with around 20 categories, and not 5.

Here are some examples of the variety of Intelligence information one should wish to represent with Semantic Networks:

- Opinions (who believes what)
- Documents (this is not a paradox: documents exist in the real life even if they are not the best way to represent Intelligence information !)
- Elements of situation (events are useful only as far they change part of the current situation)
- Attitudes
- Threats
- Intentions
- Surprises
- Forecasts
- Contradictions
- Explanations
- …..

Only open-ended methods like Semantic Networks can accommodate this variety of needs.

3.4.3 Some useful operations on Semantic Networks to assist Information Fusion Processes

In a first step, semantic representation offers a simple and natural way of accessing information by navigating from node to node (concept to concept, sentence to sentence) in the network. A more ambitious goal consists of performing automatic deductions and computations on the network, i.e. adding automatically inferred sentences to the network.

Stated with other words, this was the aim of Expert Systems and Artificial Intelligence applications in the 80's and early 90's. Most such attempts failed, for the reason that it was impossible to collect and formalize all human knowledge needed to write precise *inference rules* to automatically augment the network with conclusions. Most of the time, experts and expertise are not available. It would be the same for knowledge about information fusion.

An alternative is to infer automatically the rules themselves through automatic learning from past examples. This is a statistical approach, which implies the availability of large sets of high quality data. It needs not to be mentioned that such data will never be accessible in our case, due to the fact that each crisis is different, and that these data are not broadly published. Can we overcome this pessimistic conclusion: *no expert and no data* ? Can we anyway hope to use computers to assist analysts in inferring their fusion decisions?

We suggest that the answer may be yes, if we adopt the following principles:

- automatic operations should be only a subset of the fusion process
- what is actually done automatically should be “evident” to the analyst

We propose two steps in this direction:

- develop functions which compute similarities between objects
- develop functions which compute “subsets of interest” among the whole network.

The first point –*similarity*- means that we offer the analyst the possibility to automatically exhibit the objects which look like a given one, along with the points of commonality among them (a natural extension is to cluster a set of objects into homogeneous subsets). The essence of such an operation is just to tell the analyst: “hey, look, object A resembles objects B1,B2, … ,Bn, with respect to points P1,P2, …,Pk”. This process lets the analyst free to infer himself a potential conclusion. This functionality is simple, yet not trivial. For instance, objects Bi can be case studies, or recommendations, or mandatory procedures. Then “A looks like B” may mean: “in order to process object A, first have a look at what is said about object B !” We are not far from applying a rule, but all the “firing” initiative is left to the analyst.

The second point -*subsets of interest*-, is a generalisation of the first one. The idea is again to propose to the analyst a set of information which can be of interest for him, then let him infer a conclusion, if any. An example of “subset of interest” is what we call the “what between” operation in a semantic network. *“What between A and B”* computes all the paths linking A and B, then proposes this set of information to the analyst.

Many other definitions of “subsets of interest” may be defined, and we think they can be a valuable to fusion analysts. In the future, we wish to experiment with users which kinds of “subsets of interest” will prove to be the most useful.

4.0 SUPPORTING HEURISTIC INTELLIGENCE PROCESSING

4.1 Functional modelling

A very high level representation of the military intelligence global function appears in Figure 8 where the main actors are represented. The processing task is the main capability to answer commander's requests via the J2. Processing is done within the MIC, but shown outside to stress its importance.

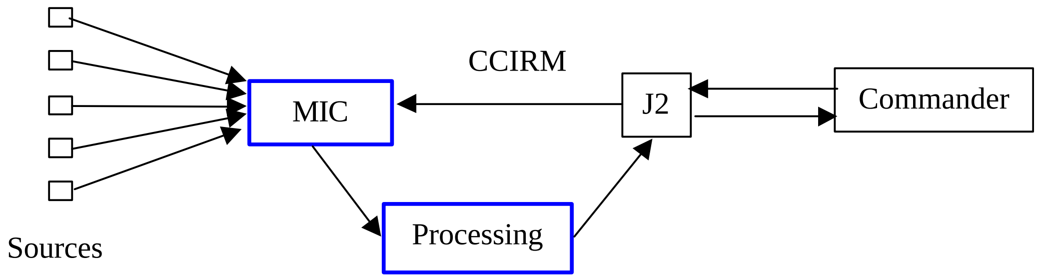

Figure 8 – Very high level representation of the military Intelligence function

The processing phase within the MIC is where the military intelligence branch needs automation to be more effective in their work when in OOTW. It was analysed and a conceptual and a functional model of the intelligence processes developed. The MIC internal functions appear in Figure 9. Those which are related to operational products are marked in blue. The complete breakdown of the UML representation of the conceptual model can be found in [11].

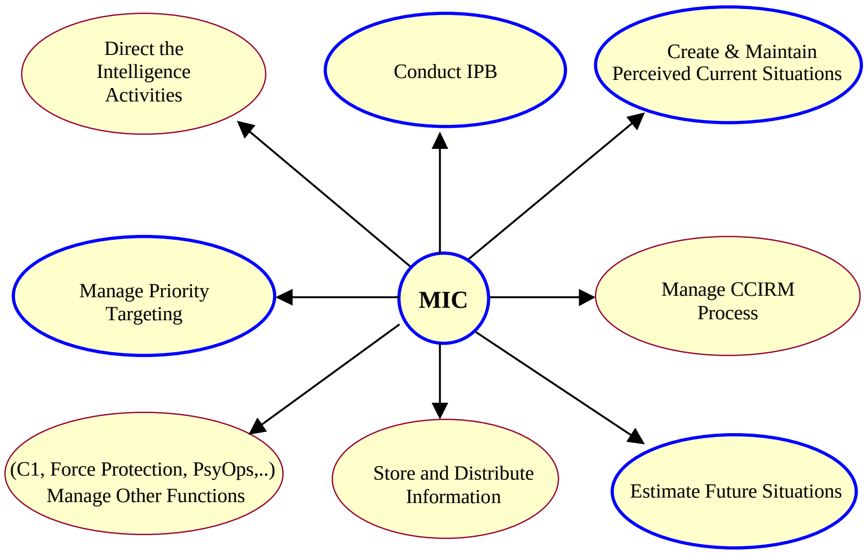

Figure 9 – Multinational Intelligence Cell internal functions

The processing phase is where the information that has been collected in response to the direction of the commander is converted into intelligence products. Processing is a structural series of actions, which although set out sequentially, may also take place concurrently. It is defined as the production of intelligence through collation, evaluation, analysis, integration and interpretation of information and/or other intelligence. A simplified way to represent the intelligence processing phase appears in Figure 10.

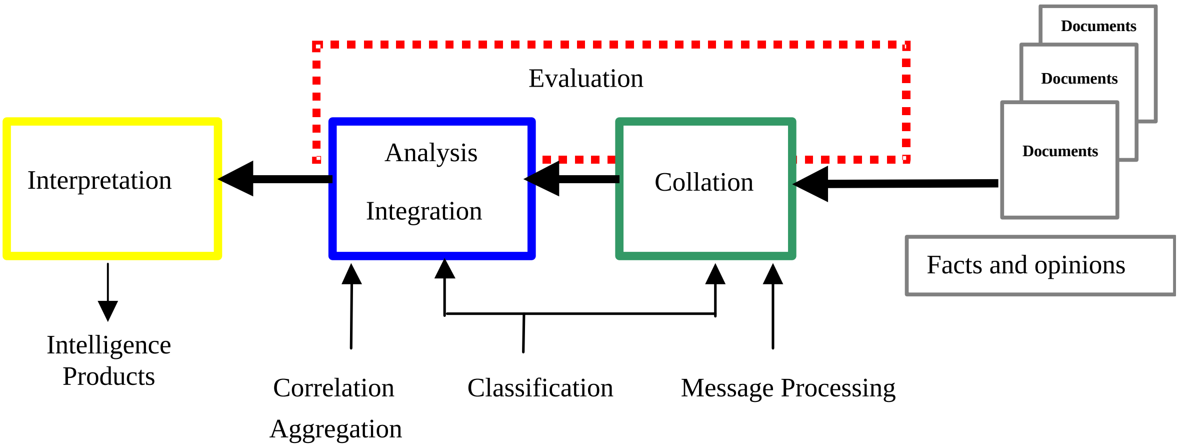

Figure 10 – Intelligence cycle processing phase – representation of processes

The military understanding of these processing steps is kept in Allied Joint Intelligence Counter Intelligence and Security Doctrine (AJP 2.0) [6] and in the NATO Glossary of Terms and Definitions (AAP6) [1]:

Collation: The registering and logging of the incoming information followed by its decomposition into individual information items. These individual information items are subject to categorizing according to either predefined categories or to new identified categories adapted to the mission. The categorized information items are finally cross-referenced each one with the others.

Evaluation: There are two types of evaluation. The evaluation of information derived from human sources that require subjective evaluation. The evaluation of information derived from technical sensors that, if accompanied by meta-data describing the circumstances of its collation, could be subjected to an automated evaluation process. One should also consider technical systems including humans, which means that incoming information does not have technical meta-data, but also subjective assessment.

Analysis: "A step in which information is subjected to review in order to identify significant facts for subsequent interpretation" It consists of a number of interacting sub-processes to bring the analyst answering questions like 'Who is it?, What is it?, What does it means? Why is it happening?, etc,', in order to recognize indicators and warnings.

Integration: "A step whereby analysed information or intelligence is selected and combined into a pattern in the course of the production of further intelligence". The process of building pictures of the current and of the predictive situations from all the gathered and analysed information.

Interpretation: "A step in the processing phase of the intelligence cycle in which the significance of the analysed and integrated information is judged in relation to the current body of knowledge."

4.2 Concept for automated support of Collation and Analysis

Out of the scope of these intelligence processes, three main interesting phases were extracted as candidates for an automated support to clearly demonstrate some new functionality.

Collation	:	*Text storage*	*Text extraction*	*Categorisation*
Analysis	:	*Correlation*	*Link analysis*	
Integration	:	*Pattern matching*	*Aggregation*	

These steps were analysed in more detail to develop a workflow of the human processing to serve as a bases for a system concept.
As shown in Figure 11, during the collation process, the incoming information is first logged and stored

into the appropriate repository. The received information is afterward decomposed into individual items, which are finally categorized and cross-referenced with previously received information items.

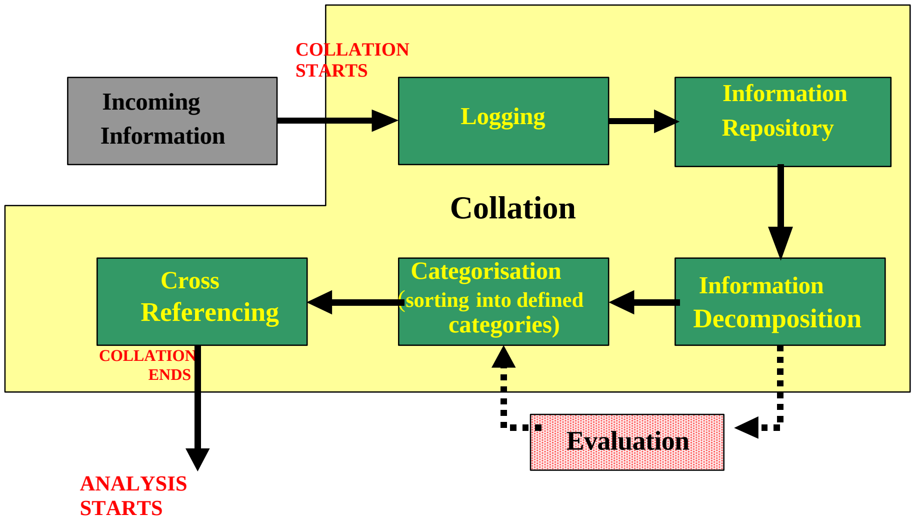

Figure 11 – Decomposition of the Collation process

During the collation step, a first semantic analysis is performed associating any new report to an appropriate category of intelligence. The number and definition of the categories depends on the type of operation which is conducted. These categories are deduced from the commanders intelligence requirements and define the main situation aspects and the information context which has to be covered to be able to give a description of the situation that sufficiently reflects the abilities and intentions of the opposite parts and to make an assessment. The association of a single report with a category of information is just an experience guided correlation. No further interpretation is done at this step. For free text reports there has to be a pre-processing to exploit their semantic content.

Figure 12 presents a further detailed view of the analysis process and its interactions with the evaluation, integration and interpretation processes. It is assumed that an initial domain representation model (a taxonomy or an ontology of a sufficient sub-set of all elements of the problem domain, including physical objects, activities, events, doctrines, known relations and dependencies) is given as part of the background knowledge.

The result of the collation process initiates the classification of the new received information items.

Classification is the function by which some level of identity of an item is established either as a member of a class or a type within a class. Classification basically is a matching procedure. It answers the question: "Who or what is it?" The appropriateness and quality of the domain representation model mainly influences the quality of the classification results. The richer the set of attributes of the (observed or deduced) object (entity, event, ..) the better a classification process can be. If the key classes which are used during the Collation Categorisation (e.g. people, places, events, organisation, and equipment) are high level classes of the domain representation model then we can interpret Categorisation as an initial step of Classification (raw classification) and include it into automation.

The classification process is followed by the correlation of the new information items. Correlation is the function by which data or information from similar sources about the same item, activity and occurrence are combined together. Different information items (classification hypotheses on objects of the problem domain) are associated / related to each other because they are sharing certain attribute values or are fulfilling specific criteria. This does not take into account that the different objects may match attributes of higher level classes of the domain representation model (a class of higher complexity different from the classes of all input objects). No "Part-of" or any alike relationship is tested because that is the functionality of the "Aggregation".

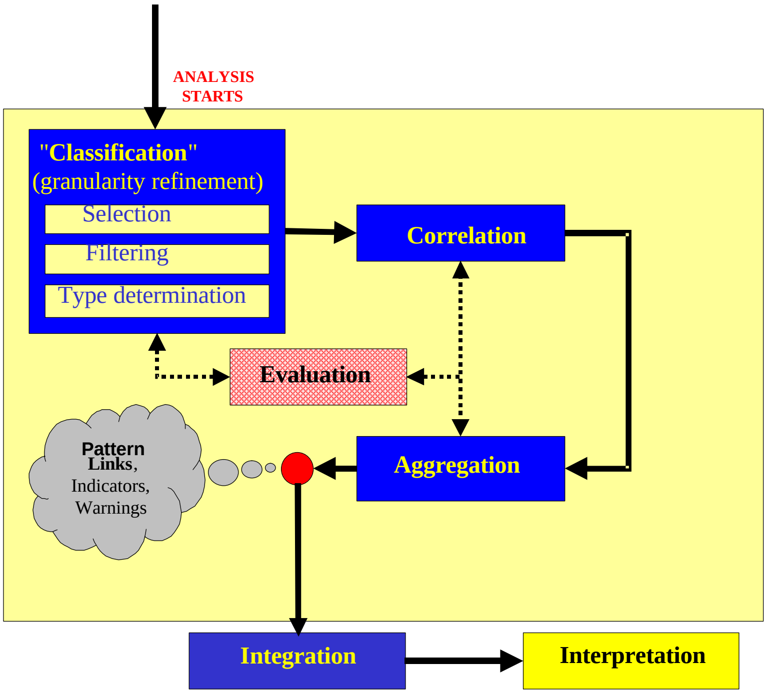

Figure 12 - Analysis, evaluation, integration and interpretation processes

The "Correlation" function is based on quantitative measures while a similar function "Link" is based on qualitative measures. Thus we can distinguish between these two functions by using the following discriminates:

Correlation: To calculate the relation between classified information items based on quantitative measures (e.g. time, distance and/or other class dependent attributes), thereby establishing numerically comparable (strong or weak) correlations between them.

Link: To establish a reference between classified information items using qualitative measures from Cross-referencing (i.e. two information items can be linked if, in any respect, they share a third information item, either observed or from background information).

If Correlation / Link results in evidence about the identity of the related objects their respective information is fused reducing the number of system objects (recognised or assumed domain elements). Cross-referencing of the Collation Step may be interpreted as an initial Link or Correlation step (raw correlation / link).

Finally, the received information items are aggregated one with the others. Aggregation/Integration will fuse different information items (elements of the problem domain) in order to deduce new and more valuable information items. The input information items that are fused together all satisfy the pattern of an object of a more complex domain class according to basic intelligence (background knowledge, doctrines, etc.) and on its related RFI (requests for information).

The following functions of Aggregation / Integration have been identified as candidates for automation:

- Carry out pattern analysis based on time and place in order to recognise activity / behaviour patterns and thereby deduce more complex activities and manoeuvres. Based on the preceding premise create aggregation hypotheses by integrating classified information, correlated / linked information, and intelligence about entity locations and behaviour with deployment and movement templates.

- Group and position information items into the organisation and thereby deduce the hierarchical structure of the organisation.
- Trigger an alarm function when new significance has been produced/ deduced/ found.

Because of the restricted resources available to the RTGonIFD it was obvious that not all above mentioned processing steps identified as candidates for automated processing could be implemented into an experimental system in the scope of this Task Group. Therefore the concept for the demonstration system is to couple available tools and demonstrators to cover as much steps as possible of the processing chain, starting with *Collation* and *Analysis*, those steps which are related to the transformation or linking of free-text into/with structured information.

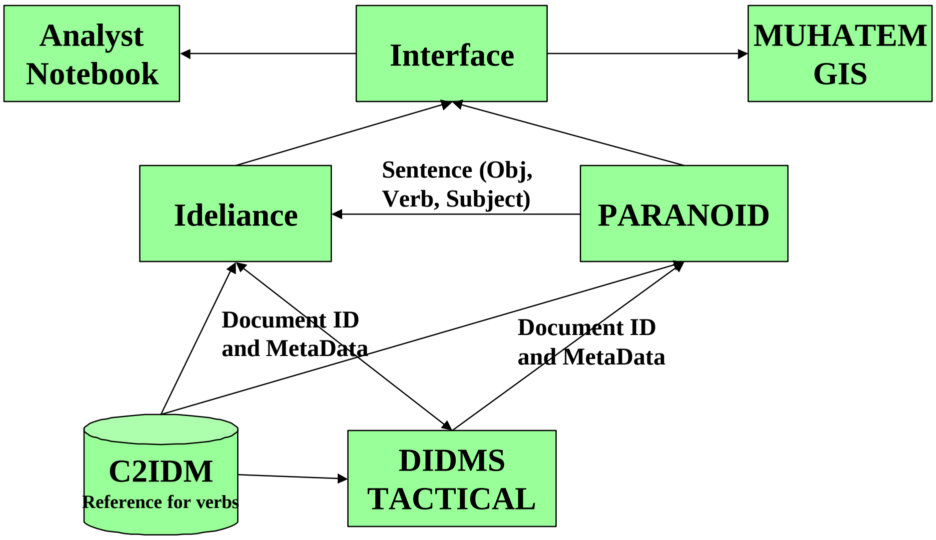

Figure 13: Demonstration high level technical design

Currently, the tools are being integrated as shown in Figure 13. After finalizing the integration process of the system components, the automated/semi-automated chain of intelligence information flow will be completed. It will then demonstrate that intelligence information coming as unstructured and/or structured data (text) can be imported, analysed, evaluated, integrated and finally represented in an efficient structured way. Thus the overall automated/semi-automated processing will produce a well-organised set of information about the situation and by that proof to be a valuable support tool for efficient and correct processing of the increasing amount of intelligence data.

5.0 CONCLUSION

The analysis of the new operational requirement makes it evident that the traditional way to process information is no longer sufficient and must be supported by new methods adapted to urban environment and unstructured data. Semantic nets, which are a highly structured data representation technique, give promising results for generating machine readable intelligence data, close to the operational way of understanding and processing information. More than that, the ability to run such a system without the assistance of technicians is a huge advantage in the context of highly evolving urban situations. Thus, it clearly appears that those parts of intelligence processing related to information management, especially in Collation and Analysis, can be performed with little human involvement, when all environment related information (immaterial, psycho-socio-cultural,…) have to pass through the operational filter. Additionally, and fortunately, the average operational personnel will be perfectly able to cope and manage support systems using semantic nets and defaults reasoning, two approaches which perfectly correlate to human mental processing. In that way, a delineation between these complementarities could be

highlighted and would ease a good conceptualisation of processing matters.
The objective of the RTO Task Group on Information Fusion Demonstration (RTGonIFD IST-038/RTG-016)) is to demonstrate that computer assisted information fusion is possible. Tools and software contributed by the participating nations will be used to proof the developed concept for automated support of Collation and Analysis in Intelligence processing. This will be done using a semantic net approach and other ontological techniques of link analysis. The demonstrator will provide a domain dependent and context oriented structured information bases to the human operator and set him free from tedious and time consuming steps of problem dependent information structuring, which is a precondition for automated high level information fusion. A most complete and thoroughly automatically established information bases will add real value to Data Fusion JDL level 2 (Situation Assessment/Integration) and JDL level 3 (Impact Assessment/Interpretation) in the conduct of the intelligence process, either in military operations or in defence against terrorism.

REFERENCES

[1] Alberts, D. S., et al., *Understanding Information Age Warfare*, DoD Command and Control Research Program, August 2001.

[2] Anderson, John R., *Kognitive Psychologie: Eine Einführung*, Spektrum, Akad. Verlag 1996

[3] Biermann, J., *Challenges in High Level Information Fusion*. In: The 7th International Conference on Information Fusion, June 2004, Stockholm, Sweden.

[4] Biermann, J., *A Knowledge-Based Approach to Information Fusion for the Support of Military Intelligence*, NATO RTO Symposium on Military Data and Information Fusion IST-040 / RSY-012, Prague, Czech Republic, 20. - 22.10. 2003.

[5] NATO Standardization Agency (NSA), *AAP-6(2002) NATO Glossary of Terms and Definitions*, http://www.nato.int/docu/stanag/aap006/aap6.htm, Dec. 2003.

[6] NATO, AJP 2.0 *Allied Joint Intelligence Counter Intelligence and Security Doctrine*, NATO/PfP Unclassified, Ratification Draft 2, 2002.

[7] Konar, Amit, "*Artificial Intelligence and Soft Computing: Behavioral and Cognitive Modeling of Human Brain*", CRC Press LLC, 2000.

[8] Davis, R., Shrobe, H. and Szolovits, P., *What is a Knowledge Representation? AI Magazine*, 14(1):17-33, 1993.

[9] *Joint Inquiry Into Intelligence Community Activities Before And After The Terrorist Attacks Of September 11, 2001*. Report of the U.S. Senate Select Committee On Intelligence and U.S. House Permanent Select Committee On Intelligence. December, 2002, S. REPT. NO. 107- 351 H. REPT. NO. 107-792, http://www.gpoaccess.gov/serialset/creports/911.html .

[10] Steinberg, A.N., Bowman, C.L., White, F.E. , *Revision to JDL Data Fusion Model*, Joint NATO/IRIS Conference, Quebec, October 1998.

[11] G. Thibault et al. (eds.), *RTO Task Group Report, NATO IST-015/ Task Group 004 On Information Fusion, Final Activity Report 2000 - 2002*, Dec. 2002.