

Revisiting the JDL Data Fusion Model II

James Llinas¹, Christopher Bowman², Galina Rogova³, Alan Steinberg⁴,
Ed Waltz⁵ and Frank White⁶

Abstract - This paper suggests refinements and extensions of the JDL Data Fusion Model, the standard process model used for a multiplicity of community purposes. However, this Model has not been reviewed in accordance with (a) the dynamics of world events and (b) the changes, discoveries, and new methods in both the data fusion research and development community and related IT technologies. This paper suggests ways to revise and extend this important model. Proposals are made regarding (a) improvements in the understanding of internal processing within a fusion node and (b) extending the model to include (1) remarks on issues related to quality control, reliability, and consistency in DF processing, (2) assertions about the need for co-processing of abductive/ inductive and deductive inferencing processes, (3) remarks about the need for and exploitation of an onto logically-based approach to DF process design, and (4) extensions to account for the case of Distributed Data Fusion (DDF).

Keywords: Data Fusion, JDL Model, information fusion, fusion models

1 Introduction

In 1999 Steinberg, Bowman, and White [1] published the first paper formally addressing various extensions to the well-known “JDL” Data Fusion process model [2]⁷. That paper began by revisiting the basic definition(s) of Data Fusion (DF) both conceptually and in terms of the “Levels” that are characterized in the original JDL model, including the introduction of a “Level 0” to the model. In addition, the notion of estimating perceptual states rather than states solely related to physical objects and entities was also introduced. The last major part of that paper described the need for and an approach to standardization of an engineering design methodology for DF processes, citing the prior works of Bowman [3], Steinberg and Bowman [4], and Llinas, et al [5] in which engineering guidelines for DF processes were elaborated. The Data Fusion and Resource Management (DF&RM) Dual Node Network (DNN) technical architecture proposed by Bowman for DF in 1980 [6, 7] and extended to response management (RM) in 1993 [8] provides software components, interfaces, and a software development engineering methodology for DF&RM.

There are various other fusion processing model-structures that have been asserted; while our purpose here is not to include a comprehensive review of such models, we comment on a few others, in chronological order as published. In 1997, Dasarathy [9] put forward ideas associated with the notion that there were three general levels of abstraction in fusion processing, the data level, the feature level, and the decision level. Accordingly he published a model that characterizes the processing at and across such levels. This model, while providing a useful perspective, is not as comprehensive in scope as the JDL model. Over 1999-2000, Bedworth and O’Brien published their “Omnibus Model” [10] that combines aspects of the Observe-Orient-Decide-Act or “OODA” decision/ control loop with the “Waterfall” software development process to make aspects of feedback more explicit, and is claimed to enhance a DF process description by combining a system -goal point of view with a task-oriented point of view.

In 2002, Salerno addressed various issues and perspectives on Information Fusion processing in [11], which focused primarily on higher-levels of abstraction in fusion-based inferencing, and draws in part on Endsley’s well-known works on Situation Assessment, e.g., [12,13]. In 2003, Blasch and Plano [14], suggested an extension to the (revised) JDL Model to include a “Level 5”, labeled as “Human (or User) Refinement”, which addressed the issues associated with the human interface to and control of the DF process; the model at this point would thus have 4 core fusion Levels (L0-L3) and 2 extension Levels (L4-L5).⁸ All of these models have some insights to offer about data and information fusion processing, and our purpose in mentioning them is not to argue for one or the other but to help put the current paper in context.

The JDL Fusion model is a functional model and was motivated by confusion in the community over the many elements of fusion processes. The model was developed to provide a common frame of reference for fusion discussions and to facilitate understanding and recognizing the types of problems for which data fusion is applicable, and also as an aid to recognizing commonality among problems and the relevance of candidate solutions.

¹ Research Professor, University at Buffalo, Buffalo, NY, USA, llinas@eng.buffalo.edu

² Consultant, Data Fusion & Neural Networks, Colorado, USA, cbowman@indra.com

³ Encompass Consulting, Honeoye Falls, NY, USA, rogova@rochester.rr.com

⁴ Technical Director, Utah State University Space Dynamics Lab, Utah, USA, Alan.Steinberg@sdl.usu.edu

⁵ Technical Director, Intelligence Programs, General Dynamics - Advanced Information Systems, Ann Arbor, MI, USA, ed.waltz@gd-ais.com

⁶ Director, Program Development, US Navy SPAWAR Systems Center, San Diego, CA, USA, frank.white@navy.mil

⁷JDL: Joint Directors of Laboratories, a US DoD government committee overseeing US defense technology R&D; the Data Fusion Group of the JDL created the original JDL Data Fusion Model

⁸C. Bowman and A. Steinberg are writing a companion paper to this one that refines the L0-L3 as Fusion Levels, defines Level 4 as being within the “Dual” resource management “Levels” and defines Level 5 as part of the human-machine interface functions; see also Section 3.1.1 herein.

Much of its value derives from the fact that identified fusion functions have been recognizable to human beings as a “model” of functions they were performing in their own minds when organizing and fusing data and information. It is important to keep this “human centric” sense of fusion functionality, since it allows the model to bridge between the operational fusion community, the theoreticians and the system developers. The framework of the model has been useful in categorizing investment in automation and highlighting the difficulty of building automated processes that provide functionality in support of human decision processes, particularly at the higher levels where reasoning and inference are required functions.

2 Motivation for the Extensions

Sensibly all the remarks contained in the present paper are motivated by both residual imprecision in the model definitions and by recent and ongoing visions of future operational concepts and associated informational needs by the DoD and Homeland Defense community. Some of the applicable vision documents include: “Navy’s Sea Power 21”[15]; Air Force’s “C2 Constellation”[16], and the Joint Forces Command’s “Battle Management Command and Control (BMC2)”[17]. The types of concepts and informational needs depicted in these vision statements include, inter alia:

“Common (or Consistent, or Relevant or Single Integrated or User Defined) Operational Picture”
“Network-Centric Warfare”
“Dominant Battlespace Knowledge”
“Operations Other Than War”
“Asymmetric Warfare”
“Information Warfare”
“FORCEnet”

Each of these topics have been addressed in many papers and other publications; a good general reference is the publications list of the National Defense University [18]; see also Joint Vision 2020 [19], the official report on Network Centric Warfare [20], and a good bibliographic site on Asymmetric Warfare [21] for expanded, detailed ideas on these concepts and technological and informational needs. All of these future visions are based on modern information and networking technologies, evolving INTERNET and Web based paradigms. The technology and concepts are evolving rapidly in the global WWW and the evolution toward client - services processes including: web services, agent based computing (broadly defined), the semantic web, with accompanying ontology development, all offer the potential for a widely distributed information environment. In this rapidly evolving domain, network services are receiving the bulk of the attention at present but the importance of the *information services* is becoming recognized. These services will center on content including data pedigree, metadata and context, among other factors; how the information services will contribute to meaning and relevance of information are key issues related to the design and development of

such services. Either explicitly or implicitly, Data and Information Fusion are cited in these various works as a central, enabling technology and information service providing capability toward the realization of these visions, and satisfaction of inherent informational needs. From the fusion community perspective, the ability to achieve successful fusion is equally dependent on the pedigree, metadata and context services and the conventions and standards that evolve to prescribe them. Therefore the functional descriptions of the fusion process as embodied in the JDL model are critical in the envisioned distributed information environment. How they will be described and how fusion can function usefully as a service in this distributed web environment is a major challenge and a principal motivator for revisiting and evolving the JDL Data Fusion Model in this paper.

The fusion community has certainly reacted to these perspectives with many relevant, focused technical papers that address many of the fine points embedded in these visionary statements, most offering specific solutions to specific problems. But so far as we are aware there has been no paper that steps back from these visions and needs and examines the underlying implications for the primary conceptual and semantic DF model for the international fusion community, which has been the JDL Model. Here, we offer our collective views on a wide range of impacts on the JDL Model of these new visions and needs of the defense and military communities. In proposing this in-depth rework we feel it is important to cleave to the original purpose of the functional model and its importance to multiple communities, including its relevance to humans who are increasingly operating in web based networked environments and struggling with these issues. This model revisit and revision then has the goal of retaining a functional flavor and structural simplicity to remain accessible to all communities of interest while revising some definitions and concepts to be more relevant and up to date.

In this paper, further extensions of the JDL Model-version (*in particular*) are proposed for community discussion, with an emphasis in four areas: (1) remarks on issues related to quality control, reliability, and consistency in DF processing, (2) assertions about the need for co-processing of abductive/inductive and deductive inferencing processes, (3) remarks about the need for and exploitation of an ontologically-based approach to DF process design, and (4) discussion on the role for Distributed Data Fusion (DDF).

3 Extensions to the Basic DF Process Model

3.1 Nodal and Fusion Level Processing

There are several subtleties that we wish to address in this section. We first reference the 5-Level DF Model of [1] that enlarged the 4-Level original JDL Model [2] with a “Level 0”, at this point choosing not to incorporate the suggested “Level 5” of Blasch and Plano [14] as, to our understanding, incorporation of this Level has not yet achieved common usage. We

also reference Bowman's DF "tree paradigm" paper in [3] for the characterization of the generic processing at a typical DF Node. Figure 1 shows the former [Note that the 1998 JDL Model shows the Level 4 and the DBMS as being only part-way in the Data Fusion domain.], and Figure 2 the latter.

As properly pointed out in [1], data association and assignment is occurring within each Level of the JDL Model, although the JDL diagrammatic structure of Fig. 1 seems to emphasize the State Estimation step of Fig 2. It could be argued however that the entirety of the "nodal" construct of Bowman in Fig 2 is

occurring within or among each Level; e.g., that Data Preparation (i.e., Common Referencing) is also necessary to the extent that new inputs to a given Level (i.e. distinct to that Level)⁹ must be reconciled with all the data inherent to other Levels, as there can be information flowing across all Levels (or nodes) all the time (more on this point later). That is, Bowman's depiction of "nodal" processing could be occurring multiple times in each level, and across multiple "Levels" within a node as well, depending on the complexity of the process.

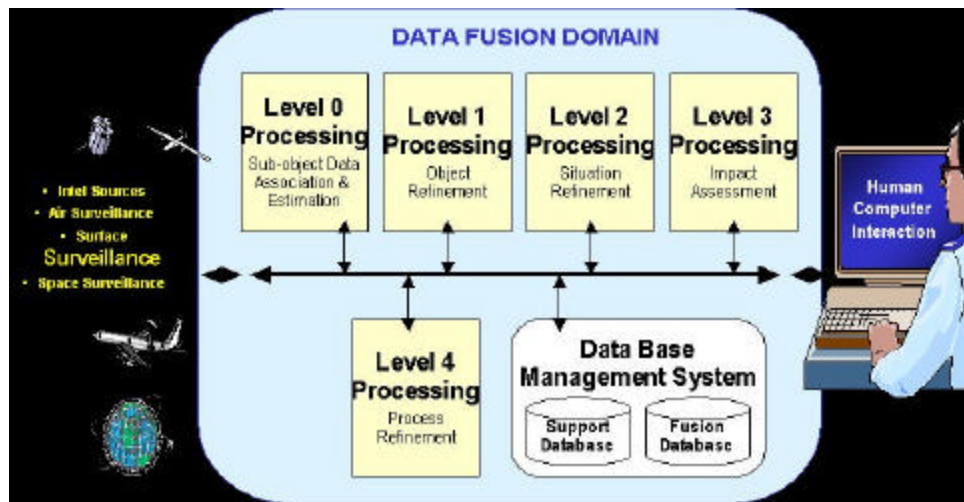


Fig. 1. JDL data fusion model (1999 revision).

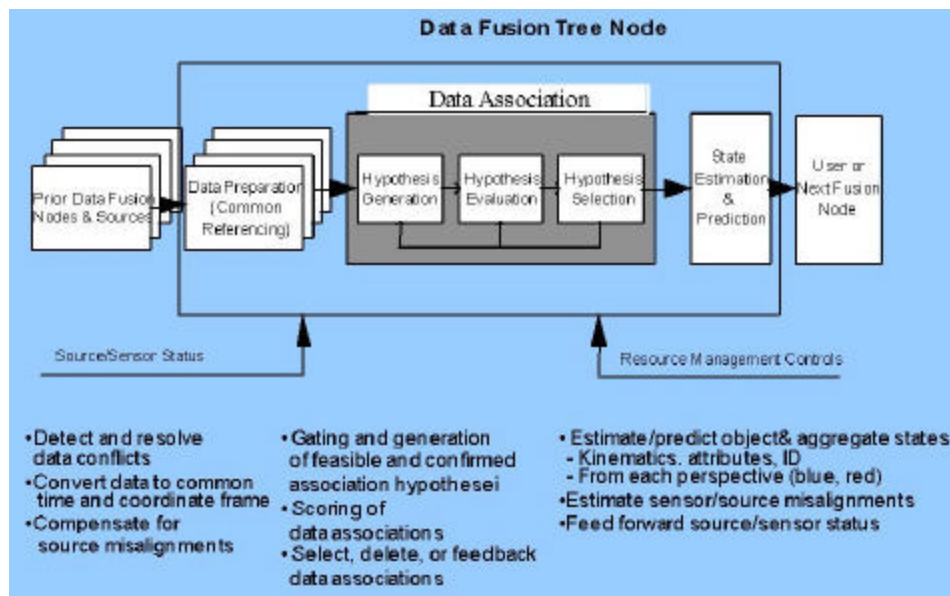


Fig. 2. Bowman's fusion node characterization.

⁹Many DF applications will have a "process manager" function or equivalent that determines which data goes to which Level, as it can be quite typical that certain data types are only useable by a given Level in accordance with the estimation needs of that Level.

3.1.1 Remarks on the Fusion “Levels” in the Current Model

The Levels in the JDL Model were originally the result of a partitioning scheme that was based on the combined and interdependent effects of (a) changing levels of abstraction and (b) changing levels of problem-space complexity. The Levels thus were defined as:

- Level 0: Estimation of States of Sub-Object Entities (e.g. signals, features)
- Level 1: Estimation of States of Discrete Physical Objects (e.g. vehicles, buildings)
- Level 2: Estimation of Relationships Among Entities (e.g. aggregates, cuing, intent, acting on)
- Level 3: Estimation of Impacts (e.g. consequences of threat activities on one’s own assets and goals)

When attempts are made to elaborate on these definitions for additional detail, various points of confusion have often arisen, e.g., subtleties about where state estimation is distinguished from prediction, or whether both estimation and prediction functions occur at all Levels or particular Levels, among various other points. In their forthcoming paper, [41], Bowman, Steinberg, and White will recount the lines of thinking associated with each of the Levels, describe Performance Evaluation as a Fusion Level, and introduce the dual resource management levels. We include here a very brief synopsis of their summary proposal regarding a reexamination of the Level structure. They propose a *partitioning by types of information being associated and by the types of estimated outputs* – they argue that this has the virtues of clarity, usefulness, and respect for existing usage. This yields the following characterization:

Table 1: Proposed process-based partitioning scheme

Data Fusion Level	Association Process	Estimation Process	Product
L.0 – Signal Assessment	Assignment (Observation-to-Feature)	Detection	Estimated Signal State
L.1 – Object Assessment	Assignment (Observation-to-Entity)	Attributive State	Estimated Entity State
L.2 – Situation Assessment	Relationship (Entity-to-Entity)	Relation	Estimated Situation State
L.3 – Impact Assessment	Evaluation (Situation to Actor’s Goals)	Game-Theoretic Interaction	Estimated Situation Utility
L.4 – Process Refinement	[Planning (Resource to Task)]	[Control]	[Action]

Under this partitioning scheme, the same entity can simultaneously be the subject of Level 0, 1, 2, and 3 fusion processes. Entity features can be estimated from one or more entity signal observations (e.g., pixel intensities, emitter pulse streams) via a Level 0 data

preparation/ association/ estimation process). The identity, location, track and activity state of an entity (whether it be a man, a vehicle, or a military formation) can be estimated on the basis of attributes inferred from one or more observations; i.e. via a Level 1 data preparation/association/estimation process. The same entity’s compositional or relational state (e.g. its role within a larger structure and its relations with other elements of that structure) can be inferred via Level 2 processes. Thus, a single entity – anything with internal structure, whether man, machine, or mechanized infantry brigade – can be treated either as an individual, subject to Level 1 observation and state estimation – or as a “situation”, subject to compositional analysis via Level 2 entity/entity association and aggregate state estimation. The impact of a signal, entity, or situation on the user goal or mission can then be predicted based upon an association of these to alternative courses of action for each entity via a Level 3 process. The 1999 “Revised JDL Fusion Model” paper [1] recognized the original Process Refinement Level 4 function as a Resource Management function (i.e., thus falling within the Resource Management model levels, see [41]) which is why an X has been inserted in the above table. Bowman and Steinberg will argue that the usefulness of these fusion Levels is due to the significant differences in the types of data, models, and inferencing necessary for each Level. Once again, processing at these Levels are not necessarily performed in order (e.g., Level 3 can be performed upon Level 1 entity state estimates), and any one Level can be processed on their own given their corresponding inputs.

In this paper we will not overlay these new Level characterizations with the largely-functional discussion about the nature of and interactions among the fusion functions, as to do so we feel would over-complicate the paper; but we certainly agree that the approach to and character of such Levels needs reexamination as well, and put the main ideas here for contemplation by the reader.

3.2 Inter-Level Information Exchange

The idea of inter-Level information and control flow is not very explicit in the traditional JDL Model; this is in part because the model is not an architecture for building fusion systems, and in part because of a desire for diagrammatic simplicity. This idea, however, has been illuminated in the DNN architecture papers [3, 6, 7] that describe the various types of interlaced fusion and management node networks. It has certainly been emphasized that the JDL Model does not imply sequential processing within or across Levels and that there is (some type of, unspecified) feedback across Levels. The notion of inter-Level “informing”, controlling, and exploitation can in fact become quite complex in certain applications, and has similarities to the complexities of peer-to-peer internetworking processes at multiple levels of abstraction. In the course of one Level informing another, there should be some sense of added value or utility balancing the negative

aspects of the additional processing complexity and time delay of enabling such feedback. Moreover, the possibility of such feed-back raises concerns for maintaining consistency in inferencing across levels. This hints at a need for inter-Level adjudication management processing similar to what (we argue later) is needed for the DDF case. So the general inter-Level processing notion can be depicted as in Fig. 3 (start at “Level n” in examining this figure).

Review of Fig. 3 reveals the type of potentially-endless, cyclic inter-Level operations that could occur, and that a “stopping criterion” is necessary (e.g., a consistent and satisficing or optimal computation of Value or Utility within a process management function) so that these operations have a sensible endpoint. So, while it can be argued that inter-Level exchange and exploitation are generally desirable, any particular approach to enabling these processing operations must be driven by cost-value/utility tradeoffs. It is such trades that are accomplished in the fusion and management node network optimization phase of the software development. The importance of this inter-Level processing should be made more visible in the basic Model as an inherent function of fusion processing.

3.3 Adjudication, Conflict Resolution, and Belief Change

The inter-Level adjudication, or conflict-resolving, processing mentioned above has two dimensions: (1) one is with regard to the specifics of discrete conflicts in what could be called “atomic” information being exchanged between Levels (such as “a” in Fig. 3), and (2) another is with regard to conflict at a meta-level of estimation or inferencing (i.e. at the nodal or “system” output point), which can imply a need for a Belief Revision or Belief Change type function at that point. Since the latter is a meta-level interpretation/ assessment issue, we feel that the Belief Change function should be applied at the system-output (i.e., state estimation) stage for any given fusion node, with “local” or discrete

conflicts resolved by the adjudication management function that employs other conflict-resolving techniques.

So, at this point, we now have the within-Level and inter-Level processing depicted as in Fig. 4.

Here, Level m is a preceding Level to Level n (i.e., involved with inferences of lower abstraction), and so Level n information provides, among other possible information, contextual information that may be helpful to improved Level m state estimation. The Level n information would first be adjudicated for consistency and checked for added value, and if of adequate value, would be incorporated in Level m’s state estimates. Details of these inter-Level processing interactions would be defined on a case-by-case basis. In turn, Level m provides upward flow of its information to Level n, with either or both of its cyclically-enhanced updates as well as data of a more atomic level.

3.4 Fusion Node or System-Level Output Processing

The JDL Model does not speak in any detail to the notion of how the overall output of a given fusion process may be generated and controlled. We believe that there are at least two factors to think about in this regard: output Quality and output Consistency. Although Fig. 4 depicts an evolving, controlled computation of “Value Added” in inter-Level processing operations, we believe that overall Value or Quality would be partitioned into a hierarchy of “local” values and of “system” quality, with the former (e.g. as in Fig. 4) gauging only how inter-Level information exchange is affecting value, whereas the latter is concerned with satisfying system-level criteria so that the nodal output is either directly satisfying system needs (in a centralized fusion, single-node case) or satisfying the needs of successor fusion nodes in a distributed fusion, multi-node case. We also see that it is system/nodal quality control that is integrated with Process Refinement operations. For example, while it is of course possible (even desirable)

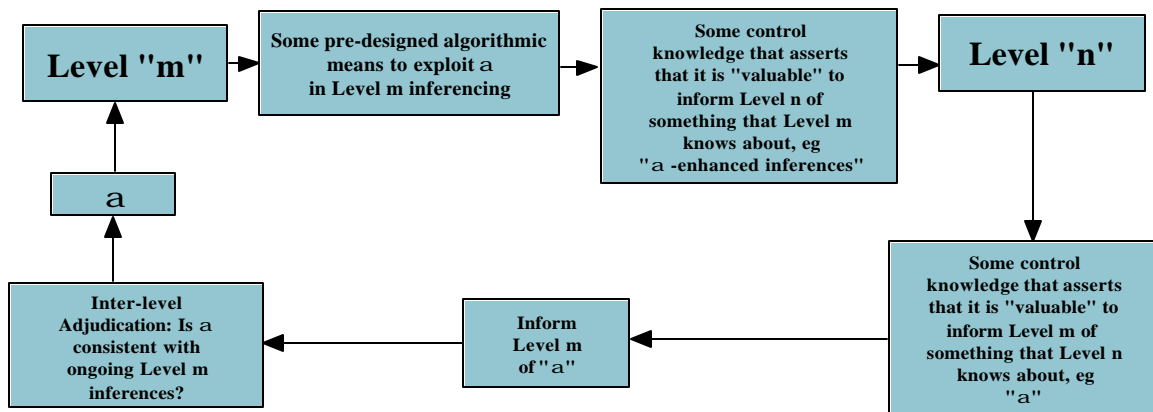


Fig 3. Notional inter-level “informing and exploitation”

that any given fusion or node Level would generate its own local “service requests” for additional information or some other need, it seems correct to have higher process management nodes arbitrate those multiple requests (for local “Added Value”) with the status of system/nodal-level output Quality, and the needs for improving that quality, i.e. with a system-level service request. While we will argue for this preference, such details are of course a case-specific design decision.

On the Consistency side, we hinted previously at the need for Belief Change (BC) functions at the output level to reconcile meta-level consistency. By introducing the concept of BC here, we imply that we mean Consistency in the sense of a rational change to the currently-existing, overall nodal state estimate, which could be considered as the node’s current “epistemic state”. We are concerned with how and whether to alter this current state based on the new information contained in the just-formed, new state estimate for this node. Sub-processes involved with BC include a means to determine that a contradiction in fact exists, what the sources of the contradiction are, and the justification for each state estimate and a strategy for reconciliation. (Notice that there may be subtle but significant differences between the operations involved with adjudication and in belief contradiction-determination for BC).

Two special cases of BC are usually considered in the literature: Belief Revision (BR) and Belief Update (BU) (see, e.g., the collection of papers in [22]). The Belief Revision process modifies existing estimates about a particular time “t” based on new information about the same point in time “t”, i.e., BR refers to adjustments in the interpretation of a state estimate at a given time. Belief Update modifies existing information about the world at time “t” as motivated by new information from time “t+1” to describe the world at time “t”, so that BU refers to a dynamic situation, in which the world is evolving in time. While BR decides

what beliefs should be discarded to accommodate new information, BU attempts to decide what changes in the world led to this new information.

Traditionally, most of the methods for Belief Change obey the following three rationality principles [22]:

1. *Consistency*: revised epistemic state should be consistent.
2. *Minimal Change*: revised epistemic state should be as close as possible to current epistemic state.
3. *Priority of Incoming Information over existing information*.

Basically, these principles suggest that the process of BC retracts some of the old information after obtaining new information to make the epistemic state consistent. The BC process is based on the concept of “epistemic entrenchment”, in which beliefs are ordered according to our willingness to give them up. If some beliefs must be removed in order to accommodate some new information and keep the belief set consistent, the less entrenched belief will be given up, while the more entrenched are preserved. In a numerical setting, a function over sets of possible worlds representing the credibility distribution over the set of possible worlds/interpretations plays the same role as epistemic entrenchment in the symbolic frameworks. The arrival of new information changes the credibility distribution (epistemic entrenchment) over possible worlds/sentences and creates new prioritization. The principle of minimum change means minimizing some kind of distance between the old and the new credibility distributions. The principle of priority of incoming information changes credibility of certain possible worlds and makes them equal to zero after processing new information.

In the case when *uncertain* information about a static world (the case of BR) is coming from different (often unreliable) sources at different times, and

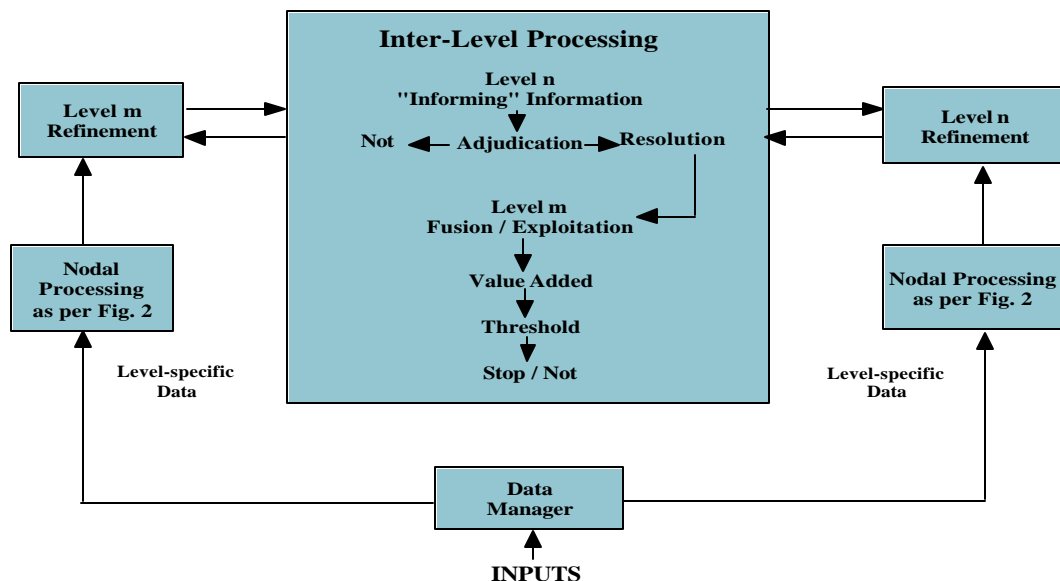


Fig. 4. Extended version of within-level and inter-level fusion processing.

especially in a distributed situation, the priority of incoming information is not justified since the chronological sequence of information has nothing to do with its importance. In [23] the principle of priority of incoming information is replaced with the principle of “Recoverability”, which “requires that any previous beliefs belonging to the current epistemic state remain with that state if it is consistent with it”. In Belief Revision referring to static multi-source distributed cases, pieces of knowledge may not only be abandoned (non-monotonicity) but also rescued (recoverability) after new incoming information. This is achieved by maintaining two types of knowledge: (1) the “knowledge background”(KB) comprising a set of all the pieces of knowledge even inconsistent available to the reasoning agent and (2) the “knowledge base”(B), a maximally consistent and currently preferred part of knowledge repositories. Revision here is performed based on overall KB, which allows also to reject information even if it is consistent with the current B.

In *dynamic* situations, a Kalman-like approach to BU at each fusion level can be adopted (“model-based” BR), see [24]. In this case revision consists of a prediction step based on a selected model of the evolution of the world and a revision step, in which the predicted state of the world is modified based on incoming information while taking into account its reliability. Transition from the old epistemic state to the new one obeys the principle of minimum change and incoming information can be rejected if this new state is too far from expected.

In *distributed* multi-agent dynamic situation both BR and BC can be justified and one of the problem here revolves around the question of how to determine whether the new state estimate, if it contradicts the current one, reflects a true change of a world state or not. In both BR and BU cases, the reliability of incoming information affects the credibility of possible worlds and has to be taken into account while building the BC process and adjusting the credibility distribution to accommodate new information. For example, new information can be rejected if it inconsistent with absolutely reliable prior credibility distribution or background knowledge.

At the nodal or system-level output level then, we have processing operations as shown in Figure 5. This figure shows the within-Level processing and value-control (details in Figs. 3, 4), along with the quality and consistency checking at the nodal output level, both interacting with and generating service requests to processing management. In the end, the nodal/system outputs will have been both quality and consistency checked before being passed to the system or inter-nodal output level. In discussing the notions of Quality and Consistency control of a fusion node’s output, we are in effect discussing when to “post” the nodal result for system use. We are aware of the issue that such control inherently implies a delay and that in many defense applications time is the most critical resource to manage. The intelligence community, for example, has been

considering the change from a “TPED” (Task, Process, Exploit, and Disseminate) intelligence cycle to a “TPPU” (Task, Post, Process, and Use) cycle, but the appropriate strategy for “posting” or outputting are case-specific.

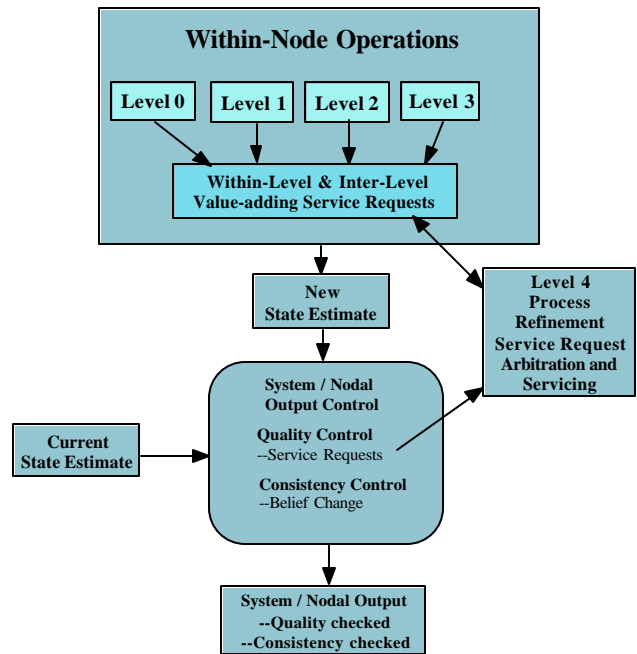


Fig. 5. System/nodal output processing.

3.5 Issues Related to Reliability

The performance of the particular fusion node as well as internode processes such as Belief Change operators, highly depends on a priori understanding of the reliability of input information, either from sensor/sources or from another fusion node. The concept of reliability, however, has various interpretations and representations, with no well-established consensus. Reliability, in a sense reflecting a second-order uncertainty in the quality of an input, is important in fusion applications because most of the common fusion operators are symmetrical and based on optimistic assumption that the sources are equally reliable [25]. Clearly this can often be an incorrect assumption, so that it is very important to evaluate the reliability of information (absolute or relative) coming into fusion processes. Equally, given a set of heterogeneous inputs as regards their reliability, it is desirable to have an estimation process that computes, estimates, or assigns the reliability of a node’s fusion outputs, to propagate these values into the overall fusion process for improved performance and effectiveness. Reliability of information depends on how well input data are represented, how good and adequate are the uncertainty and fusion models used, and how accurate and appropriate or applicable the prior knowledge is. There are different situations that can be considered while dealing with reliability of sources [26]: (1) it is possible to assign a numerical degree of

reliability to each source, (2) a subset of sources is reliable but we do not know which one; (3) only an order of the reliabilities of the sources is known but no precise values. Dealing with these situations calls for one or all of the following strategies of incorporating reliability of data, knowledge, and information into fusion processes:

- Strategies for identifying the quality of data input to fusion processes and elimination of data of poor quality.
- Strategies for modifying the data and information by considering their reliability before fusion.
- Strategies for modifying the fusion process to account for the reliability of the input.

Selection of one of these strategies depends strictly on the problem to be addressed, the fusion process to be considered, and the global knowledge about the sources and the environment. Incorporation of reliability into the fusion process gives “richer behavior” to the fusion system while producing many theoretical and practical problems not very often addressed in the data fusion literature, largely concerned with modeling information credibility. Among these problems are the problem of estimation of reliability of sources and their temporal analysis; the problem of interrelationships between reliability of information sources and their number and fusion results; the problem of incorporating contextual information into evaluating source reliability; and the problem of incorporating reliability into fusion processes.

4 Abductive/Inductive and Deductive Inferencing

Overall Intelligence and targeting processes (that include humans) have always included the precursor process of inductively *discovering* the characteristic signatures of target classes, followed by the deductive *detection* of those targets using the previously characterized signatures. The data fusion process implements the latter, largely deductive, detection process. The former process of discovery of previously unknown signatures is performed by analyzing known representative targets (e.g. captured military equipment, or intercepted military signals), or by the analysis of raw data about unknown or hidden targets (e.g. the communication and financial transactions of a single terrorist cell).

Data fusion processes have been largely applied to symmetric military warfare in which long-term strategic target development processes (in military labs, on ranges) have developed the signatures or deductive model-based templates describing the component targets of fielded adversary forces. Asymmetric adversaries, on the other hand, are quite unpredictable in their behavior, tactics, weapons, and choice of targets. In these environments, the deductive and model-based traditions of data fusion processing simply will not provide the capabilities needed in support of military planning and real-time decision-making.

Waltz has linked the inductive discovery process (known as “data mining” or “knowledge discovery” in commercial applications) to the data fusion process to provide an integrated process appropriate for difficult (hidden, unknown, complex, or adaptive) targets in which the discovery and detection processes must be tightly coupled. [27] The process follows the typical sequence of scientific discovery and proof, using a sequence of steps to conjecture, hypothesize, generalize and validate.

Consider the three step processes of discovery and detection, summarized for a simple counterterrorism example in Table 2.

1. **Discovery** – The analyst applies data mining tools to locate patterns of meaningful relationships in contacts, financial exchanges, associates, and concurrent activities of a terrorist cell. Correlated patterns are examined for relevance (e.g. causality, behavioral or intentional association, organizational or functional relationships, etc.), and *specific* potential terror cells are discovered. This discovery process requires the abductive, innovative form of reasoning to achieve the best explanation for the complex relationships in data. The hypothesized pattern generation, evaluation, and selection occurs within the data association function of a fusion node.
2. **Generalization and Validation** – The analyst generalizes, in a model, the discovered patterns of relationships and validates the model in the specific cases discovered. This stage applies inductive generalization, developing a general model from the small set of discovered cases. The model parameters are estimated within the state estimation function of the corresponding fusion node.
3. **Detection** – The validated model provides a target detection “template” in the automated data fusion process that monitors real-time raw data. The process automates deduction, to detect the presence of evidence that matches the template to detect other similar terrorist cells.

Most current data mining and generalization processes are very analytic and manually intensive, requiring a high degree of interaction between the analyst and the discovery tools. Others such as the abnormality detection tools built by Desrocher et al [28] have automated the pattern discovery, generalization, validation, and detection processes. Once a target class (e.g. a class of terrorist cell) can be described and modeled, the data fusion process permits a high degree of automation of the detection process for subsequent similar terrorist cells using multiple sources.

In Waltz’s earlier paper, the functional processes of an integrated data mining and model-driven fusion were depicted, and Fig. 6 illustrates one example operational implementation of the process. Real-time data feeds from three sources build three operational data stores, which

form the basis for a traditional data fusion vertical pipeline that derives objects (entities, their identities, and behavioral tracks), then situations and their impacts. The output of the process is a real-time visualization of the present situation. Concurrent with this process, relevant data from the operational data stores are extracted, transformed and loaded into a long-term data warehouse. The warehouse data are further cleansed and transformed to a common multidimensional data set to allow entity-relationship clustering by a data mining engine. The

mining process allows faint and complex signatures to be discovered, modeled and validated for insertion back into the data fusion pipeline.

(This process is called “publishing a solution” in the commercial data mining tool, Clementine by SPSS, Inc. The published solution is a data model that may be used in the data fusion operations, referred to as the “production processes” in the commercial terminology; see [29]).

Table 2: Integrated data mining and data fusion processes.

Step	Process	Reasoning Process	Example use of Typical Automated Tools
1. Discovery	Data Mining – Discovery of potential specific target and its characteristics in raw data sets	Abduction – Reason about a specific target, conjecturing and hypothesizing to discover best explanation of relationships to describe target. (Hypothesis creation)	Analyst uses data mining tools to locate patterns of relationships in contacts, financial exchanges, associates, and concurrent activities of terrorist cell.
2. Generalization and Validation	Target Modeling Generalization – Characterize target class in general model	Induction – Generalize fundamental characteristics of the target in descriptive model. Test and validate characteristics on multiple cases. (Hypothesis validation)	Analyst develops and refines quantitative model of terrorist cell behavior. The model is tested on additional data to evaluate its detection value using data mining tools.
3. Detection	Data Fusion–Detection of subsequent occurrences of target based on comparison with target models	Deduction – Test real-time and massive volume data against multiple target templates to detect (deduce) the presence of targets. (Hypothesis testing)	Real time raw data are ingested by automated data fusion tool to detect presence of evidence for other similar terrorist cells.

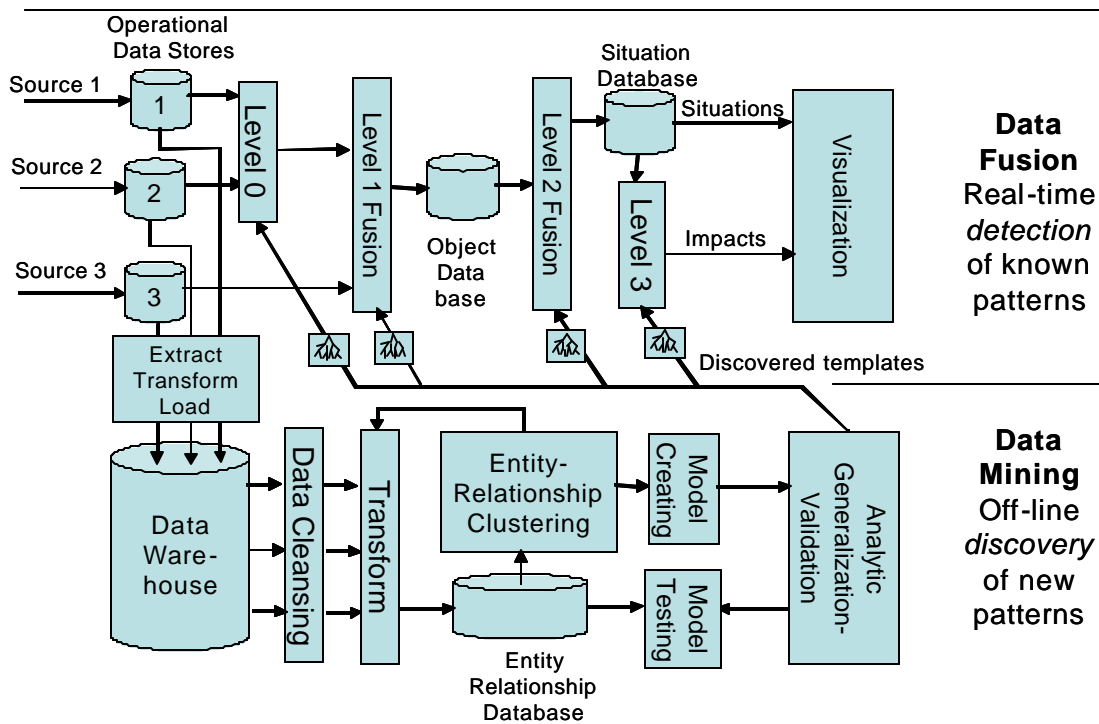


Fig. 6. Co-processing of abductive/inductive (data mining) and data fusion operations (based on Waltz, [27]).

There are several major challenges in incorporating the abductive/inductive techniques into a robust and automated data mining-fusion system. One issue is the development of a reliable method for automated discovery of relevant patterns in the flow of real-time data (not the highly manual data-base data mining operations seen in the mining literature). This is in essence a “locally abductive” process that results in the discovery of local, ephemeral patterns of interest; but even if that capability exists, there is the linked concern for whether decisions and/or actions would be taken on the basis of the discovery of such a pattern—this is a concept of employment issue, and is associated to the reliability of such discoveries. Related to this is the subsequent inductive process which would provide a framework for the generalization of detected patterns (or qualification of the detected patterns in some other context); this is a model-control problem, affecting the procedure by which a new pattern and related new model knowledge is “passed over” in a sense to the “deductive side”. Exactly how such newly-discovered patterns would be qualified as “valid” is a function of the problem domain. We suggest that these functions, necessary for fusing data in the modern, “post 9/11” world, be integrated and visible in an extended version of the JDL Model.

5 Ontology-based Data Fusion

5.1 Ontology (Ontologies)

The topic of ontology is far too large to elaborate on in any meaningful way here but we nevertheless start with definitional issues. As mentioned in Bowman et al [30], “Guarino and Giaretta [31] noted that the term “ontology” had at least seven distinct meanings in the literature, ...”. So, we will not assert a hard position on this point but say that ontology is a philosophical discipline, part of a branch of philosophy that deals with the nature, representation, and organization of reality. The development of “an” ontology rests on some logical or axiomatic-grounded theory which gives an explicit, domain-specific account of a conceptualization of things in the real world. There are also different kinds of ontologies such as content ontologies and process ontologies etc—those that are descriptive of the elements of interest in a domain are content ontologies, and are the focus herein.

One main benefit typically cited for basing the development of an information process on an ontological footing is: interoperability with other local and also external processes, which leads also to shared understanding.

5.2 The Role for Ontologies in Fusion Process Development

A content ontology specifies the concepts in the real-world as they are considered to be in an inherent, fundamental sense, i.e. it is a description of what is believed to be the true nature of things in the real-world

that are of interest. What is of interest in the real-world in the context of developing a fusion process are the spectrum of (true) states of interest as associated with a given task or mission. That spectrum is bounded by, and related to, the informational needs of users. So, one could say that the techniques used for developing a specification of user information needs, such as Cognitive Systems Engineering and Cognitive Work Analysis (see, e.g., [32]) used by the human engineering community, would *inform* the ontologist in developing a correspondingly-bounded ontological description of those needs as reflected in a set of ontologically-described states. There results from this exercise an ontological specification of the states of interest couched in an ontological representation language.

We consider the *relationships* within and between these states a crucial element of this ontology because they form the basis of the creation of *theories about the states* and inter-state relationships. Smith and Grenon [33] note that “There are many candidate formal-ontological relations, for instance: identity and difference, parthood and overlap, inherence and dependence, participation and location.”; the spectrum of relationship-types is well-summarized in [33], which reveals the richness in the range of relationships that is available to formally specify the states of interest in the real-world. In addition to the spectrum of relationship-types, there is also the choice of granularity with which such descriptions are formulated. As Smith and Grenon [33] say, “Thus we can examine the human body from a molecular, cellular, body tissue, organ, and whole organism perspective, and each of these can have equal claims to veridicality.” The theories mentioned previously are formal models of a mathematical or logical nature that are adequately-representative of the relationships defined in the ontology. Consider a “convoy”. A convoy could be specified as an aggregated set of objects (in an “aggregation” relationship) or it could be defined as what some ontologists call an “integral whole”, in which there are certain critical interrelationships among and between the vehicles such that, for example, if an appropriate “lead vehicle” is not present, then the convoy concept specification is unsatisfied, and thus the convoy is not considered to exist. In the first case, a clustering-model theory could be nominated to represent the notion of the “aggregation” relationship, but clearly in the second case a different theory or model would need to be formulated.

It is important to understand that these theories or models are couched in the real-world. A fusion process however never sees this real-world; it sees the sensor-and source-provided world, i.e. the observed world. As a consequence, the *fusion algorithms* that attempt to estimate the states of interest need to be fashioned as *observation-constrained variants of the real-world based theories*. Consider the convoy again. If our ontological specification of a convoy was as an aggregation, and our theory was a clustering-based theory, but there were certain limitations in our observational capability that did not provide all the information the theory required, we

would need to build a variant of the clustering model that accounted for these constraints but that was as “faithful” as possible to the aggregation-theory in some formal sense. The fusion algorithm is thus an observation-constrained variant of the theory.

What needs to be further developed is a deeper insight into the spectrum of relationship-types that are typical of those for any given fusion application. In particular, we see ontology as aiding the fusion community in moving ahead with Level 2 and 3 capability development because it will provide adequate *specificity* in defining the L2, L3 states and the relationships within and among those states, e.g. exploiting the range of relationships described in [33]. That is, we feel that a major constraint to moving forward in L2, L3 development has been a lack of specificity in state definition, e.g. not adequately analyzing and partitioning notions of “situations” into specific forms for which theories, models and, eventually, algorithms can be formulated.

The overall engineering methodology to develop the appropriate ontological specification of states of interest, the associated theories and, ultimately, the associated fusion algorithms also needs to be better defined (sensibly as an extension of the ideas in [3,6,7]). That methodology would no doubt involve both ontological engineering and cognitive systems engineering techniques, but also has to result in the construction of a software-based fusion process that reflects all of the principles applied. As these definitions and specifications unfold with higher specificity, they can be used either explicitly or implicitly in extended definitions of the JDL Model.

6 Distributed Data Fusion (DDF)

6.1 A Perspective on DDF

Sensibly all of the visionary documents mentioned in Section 2 describe an information environment for future mission operations that is distributed in some sense, so there is a clear need for the fusion community to address the issues and aspects of a distributed data fusion (DDF) information framework for evolving and future defense applications. Many recent papers have addressed various of the architectural aspects and algorithmic aspects for DDF, but the design space for a DDF is so wide in a parametric sense that it is difficult to assert any generalizations from these works. It also appears equally difficult for the military to define specifications for DDF (apart from broad generalizations such as achieving shared awareness, or the Common Operating Picture, etc) that can be engineered to, and so not unexpectedly the overall R&D picture and assessment of our level of understanding for engineering these systems in an optimal way is somewhat muddled.

6.2 DDF Issues and the JDL Model

Our purpose here is not to “solve” this dilemma but to nominate some DDF functions and factors that could be

added to the JDL Model so that these functions and factors are more broadly visible to the DF community, and possibly aid in R&D planning and execution. We divide the issues into architectural issues and algorithmic issues. The architectural issues we will address are: adjudication and pedigree, information-sharing strategies, and dynamic topology management; the algorithmic issue is that of a generalized need for local and network-specific algorithms that are *assuredly correct in the formal sense*.

6.2.1 Architectural Issues in DDF

One “message” in this section is that DDF cannot only be addressed by considering algorithmic requirements. If we consider that any node in a DDF architecture can only fuse two things: it’s locally-controlled (“organic”) inputs and the inputs somehow arriving from the network at large, then the entire framework by which any network-provided inputs arrive at any node is part of the DDF problem. Here we address just two of those factors, which generally relate to the nature and operation of the network communication infrastructure: dynamic topology management (DTM) and information-sharing strategies (ISS). DTM is required if for nothing else than to provide needed redundancy and graceful fault tolerance in a DDF wherein jamming, node destruction etc., can be ongoing and also, when the network is healthy, to provide for good performance as related to, for example, maximum nodal coverage, effective use of bandwidth, etc. As shown by Grime [34], DTM can in fact be used in certain circumstances as a basis for maintaining a provably-correct DDF algorithmic solution as well, when external factors alter the network topology. Alternately, ISS’s are the specified means by which the detailed flow of information circulates in the network; an ISS specifies who sends what to whom, how often, in what format, etc. It can be appreciated that an ISS is not only the result of technical considerations but also of socio-organizational type considerations (e.g., flow of command authority), and also that it too, like DTM, must have an adaptive part for the same reasons that a DTM must have one. As DTM most readily (but not only) relates to a wireless network, much of the research on DTM has been for “ad hoc wireless networks”, e.g. [35]; ISS’s have also been contemplated as enabled by so-called “publish-subscribe” strategies and by a variety of inter-agent communication strategies by the intelligent agent research community.

The third and final elements we address here are the adjudication and pedigree functions. The estimated state adjudication management function we relate to here, mentioned in Section 3.4, is that which reconciles system or nodal outputs, i.e. inter-nodal adjudication. We believe this would operate as part of a Belief Change function as described in Section 3.4. We define Pedigree as “an attachment to a message or communication between nodes that includes any information necessary to the receiving node(s) such that the receiving node fusion processing *maintains it’s formal and*

mathematical processing integrity". In regard to integrity, the two primary issues that have been addressed by the DF community that can create possibly pathological processing at a receiving node are: (1) double-counting (aka rumor-propagation, self-intoxication), wherein previously-accounted for information is processed as if it were new information, and (2) statistical dependencies, wherein a receiving node's fusion process is corrupted by the violation of statistical independence assumptions in its algorithms. We argue that role for both of these important functions with respect to other fusion and management functions should be made clear in the fusion model.

6.2.2 Local and Network Fusion Algorithms

In the most general case, any node in a DDF architecture can be either or both a sensing node and a fusion node, i.e. the node would have organic sensing and would fuse both that local data as well as the local data with network-provided data (or estimates). Since a given node would have the necessary insight to: (a) local sensor operating conditions, (b) flexibility in local sensor/process control and adaptation, and (c) required responsiveness, it is very likely that there will be a family of algorithms designed just for this local processing. Alternately, the knowledge about the network nodes, even network connectivity etc., at a local node may be quite limited, so the algorithms that either receive and process external data/estimates and/or combine those estimates with the local estimates are likely of a different sort than the local algorithms. For the case of target tracking, the community has for example developed the "Covariance Intersection", "Tracklet", "Information Filter", and other methods as special techniques for internodal fusion-based processing, in these cases to primarily deal with the pathologies of statistical dependencies as mentioned before (e.g., see [37,38, 39]). We suggest that an extended JDL Model should be developed that overtly depicts these additional functions necessary for the DDF case.

7 Summary

Realistic implementations of Data Fusion processing can be and often are complex. The JDL Model (and other models) however are not (and should not be) attempts to generalize a detailed processing architecture; their purposes are to serve the broad DF research community as frameworks of common understanding and pedagogy. But in addition, these models should make visible to all what the basic functions and issues are when thinking about real-world implementations; that is, they should serve researchers and developers as frameworks for contemplation about fusion processes and for sensitizing them to critical issues in advance of particular research initiatives or prototype developments.

Developing a model description that is adequately-detailed to encompass the issues and functions considered minimally necessary and sufficient to serve

these multiple purposes is no easy task. There is also the issue of how to achieve community consensus on any model description; this is a community infrastructure issue that has been raised previously [40]. This paper is an offering about issues and functions considered to be important to any generalized DF Model description for modern-day applications, and as a possible input to what we hope would be a community-wide effort to establish and control a community-standard model.

8 References

- [1] Steinberg, A.N., Bowman, C.L., and White, F.E., "Revisions to the JDL Data Fusion Model", in Sensor Fusion: Architectures, Algorithms, and Applications, Proceedings of the SPIE, Vol. 3719, 1999
- [2] White, F.E., "A Model for Data Fusion", Proc. 1st National Symposium on Sensor Fusion, 1988
- [3] Bowman, C.L., "The Data Fusion Tree Paradigm and its Dual", Proc. 7th National Symposium on Sensor Fusion, 1994
- [4] Steinberg, A.N. and Bowman, C.L., "Development and Application of Data Fusion Engineering Guidelines", Proc. 10th national Symposium on Sensor Fusion, 1997
- [5] Llinas, J., et al. "Data Fusion System Engineering Guidelines", Technical Report 96-11/4, USAF Space Warfare Center (SWC) Talon-Command project report, Vol.2 1997
- [6] Bowman, C.L. and Murphy, M.S., "Description of the VERAC N-Source Tracker/Correlator", Naval Res Lab. Report R-010-80, March 1980
- [7] Bowman, C.L. and Morefield, C.L., "Multi-sensor Fusion of Target Attributes and Kinematics", 19th IEEE Conference on Decision and Control, December, 1980
- [8] Bowman, C.L., "The Role of Process Management in a Defensive Avionics Hierarchical Management Tree", Tri-service Data Fusion Symposium Proceedings, Johns Hopkins Univ., June 1993
- [9] Dasarathy, B., "Sensor Fusion Potential Exploitation-Innovative Architectures and Illustrative Applications", IEEE Proceedings, Vol. 85, No.1, 1997
- [10] Bedworth, M. and O'Brien, J. "The Omnibus Model: A New Model of Data Fusion?", IEEE Aerospace and Electronic Systems Magazine, Volume: 15 Issue: 4, April 2000
- [11] Salerno, J., "Information Fusion: A High-Level Architecture Overview", Proc. Intl Conf on Information Fusion (FUSION2002), Annapolis, MD, June 2002
- [12] Endsley, M.R., "Theoretical Underpinnings of Situation Awareness: A Critical review", in M.R. Endsley and D.J. Garland (Eds) Situation Awareness Analysis and measurement, L. Erlbaum Assoc. Publishers,
- [13] Endsley, M.R., "Toward a Theory of Situation Awareness in Dynamic Systems", Human Factors J., 37, 1, March 1995
- [14] Blasch, E.P. and Plano, S., "Level 5: User Refinement to aid the Fusion Process", in Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2003, B. Dasarathy (Ed.), Proc of the SPIE, Vol. 5099, 2003
- [15] Sea Power 21: Operational Concepts for a New Era (Washington, D.C., Department of the Navy, June 2002).
- [16] See article in Hansconian, news paper of USAF Hanscom AFB, Feb 21, 2003
- [17] Derganc, M., "Joint Interoperability and Integration and Joint Battle Management Command and Control Overview", briefing to USJFCOM, April 2003, <http://www.dtic.mil/ndia/2003interop/Derganc.pdf>
- [18] <http://www.ndu.edu/inss/press/nduphp.html>
- [19] <https://www.doctrine.usmc.mil/jv2020/jv2020.pdf>

- [20] <http://www.defenselink.mil/nii/NCW/>
- [21] <http://www.au.af.mil/au/aul/bibs/asm/asw.htm>
- [22] Gärdenfors, P., *Belief Revision*. Cambridge: Cambridge University Press, 1992.
- [23] A. Dragoni, *Belief Revision: From Theory To Practice*, in: *The Knowledge Engineering Review*, V. 12 n° 2, Cambridge University Press, 1997.
- [24] Benferhat, S., D. Dubois, Prade, H., *Kalman-like Filtering in a Possibilistic Setting*, in the Proc. of ECIE-2000, Berlin, 2000.
- [25] Delmotte, F., Dubois, L., and Borne, P., "Context-Dependent Trust in Data Fusion Within the Possibility Theory", in Proc of the IEEE Intl Conf on Systems, Man, and Cybernetics, Vol. 1, 1996.
- [26] D. Dubois and H. Prade, *Combination of fuzzy information in the framework of possibility theory*, in: Abidi M.A., Gonzalez R.C (eds.), *Data Fusion in Robotics and Machine intelligence*, Academic Press, Inc, 1992, 481-505.
- [27] Waltz, Edward L., "Information Understanding: Integrating Data Fusion and Data Mining Processes", Proc. of IEEE International Symposium on Circuits and Systems, Monterey CA, May 31-June 4, 1997. For a more detailed description of the integration, see, Waltz, Edward, *Knowledge Management in the Intelligence Enterprise*, Norwood MA: Artech, 2003, Chapter 8.
- [28] Desrocher D., C. Tschan, H Koons, and C. Bowman, "Data Exploitation Techniques for Enhanced Satellite Operations", AIAA 2001-4802 and 2002-1811, SatMax 2002, Arlington, VA, April, 2002
- [29] Clementine ® 7.0 Users Guide, Chicago, IL:SPSS Inc. 2002. See Chapter 12 for a discussion of the process for publishing data mining models to the data fusion process.
- [30] Bowman, M., Lopez, A., and Tecuci, G., "Ontology Development for Military Applications", Proceedings of the Thirty-ninth Annual ACM Southeast Conference, Athens, GA, ACM Press: New York, NY, 2001.
- [31] Guarino, N. and Giaretta, P., "Ontologies and Knowledge Bases: Towards a Terminological Clarification", in N. Mars (ed.) *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing 1995*. IOS Press, Amsterdam, 1995.
- [32] Rasmussen, J., Pejtersen, A.N., and Goodstein, L.P., "Cognitive Systems Engineering", Wiley-Interscience Publishers, 1994.
- [33] Smith, B. and Grenon, P., "The Cornucopia of Formal-Ontological Relations", forthcoming in *Dailiectica*.
- [34] Grime, S., "Communications in Decentralised Sensing Architectures", PhD Thesis, Oxford University, 1992.
- [35] D. J. Patel, R. Batta and R. Nagi, "Clustering Sensors in Wireless Ad hoc Networks Using a Dynamic Expected Coverage Location Model." , submitted to MORS Journal, 2003.
- [36] Durrant-Whyte, H.F., M. Stevens and E. Nettleton, "Data Fusion in Decentralised Sensing Networks", In Proc. Fusion 2001, Montreal Canada, July 2001.
- [37] Chong, C.Y., et al, "Distributed Multitarget Multisensor Tracking", Chapter 8 in *Multitarget Multisensor Tracking: Advanced Applications*, Y. Bar-Shalom (Ed.), Artech House, Norwood, MA., 1990.
- [38] Julier, S.J. and Uhlmann, J.K., "Non-divergent estimation algorithm in the presence of unknown correlations", In Proceedings of the American Control Conference, volume 4, pages 2369–2373, Piscataway, NJ, USA, 1997.
- [39] E. Drummond, "Track and Tracklet Fusion Filtering," *Signal and Data Processing of Small Targets 2002*, Proc. SPIE Vol. 4728, pp. 176-195, 2002.
- [40] Llinas, J., "Developing a Cost-Effective Approach to Information Fusion Research and Development", Keynote Speech, Information, Decision, and Control Conference, Adelaide, Australia, 1999.
- [41] Steinberg, A.N., Bowman, C.L., and White, F.E., "Rethinking the JDL Data Fusion Model", to appear in *NSSDF Conference Proceedings, JHAPL*, June, 2004.