

Challenging Old Assumptions in Global Information Management

Presented at Sixth
Annual ONR/CADRC
Decision Support
Workshop, Quantico,
VA, Sept 8-9, 2004

Dr. H. Van Dyke Parunak
Chief Scientist
Altarum Institute
3520 Green Court, Suite 300
Ann Arbor, MI 48105-1579
van.parunak@altarum.org

INTRODUCTION

Many fundamental assumptions in information management are driven by the nature of problems in the business world, and by the kinds of technology that have been available. The distinctive nature of combat, and new technical developments, invalidate some of these assumptions. This paper discusses three of these assumptions:

- That users need access to everything;
- That the Global Information Grid (GIG) needs a global semantic standard;
- That information is passive (and only people are active).

It explains why each assumption is invalid, and outlines emerging technologies that suggest new directions for addressing the needs that these assumptions identify.

ASSUMPTION 1: UNIVERSAL ACCESS

The vision behind the “common operational picture” (COP) is that information technology will enable every user to have access to any piece of desired information, and that this information will be consistent across all users.

There is no question that many users *want* access to all information, at least in principle. However, actually delivering such access may be technically intractable and psychologically undesirable, and new research is pointing the way toward mechanisms that will enable information systems to select the best information to send to each user.

Why is Universal Access Inachievable?

The vision of making all information available to all users faces two fundamental limits, one psychological and the other technical.

Psychologically, there is a limit to the amount of information that a human being can process. This limit is illustrated every time someone searches the World Wide Web using a search engine such as Google. These searches routinely return tens of thousands of results, but most users consult only the first few links on the first page.

This anecdotal experience is borne out by two lines of research. The cognitive limitations of the human organism were highlighted nearly fifty years ago in George Miller’s classic

study, “The Magic Number Seven, Plus or Minus Two” [3]. Reviewing a broad range of psychological studies, he found a sharp decrease in performance when people were asked to manipulate more than about seven items of information concurrently, reflecting the intrinsic limits of human attention. More recently, studies in artificial life [5] show that as an agent’s knowledge of its environment grows, its performance first increases, then decreases as its sensors become overloaded. Both results suggest that an important function of an information system is striking a balance between the amount of information available and the capacity of the human to process it.

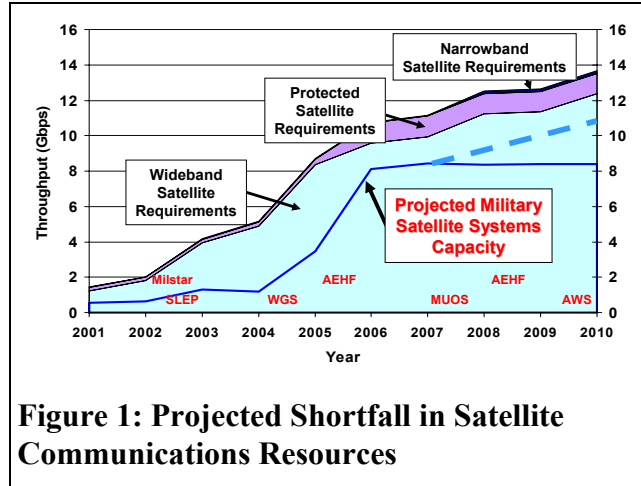


Figure 1: Projected Shortfall in Satellite Communications Resources

There is also a technical limitation to global information access. It is sometimes assumed that steady technical progress in telecommunications will eliminate any constraint on bandwidth, and in the abstract, great gains are being made in our ability to move information around. Practically, though, it is unlikely that this technology will be available to operators in the field in a reasonable time frame. Figure 1 summarizes a recent study by the Office of Net Assessment [1] on the availability of commercial and planned military satellite communications resources that could be used in case of military emergency, compared with projected needs. Commercial providers have moved away from satellites and toward fiber for their backbones. While satellites could support military communications in areas far removed from those they were originally intended to support, new fiber is tied to specific regions, and is not available to support conflicts in (say), the empty quarter of Iraq.

Thus, fundamental limitations in both human and technical capacity make it unwise to plan an information system around the assumption that every user can in principle have access to all information.

What is the Alternative to Universal Access?

If we cannot provide all information to every user, we must be selective, filtering the available information to match the user’s needs. Two basic approaches are available: deterministic and stochastic filtering (Figure 2).

Deterministic filtering uses a formal logical analysis (often based on symbolic artificial intelligence) to match the user’s interests against

Deterministic	Stochastic
“NP-Hard”: as size of problem grows beyond toy examples, the processing approaches the age of the universe (or longer!)	“Weighted Monte-Carlo”: Select randomly, Guided by self-organizing mechanisms inspired by natural insect colonies (“stigmergy”)
Time to solution grows exponentially with size 	Quality of solution improves exponentially with time

Figure 2: Two Approaches to Filtering Information

available information and select the information that is most likely to be relevant. This approach is attractive because the user can understand the logical rules used to select the relevant information, and thus gain confidence that the system will deliver the required material. Unfortunately, as the volume of available information grows, this confidence can be disappointed, because of the computational complexity of the processing involved.

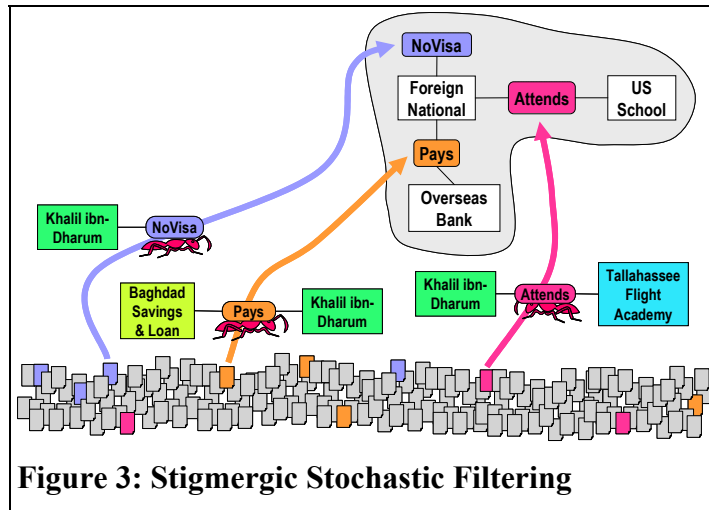


Figure 3: Stigmergic Stochastic Filtering

It is intuitive that as the amount of information that needs to be reviewed increases, the time necessary to review it will increase as well. If the processing time increases as a polynomial of the size of the input, it is reasonable to rely on more powerful computers, or larger number of computers, to meet the challenges of expanding information. However, deterministic logical methods, especially those that confront realistic representations of semantic structure, tend to be NP-hard. That is, the length of time required to execute the logic increases exponentially in the length of the input, and problems that are larger than toy examples will not be able to complete in a reasonable length of time even on the fastest computers we can imagine.

Deterministic filtering is intractable, but *stochastic filtering* is not. In its simplest form, stochastic filtering simply means that we select randomly from the available information. Purely random selection is unlikely to provide any information that is relevant to the user's interests, but it is possible to weight the selection in such a way that the retrieved information is more likely to be relevant to the user's interests. We have been developing mechanisms for stochastic filtering, inspired by the mechanisms that insects use to sort their nests and coordinate their actions [6]. These mechanisms, known collectively as "stigmergy," use the environment in which insects live as a locally indexed communication mechanism [4]. In this approach (Figure 3), digital ants *swarm* over a massive collection of documents, *recognize* fragments of a concept map that represents the user's interests, and *self-organize* to populate the map with documents relevant to its underlying assumptions. Stigmergic mechanisms have several desirable features.

- For many such processes, the quality of the solution (for example, the number of relevant documents retrieved) increases exponentially over time. That is, initially the number of documents grows very rapidly, providing an initial basis for the user's decisions. If the user has longer to wait, the process will continue to yield improved results, although at a slower rate.
- The process can easily be distributed over many machines, without the need for central coordination.
- The process continues to operate even in the face of dynamic change (for instance, shifts in the user's interests or in the body of information available), without the need to be restarted.

ASSUMPTION 2: GLOBAL SEMANTIC STANDARD

One of the greatest contributions of the Internet has been to lower the walls separating different bodies of information. Information that twenty years ago required a trip to the library can now be accessed by a few keystrokes from one's desk. Many different information producers have eagerly made their resources available, in a move toward a single global information grid, or GIG.

Unfortunately, we have learned that exchanging the *form* of information is much simpler than exchanging its *meaning*. Any body of information rests on an assumed way of conceptualizing the world, called an "ontology." For example, people who talk about automobiles agree in advance that a car consists of a body, a power system, wheels and suspension, and an interior, and that options for power systems include gasoline, diesel, natural gas, electric, and hybrid, while veterinarians concerned with horses focus on legs and hooves, hair and skin, and internal organs.

Experience has shown that there is no single "right" ontology for describing the world. Different information producers tend to have different ontologies, making it difficult to combine their information in a single application. A traditional way to deal with this divergence is to attempt to outlaw it, by developing a single ontology that all producers agree to use. This approach is impractical. Fortunately, new technical tools are making it possible to combine information from systems with globally incompatible ontologies.

Why is a Global Semantic Standard Impractical?

There are three obstacles to the vision of a single global semantic standard: the existence of established inconsistent ontologies, the complexities of translation, and the fact of continuous, rapid change in local contexts.

First, there are many established communities of practice, each with its established view of the world that makes its own activities efficient and internally coherent. As valuable as a global standard may be for people who work across different communities, it will impose severe costs on the existing ontologies that it displaces, costs that are often high enough to dissuade the users of existing systems from abandoning them.

One might hope that an automatic mechanism could be found to translate automatically among different ontologies, thus permitting people to use their existing systems while keeping them aligned to the global standard. The second obstacle to a global semantic standard is that the problem of aligning ontologies belongs to the class of NP-hard problems described in the previous section [2]. That is, for ontologies of realistic size, a program to align them with one another would not be able to deliver results in reasonable time.

The third obstacle is that ontologies are not static. They are constantly evolving to support the needs of the communities that use them. In a dynamic environment (such as the US military in a time of transformation), there will be strong pressures on local groups to specialize or refine the portions of the ontology that they use the most. If a global ontology outlaws such specializations, it will severely limit the productivity of its users; if it does not, it will soon become irrelevant.

The fundamental issue is that thought is dynamic, constantly exploring new combinations and relationships. Any attempt to codify the infrastructure of thought runs the risk of limiting creativity and mental productivity.

What is the Alternative to a Global Semantic Standard?

It is a fact of ontological life that the overall semantic structures underlying different systems will differ from one another. Instead of trying to impose a global standard, an alternative approach recognizes that most processes that cross systems require agreement only in a local region of the ontology. Figure 4 illustrates this graphically. Even though the two structures diverge from one another, they are aligned in a local region, and as long as the interactions between them affect only this region, the more remote differences have little impact.

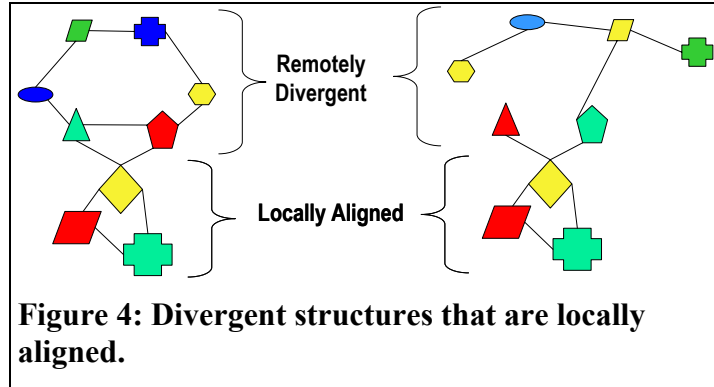


Figure 4: Divergent structures that are locally aligned.

Altarum’s Living Ontologies technology supports this alternative approach by enabling knowledge workers (for example, system modelers) to integrate different ontologies in their activities. The technology supports several processes that can be interleaved with one another.

- Modelers can *search* existing ontologies for concepts and substructures that are candidates for concepts that they need to represent.
- They can *compare* these candidates with one another and with new constructs to assess the degree of consistency among them, thus promoting agreement where possible without imposing a single global standard.
- They can then *construct* new systems by combining the candidate structures they have retrieved and evaluated, all the while monitoring the growing model for its consistency with existing structures.

ASSUMPTION 3: PASSIVE INFORMATION

Before the computer era, information took the form of ink on paper. It was completely passive, and depended on people to file it, retrieve it, transport it, interpret it, and act on it.

In many ways, computer systems have simply automated this view of information. The database or web server has replaced the filing

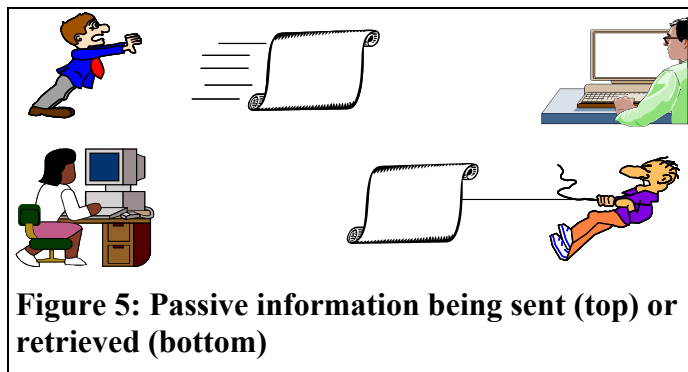
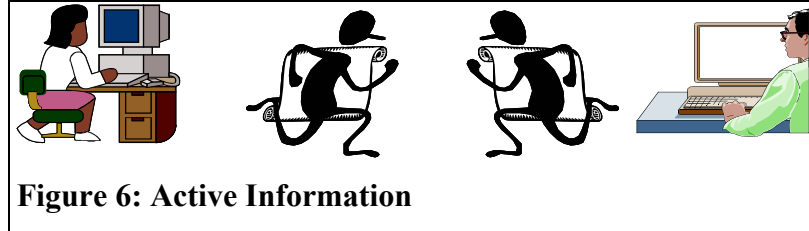


Figure 5: Passive information being sent (top) or retrieved (bottom)

cabinet or library, and email has taken the place of physical couriers, but people still must make the decisions and take the actions to manipulate knowledge. In the email



model (Figure 5, top), the current holder of information must know who would find it useful and send it there. In the web model (Figure 5, bottom), someone who needs information must know where it is, and retrieve it. In both cases, information is passive, and all decisions about it must be made by people.

Why is Passive Information Undesirable?

Passive information was the only option in the days of ink on paper. Today, the line between data and process is extremely fuzzy. Both consist of a machine-readable list of ones and zeroes, and there is no intrinsic reason why one (the process) should be active while the other (the information) is passive. We can imagine an architecture in which each packet of information is self-conscious about its contents and is actively seeking users who might find it valuable (Figure 6). There are two reasons to pursue such a vision of active information.

First, as we have already noted, the cognitive burden on users is increasing as the amount of information available continues to grow. Anything we can do to reduce this cognitive burden will make information systems more effective.

Second, the passive information model (Figure 5) forces a rigid and often unrealistic distinction between the *producers* of information, who need to identify possible users to whom they can send reports, and the *consumers* of information, who need to locate likely sources to which they can send queries. This distinction is often inappropriate. Most people who are seeking information, do so in order to support their own information production tasks.

What is the Alternative to Passive Information?

Altarum has developed an alternative to passive information known as PARTNER (Population of Agents for Real-Time Networking with Emergent Routing). A PARTNER system has three species of agents.

1. *Domain agents* produce and consumer information. They may be people, on-line sensors, legacy databases, or similar entities.
2. *Message agents* are packets of information that circulate in the network. A message agent is neither a query nor a report, but a packet containing some known details and a description of some unknown details that could naturally be related to the known details. Its mission in life is to find other message agents that complement its own knowledge, and bring them to the attention of domain agents whom it knows and who might be interested in information related to its own contents.
3. *Network agents* are the computers that populate the network. They provide three services to the message agents. Like routers in any communications system, they

move them around and store them when they are not moving. In addition, they provide processing resources that message agents can use to compare themselves with one another.

PARTNER agents perceive, react to, and reinforce two fields in their environment, simultaneously coordinating their activities and controlling the use of resources (Table 1).

A *semantic* field imitates the use of pheromones in insect societies, and guides mutually relevant messages toward each other through the network. An *economic* field, modeled on naturally occurring markets, helps messages use network resources responsibly as they travel.

The semantic field is based on a *signature* that can be associated with any agent, and used to assess its similarity with other agents. The data in a signature can be either explicit (the role and ID of a domain agent; metadata associated with a message agent; the location, capacity, and connectivity of a network agent) or derived (a keyword or concept vector, the texture of an image). These examples can all be precomputed, but signatures can also contain dynamic information developed during the lifetime of a message agent, such as profiles of what users have found a given item interesting.

A message's sender initializes its behavior in these two fields, assigning its contents (thus initializing its signature and determining its semantic behavior) and its budget (determining its economic behavior). Network agents maintain the two fields based on the messages they handle. Message agents move through the network, leaving traces of their signature on the nodes through which they pass, and following such traces to find other similar messages. When two message agents with similar signatures meet, they compare themselves, and if they are indeed similar, they notify the domain agents who originated them. If the match produces useful information, the domain agents pay the message agents a reward, increasing their budgets. At the same time, every action that a message agent takes (moving from one network agent to another, occupying storage, comparing itself with other message agents) decrements its budget. Message agents whose budget reaches zero are removed from the system, thus preventing congestion.

Such an information ecosystem has several useful global behaviors.

- The lifetime of information in the system is not fixed by policy, but emerges dynamically, based on the initial budget and the subsequent rewards that domain agents pay to each message agent. Information that proves to be valuable will persist, while information that is not used will be removed (or archived to a server that itself can serve as a domain agent).
- The semantic field will lead to message agents naturally congregating at certain network agents based on their signatures. Such concentrations will strengthen the semantic field around those network agents, enabling other message agents to find them more easily. Thus the system naturally self-organizes to provide efficient access to related information.

Table 1: Two fields drive PARTNER.

Enabling Flow Field:	Semantic	Economic
Challenge:	Effectiveness	Efficiency
Manages message ...	Relevance	Cost
PARTNER Mechanism:	Digital Pheromones	Budgets

- The flow of currency through the system is a natural way to monitor the value contributed by different components. For example, network agents that accumulate more usage fees than others could use these fees to supplement their physical resources, thus guiding infrastructure investments based on actual usage patterns.
- Perhaps most important, the system eases the cognitive load on human domain agents. Information producers do not need to know who could use their products, nor do consumers need to know whom to ask. Users simply introduce message agents into the system, and the system will tell them when those packets of information have encountered other packets that might be useful to them. Instead of a synchronous process of query and response driven by human action, the system learns the interests of its users and proactively helps them find both information and other users that can support their activities.

SUMMARY

Current information management systems are constrained by assumptions developed in the days of stand-alone computer systems, or even pre-computer paper and ink repositories. These systems cannot accommodate the dynamic requirements of today's military. Fortunately, new computing technologies, such as stochastic filtering, living ontologies, and active information, permit us to challenge the old assumptions and develop new systems that offer a revolutionary advance in effectiveness and efficiency. The Altarum Institute has demonstrated many of these technologies in its research programs, and is actively pursuing their deployment in real-world applications.

REFERENCES

- [1] D. Anhalt. The Changing Nature of Commercial Satcom and its Impact on US Military Advantage. *Satellite 2001*, Office of Net Assessment, Washington, DC, 2001.
- [2] A. Elmagarmid, M. Rusinkiewicz, and A. Sheth, Editors. *Management of Heterogeneous and Autonomous Database Systems*. San Francisco, CA, Morgan Kaufman, 1999.
- [3] G. A. Miller. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *The Psychological Review*, 63:81-97, 1956. Available at <http://www.well.com/user/smalin/miller.html>.
- [4] H. V. D. Parunak. 'Go to the Ant': Engineering Principles from Natural Agent Systems. *Annals of Operations Research*, 75:69-101, 1997. Available at <http://www.altarum.net/~vparunak/gotoant.pdf>.
- [5] T. Takashina and S. Watanabe. The Locality of Information Gathering in Multi-agent Systems. In *Proceedings of Second International Conference on Multi-Agent Systems (ICMAS-96)*, pages 461, 1996. Available at http://homepage.mac.com/t_takashina/paper-dir/ICMAS96_takashina.pdf.
- [6] P. Weinstein, H. V. D. Parunak, P. Chiusano, and S. Brueckner. Agents Swarming in Semantic Spaces to Corroborate Hypotheses. In *Proceedings of AAMAS 2004*, pages 1488-1489, 2004. Available at <http://www.altarum.net/~vparunak/AAMAS04AntCAFE.pdf>.