Analysis and the Knowledge Gap
Drowning in Data

I am unashamedly pro-analyst, and as such any reform effort aimed at the intelligence community that doesn't focus on empowering our most precious resource doesn't interest me. Analysts are drowning in data, a flood of it. But, what do we do? We ask them to answer horrendously complex 21st century questions with Korean War era analytical tools. Why then are we surprised when their products fall short? Part of the problem is our fault. Managers told them that their number one customer was the President of the United States . Guess what? They took us at our word. Everyone writes for the President's Daily Brief (PDB). Unfortunately, in our zeal to please the "man" we seem to have forgotten that the quality of current reporting the President receives depends on the existence of a vibrant, dynamic basic research program. Of course, you can produce the PDB everyday without one, that is what we essentially do now, it just won't be very good, especially when it comes to military and technical subjects. If we want to dig our way out of this hole, there is only one way to do it. We must do a much better job of turning information into knowledge.

Creating New Knowledge Is Not Easy

For as long as I can remember, I have complained without success to anyone who would listen my fear that current reporting techniques were crowding out all other types of analysis. Now it has, and I still can't get anybody to care. My powers of persuasion obviously weren't up to the task. The "wise man" approach favored by current analysts has spread like kudzu relentlessly overgrowing everything in its path. Researchers don't do research any more. Estimators don't estimate. Analysts mostly absorb information through osmosis from reading and studying the avalanche of data being collected, compare their findings against conventional wisdom, and pass on their conclusions in written reports or oral briefings.

Maybe my real gripe with this sort of conjuring is nothing more than envy. It's certainly not one of my strong points. I find working out simple problems in my head a challenge. For me counting is still a fingers and toes exercise. I can remember things in short bursts okay, but throw a bunch of variables into the mix, or add other sorts of complexity, and I usually end up with a mind melt down. I also don't have a gift for producing elegant prose, especially when I'm working under tight deadlines, or in those instances where the telling is

"The 'wise man' approach favored by current analysts has spread like kudzu relentlessly overgrowing everything in its path. Researchers don't do research any more. Estimators don't estimate."

as important as the discovery itself. Many of my colleagues do have such a gift, and frankly, I hate them for it. Unless my findings speak for themselves, I'm in big trouble. So yes, my criticism and concerns about the dominance of current reporting techniques may contain an element or two of jealousy.

Or again, maybe it is partly a carry over from my early days as an analyst at DIA, where my first mentor, Eva Watkins, impressed on me the value of research and record keeping. She possessed an uncanny ability to identify and locate military formations by tracking mentions of the PLA's military unit cover designators (MUCDs) appearing in the Chinese media, and others she found buried in classified reports. Not very glamorous work, but a necessary step for doing any serious work on the status and capability of China's military forces. If the Chinese had known how good she was, I doubt they would have bothered with the hassle and expense of frequently changing the numbers in an attempt to deceive us.

I find fewer and fewer like Eva working in the intelligence community. Plenty of excellent analysts remain; they just aren't doing much basic research anymore. Everyone seems focused on current reporting. We were at that odd place where even as our collectors produce substantially greater amounts of information, we seem to use less and less of it. Tons pile up in the archives. The bigger the piles become the harder it is to find what we need. A few diehards still try to do research, but get little or no support from management. R&D and investment dollars have largely been spent on collection not analysis. Many of my colleagues just gave up. Some left government. Some retired. Others joined the ranks of current reporters. What went unnoticed amongst all these changes, at least at first, was that with the demise of basic research the substantive quality of virtually all types of analysis, including current reporting, suffered. Supporting evidence for military and technical questions, such as Iraq WMD, disappeared. Answers became three parts witchcraft and only one part analytical tradecraft.

The situation has gone on so long that current members of the intelligence community sometimes have difficulty understanding my complaints. Over twenty years ago regional geographic offices replaced CIA's Office of Strategic Research (OSR), my old organization, and its counterparts, the Office of Political Research (OPR), the Office of Economic Research (OER), and a few others. Not long after most basic research began to die out. Today, it is not unusual to find senior managers in the community who not only have not done research themselves, but also don't know what it is, and certainly couldn't

"When a unit moved or the PLA decided as a security precaution to change its cover designators, something it did regularly, Eva sprang into action to perform her magic. She tracked changes in a dog-eared notebook looking for patterns and keys to the new MUCD assignments. All it took was a few new numbers, and Eva had cracked the code."

teach others how to do it. Accordingly, much of our analytical work is a mile wide but only an inch deep. Smart people tell us what they think, but their conclusions, even if highly accurate, come with little in the way of evidence either hard or circumstantial. Data everywhere, but creating new knowledge remains problematic.

Finding the Formula for Generating More Knowledge

The intelligence community is not the only one suffering from this condition. Others in government, the military, and business are experiencing the same problem: We all are drowning in data. It comes in torrents from a wide range of sources, much of it in machine format, but with little or no thought on how it might be used by others, particularly mind workers such as analysts. All of us are struggling to turn the river of information into knowledge. So far, no one in the intelligence community, the field I know best, has had much success.

Many of the IC's early attempts at generating more knowledge centered on acquiring larger amounts of information. Made good sense. Collecting more data should increase the chances of finding the answers you seek. They also apparently believed it reduced the need for basic research. Unfortunately, it has not worked out exactly that way. In some respects creating new knowledge in the IC has gotten harder, not easier. Our recent experience in Iraq, for example provides a graphic lesson of what can happen when you have a lot of information but too little knowledge.

Creating New Knowledge Is Not Easy
"... much of our analytical work is a mile wide but only an inch deep. Smart people tell us what they think, but their conclusions, even if highly accurate, come with little in the way of evidence either hard or circumstantial. Data everywhere, but creating new knowledge is problematic."

It's hard to argue that more collection won't make you smarter. I certainly won't. Every good analyst I know always wants more data. In my experience, however, the quality of the end product invariably depends more on the analyst than on the amount of data available for the study. Fact is, those in the intelligence community who favor more and better collection have held sway for many years now. Their effort has paid huge dividends. Because of their efforts the U.S. today possesses collection capabilities, both technical and human, the envy of the world. Because of their foresight American analysts are blessed with an unsurpassed abundance of rich data to work with, more than at any other time in our history. Despite great people, good intentions, and tons of information we still couldn't accurately assess the status of Iraq's weapons of mass destruction program.

How Can We Close the Knowledge Gap?

Recognizing that the IC is performing below its capability, of course, does little to solve the information-knowledge problem. To do that, I think we should take a more detailed look at analysis, and the role that data plays in knowledge creation. You would think that much work has already been done on this subject. All we need do now is review and restate the conventional view. Unfortunately, we don't yet have a taxonomy of analysis that we can all agree on for our starting point. For now I will proceed with my own conceptional scheme leaving the debate of its accuracy and value until a later time.

In my schema analysis comes in different forms, but exists solely to support the needs of policy makers. Sometimes their requirements can be met by simply passing on reports from the collectors, or adding a bit of area or subject knowledge when needed. Providing news of fast breaking developments and other information accurately is an especially important function. Such written and oral reports constitute the vast majority of what policy makers see and hear. Not surprisingly, such reporting can generate a host of questions most of them complex and not easy to answer. Indeed, policy makers seem to value the answers to these follow up questions and requirements more highly than the news, but without ever fully understanding the analytical challenges this poses.

Directed research of this sort, unlike current reporting, starts with a question not a new piece of information. True a current analyst doesn't just pass on new data, and must search for other pieces to provide context and hopefully find information that adds greater explanatory details to what is known, but in a pure sense is not creating new knowledge. In large part this is a function of the data itself. Only a tiny fraction of the material we collect, what I call the "obvious" bits of information, initially ends up being captured by the current intelligence analysts – enough to support basic current reporting requirements, but not much else. The readily observable segments -- those you immediately recognize as important and relevant -- probably add up to something less than 5% of what we collect.

After skimming off the "big" pieces of data that leaves considerable knowledge undiscovered in the haystack. Finding it isn't easy. Some pieces, the "less obvious" bits, are missed because of the sheer volume of information collected daily. More fundamentally, much of the potentially useful data segments are so fragmented that their value is not easily discerned by the human mind. Context and association with other data segments are necessary components for determining the relevance of these "hidden" bits of information. Unless something triggers immediate recognition by the analyst, any potential knowl-

edge the "less obvious" data may contain is lost, and finding the "hidden" pieces without a rigorous disciplined search is all but impossible.

Directed research can uncover some of the less obvious and hidden pieces of data, but not nearly enough. The foundation for current reporting and directed research is built on the products of "basic research" requires a more in-depth examination of the materials being collected by research analysts. Basic researchers focus on micro issues and details often of little interest to policy makers such as orders of battles and technical manuals, but provides the new knowledge essential for producing useful analytical products. Basic research also determines the quality of the work current reporters and analysts focused on answering the questions of policy makers, what I call directed research, produce. Directed research unlike current reporting starts off with a question not an event. Such a study might examine the pace of developments in Iran's nuclear weapons program today, for example, compared with the pace observed in past years. The raw materials found in a study of this sort -- less obvious and hidden pieces of intelligence data -- are the product of directed research and set it apart from current reporting. Directed research also depends on the existence of a body of basic research performed by others.

This first step in the analytical process -- adding human understanding to data flow -- is at the heart of what analysts do. Their job is to filter the information collected for possible relevance, extract those points they consider important, and load them into some form of a repository; usually one that allows them to view the data from a variety of perspectives. Surprisingly, this process has hardly changed from the procedures my colleagues and I practiced forty years ago.

21st Century Analysts Wear Old, Hand Me Down Clothes

Despite a much larger volume of data, and hundreds of millions of dollars spent on equipment and technology, analysts still manage most data by hand. They manually scan for meaningful nuggets, and store those they find, usually in nothing more than a digital or paper "shoe box." Anything more elaborate or mechanized is the rare exception. Indeed, loading their filtered data into a format machines can manipulate is so laborious and unwieldy that most quickly go back to their old "shoe box" approach, even when this means greatly limiting the depth and breadth of their inquires.

Reporting the news headlines takes you only so far. Answering the complicated questions requires an analyst to dig much deeper. Good analysts find insights or trends hidden in the archives, and make com-

parisons in the data over time. Most times, much of what might prove useful is not easily found. It lies hidden in the vast stores of information stacked up in each agency's data bases. New innovative techniques, such as data mining help a little, but such tools as currently exist, experience the same difficulty in manipulating unstructured data as we humans do, and produce even more pieces of information for the analyst to process. Doing it by hand takes too long. Accordingly, although we should, we often don't attempt such research. Analysts concentrate their time and effort instead on current reports, such as the PDB, or write other more lengthy products but ones that lend themselves to current reporting techniques.

That's fine, if the questions you need answered don't involve military or scientific and technical issues. Making judgments on Iran's nuclear program or the pace of China's military modernization, however, are not possible absent a vigorous research effort. Pretending that we can makes matters worse. Consumers mistakenly assume that the answers they receive, especially those on matters of national security, are based on in-depth and sophisticated analytical studies. The IC can report the news. Anything beyond that, however, is problematic. It need not be that way, but it will require changes in the way analysts do business, and the support they receive from their managers.

Tackling the information-knowledge problem has proven anything but easy. Most would agree that the situation in the IC is getting worse, not better. It is not clear to me why we haven't equipped our analysts with more technology. Billions have been spent on collecting and storing data, but little on transforming it into knowledge. None of the major existing databases, for example, are analyst friendly, primarily because they were not designed with analysis in mind. Most originally started out as structured data for use by a collecting agency, or was formatted for storage by IT offices. Not surprisingly, each database differs considerably depending on the requirements of the organization creating the data, e.g. NSA, NGA, CIA/DO. Works great for them but is totally unsuited for analysis. In most cases, the structure of the existing data repositories actually impede the work of analysts. So much so, that for all practical purposes analysts don't use much of the data.

I sometimes think people must secretly equate giving our analysts technology with cheating -- a chess master during a match communicating secretly with a panel of experts, or crooks using hidden cameras and wireless devices in a gambling casino. When it comes to better understanding problems such as North Korea's nukes or Iraq's

Only a tiny fraction of the data collected, what I call the "obvious" bits of information, initially ends up being captured by the current intelligence analysts – enough to support basic current reporting requirements, but not much else. The readily observable segments they use -- those you immediately recognize as important and relevant -- probably add up to something less than 5% of the total volume of information being collected.
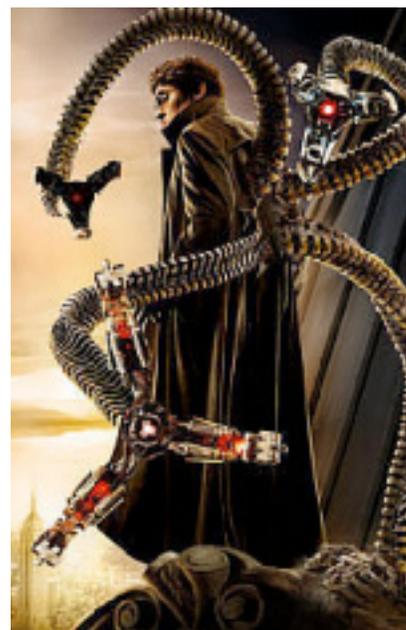
sectarian violence, however, I hold that cheating is more than fair. It is essential. We should arm our analysts with every technical trick available today, and then go back into the lab to build more. Outsmarting terrorists and other adversaries is not a bad thing.

Empowering Analysts Technologically

Current intelligence is the easiest place to start. Dramatic increases in the quantity and quality -- as much as a doubling -- of the current reporting information available to our analysts can be accomplished quickly and easily. No need to collect more information or build new systems with greater capability just yet. We are choking on raw data. Before we collect more stuff it makes sense for us to look for better ways to use what we already have. Squeezing more knowledge from the materials we collect is not rocket science. If analysts capture more pieces of the puzzle, they can make better guesses. The technology to do a better job exists today. We have simply chosen not to use it. Arming our analysts with the best tools available has never been one of our priorities. That has to change.

As a first step this means supplementing the existing data distribution scheme. (See the Distribution Diagram at the sidebar). Analysts in this example would continue receiving raw traffic as they do currently, but they would also be provided additional materials from a rule based filter. Subject experts -- our best and brightest -- would compile a list of best practices and lessons learned to incorporate into the base filter. Additionally, analysts would be allowed to make their own changes, additions, and other modifications to these initial rules on the fly. The filter could be a stand alone module tailored to suit the individual analytical offices it serves, but would be most effective if its results were shared collaboratively on line with all authorized users. In either event, the system must allow for adding analyst derived data into the system, and appending comments on other data segments obtained from the filter. The filtering mechanism for all source analysis must also accept archived data in any format -- text, graphics, and other digital materials. That is what the all source approach is all about.

The best filtering system in the world doesn't do much good unless the new materials can be exploited effectively by the analysts. The more data you have, even though it may have been partly processed, the harder it is to get your arms and mind around it. An analytical toolbox must accompany our attempts to squeeze more knowledge from the information collected. At a bare minimum, it should allow analysts to manipulate the filtered data segments by rule, time, loca-

tion, keyword, etc., and the results obtained displayed in time intervals of an hour, day, week, month, etc. The ability to do trend analysis on the results is needed, but the complexity of building such a tool, should not hold up the introduction of the toolbox concept. As with the filtering mechanism, the toolbox should come with best practice queries pre-packaged. Once analysts become more proficient in using the tools, the system should anticipate that the users will generate additions, modifications, and changes of their own. Ideally, the system, like the rule based filter, would allow for collaboration and sharing of these new tools among analysts. Likewise, the system must also provide a capability for the analyst to append selected queries, notes, and manipulation results to data segments.

I can hear the screams of protest from intelligence community managers already. They will claim I'm out of touch. Hundreds of millions of dollars have been devoted to analytical tools and other analysts centered activities over the last several years they would argue -- "Carl you just don't understand." Maybe I don't. Maybe I haven't seen all the on-going work. Maybe the absence of such tools aren't at the heart of our analytical problems. What I do know, however, is that the quality of our current reporting today is not worth any where near the $40 billion we pay for it. We can do better. Empowering our analysts requires more than buying them things. Adopting new analytical software or training techniques is not enough. If it were that easy, someone would have found solutions for our problems already. Fact is little has changed. Indeed, ignoring the needs of analysts has persisted for as long as I can remember. Neither as an analyst nor as an intelligence community manager can I remember anyone ever asking for my input on how to improve the quality of the analytical effort. Sometimes collection, but never analysis. What's more, I'm not aware that of any of my colleagues have ever been consulted either. This doesn't mean that it doesn't happen, I'm sure it must, but it does suggest that it is not a common place occurrence. It's as though people believed that everything was fine, that we had already mastered the art and science of intelligence analysis. There was nothing more to learn. Of course, our track record in recent years argues otherwise.

Fixing basic research will not be nearly as easy as improving current intelligence. Most analysts know how to do current reporting. Give them more materials to work with and they will do a better job. Not so in the case of research. So little is being done today most of the intelligence community will have to start over from scratch. Small pockets of excellence exist, but far less than is needed to support a credible research program. With few qualified to train others and a

Distribution Diagram
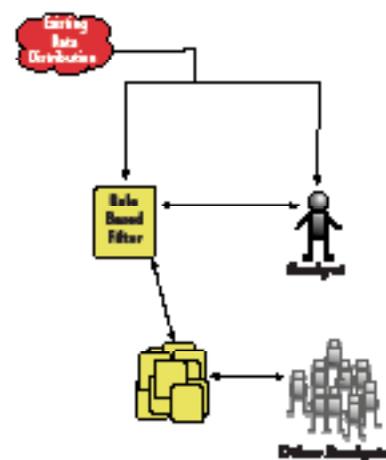
Analytical Toolbox
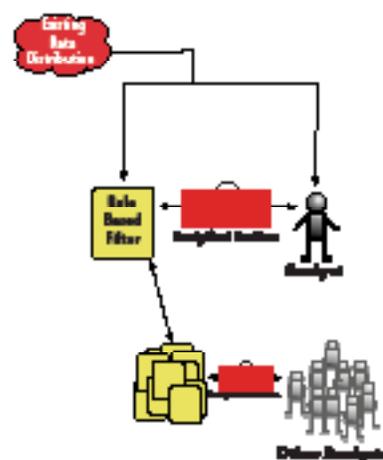
senior management packed with current intelligence analysts, progress in building new capabilities will undoubtedly be slow. But, be wary of quick fixes. A too ambitious reform effort or doing nothing are both dangerous options. That leaves us with a gradual, incremental strategy as our wisest choice.

Since I don't know of any existing research tools, I can only speculate what they might look like. I start from the premise that at a minimum, if we can do a project by hand, tools can be designed to do it faster and more efficiently with machines. This step alone would expand the type and complexity of analytical questions we could tackle enormously. In addition, although I don't know for certain, I suspect that we will also find that such tools will allow us to tackle problems that thus far have been beyond our reach. This is not to say that the technical hurdles to build these new tools are trivial. But with sufficient effort and resources building such tools should be possible.

We never have really applied technology to research in the intelligence community like it has in many other intellectual fields. Except for a few economic studies at CIA, and some of my OSR colleagues during the Cold War, who analyzed Soviet force movements using a super computer, we've never really applied technology to basic intelligence research, and nothing recently. Unlike current intelligence, where some first steps have been taken, research has received no attention, and instead in most agencies has largely disappeared. This means we will have to relearn the fundamentals of research along with applying twenty-first century techniques to our work.

I also suspect that one size fits all tools won't work. More likely the question being answered will dictate a tool's characteristics not our current practice of limiting the problems we tackle to what technology we have at hand. Classes of tools that can be modified for individual projects may be possible, but I don't envision someone building a data mining-like tool for basic research problems. Virtually every question a researcher encounters will likely differ somewhat from the last. Each will have their own set of technical requirements. Applying existing tools to new problems will work sometimes, but more times than not a new question will require us to make modifications to the tools we have on hand or build new ones. It also means we will likely have to assign technologists directly to analytical units rather than keeping them in general support.

But, I don't think building a research capability requires that we hire more people or increase the intelligence budget. Our current effort is so wasteful, that some improvements can be achieved by a redistribu-

"Squeezing more knowledge from the materials we collect is not rocket science. If analysts capture more pieces of the puzzle, they can make better guesses. The technology to do a better job exists today. We have simply chosen not to use it. Arming our analysts with the best tools available has never been one of our priorities. That has to change."

tion of the existing analytical resources. No need to cut collection expenditures, or cut corners in our other reform efforts. The simple fact is too many people are doing current reporting. The amount of duplication in the intelligence community is staggering. It simply does not take this many people to do so little. Most days a handful of analysts can produce more current reports than our consumers can digest in a week. And, it's not as if the other analysts were off doing something different. They aren't. They are all doing the same thing. The number of analysts, however, far exceed the number of reportable events. Indeed, we could send two thirds of our analysts to the dark side of the moon and only family and close friends would notice that they were gone. Historically, the intelligence community produced a comparable amount of reports with a third the number of analysts. We can do it again.

<div style="color:red">**"I start from the premise that at a minimum, if we can do a project by hand, tools can be designed to do it faster and more efficiently with machines."**</div>

Final Thoughts

Let me end with a caution. The challenges facing the intelligence community do not lend themselves to a strategy of playing it by ear. We need to clearly understand what brought us to this point and devise a blueprint for how we are going to fix it. In this regard, I think we could learn much from Dr. Richard F. Elmore's work on educational reform, especially an article he wrote in 2004 entitled "Change and Improvement in Education Reform." He observed that:

> "The premise that educators know what to do and all they need are the correct incentives to do it is essential wrong. Some educators know what to do; most don't. Some are able to learn what to do on their own; most are not .... Educators, like most practitioners, learn most of what they know from what they do. It should not surprise us that the best predictor of what they will do at Time 2, other things being equal, is what they were doing at Time 1 .... educational reform can become a kind of conspiracy of ignorance: policy makers mandating results they do not themselves know how to achieve, and educators pretending they do know what to do but revealing through their actions that they don't"

I quote Dr. Elmore at length because his observations ring true for our current, and past, reform efforts in the intelligence community. In reaction to our poor showing on WMD in Iraq, both the Executive Branch and the Congress have rightfully demanded changes in how we do business. They mandated a significant change in the structure of the community including the downgrading of the role of the Director of the Central Intelligence Agency whose community responsibilities as DCI have been given to the newly established Director of

National Intelligence (DNI). Intelligence professionals have responded with promises to do much better in the future. We have even gone as far as asking that policy makers hold their criticisms so that we can give full attention to our important responsibilities for reform and change.

While both policy makers and practitioners recognize the need for change, neither talks much about what we need to improve collection and analysis. Both apparently believe that the bureaucracy that brought us to this point knows what's wrong with intelligence, and more importantly has a plan to fix it. I think they are wrong on both counts, but I am especially skeptical that we have a detailed plan for making the needed improvements. For many of us this is not the first time a commission, or some other group, has demanded change in the wake of an intelligence failure. Despite their best intentions and excellent suggestions each new reform effort failed to make much of a dent in the community's problems. Reminiscent of Yogi's comment about it being "deja vu all over again", the current effort has already lost much of its urgency and steam. The reality is that change is easy, but improving our capabilities is hard, especially when you don't have a detailed game plan for fixing what is broken. If such a plan exists, I would like to see it. Policy makers and practitioners should debate and argue over its details. So far, this wave of reform unfortunately appears to have more in common with our past efforts rather than a brave new start.

"The reality is that change is easy, but improving our capabilites is hard, especially when you don't have a detailed game plan for fixing what is broken."