Israel Intelligence Community Commemoration and Heritage Center
Institute for the research of the methodology of intelligence

INTELLIGENCE IN THEORY AND IN PRACTICE

a journal on intelligence methodology

Issue No. 3, OCTOBER 2018

BIG DATA AND INTELLIGENCE

INTELLIGENCE IN THEORY AND IN PRACTICE

a journal on intelligence methodology

Issue No. 3, OCTOBER 2018

3

The journal Intelligence – in Theory and in Practice is intended to be a forum for a conversation on intelligence methodology for members of the intelligence community (present and past) and supporters of intelligence in Israel and the world.

Graphic design: Ze'ev Eldar
Publisher: Effi Melzer Ltd.
Translator: David Hornik
Printing: Moshe Asayage Ltd.

Israel Intelligence Community Commemoration and Heritage Center
Institute for the research of the methodology of intelligence

Address for sending responses and articles:
dudi.st591@gmail.com

This issue is dedicated to the memory of Effi Melzer, who published the previous issues.

Effi Melzer Research Publishing Ltd.

TABLE OF CONTENTS

Foreword

In the era of the information explosion and a teeming, dynamic arena, the world of big data and cyber poses increasingly complex challenges to the intelligence community. Pathbreaking strategic thinking, professionalism, technological innovation, and smart and precise technical management of intelligence data are critical to the security of Israel and to its ability to defeat its enemies. Nowadays this imperative guides the activity of the Israeli Security Agency, which views the information technologies as a resource that is vital to the work of collection and prevention when confronting foes in the different arenas.

In recent years the agency has taken important measures to adapt its technological capabilities to the new needs. Such needs range from storage capacities to meeting complex challenges that require the ability to use textual, visual, or vocal data automatically, and, of course, to identify and extract, from an ocean of data, what is relevant, precise, and can lead us to the right path in carrying out our role.

The main characteristic of a preventive intelligence entity is the need, on the one hand, to concentrate on routine collection processes that are wide and deep, and, on the other, to ensure a very rapid response to signs of the enemy's intention to cause damage to the country and its citizens. The Israeli Security Agency, both by nature and in its methods, is oriented to preventive activity and to all the processes of collection, processing, and research. These, in turn, are based on supportive technologies that facilitate exhausting the huge quantities of incoming data to make accurate assessments.

As we look ahead, our enemies are hardly treading water, the world of big data is developing and expanding, and the technologies are getting more and more sophisticated by the minute. What this means for us is that the challenges and the threats will only continue to intensify. Thanks to an organizational culture that promotes technological innovation, to high-quality and devoted human capital, to cooperation with the intelligence community, and to large-scale investment in advanced technologies, the Israeli Security Agency and the intelligence community are now at the technological and operational forefront in Israel and the world. These are also the keys to our future achievements and to our ability to guarantee the wellbeing and security of the state of Israel.

In light of the centrality of the subject and its importance for the intelligence organizations, I am confident that the readers of this issue will gain a far-reaching and in-depth perspective on the challenges of the hour. I wish you enjoyable reading, and I wish all of us success.

Nadav Argaman, head of the Israeli Security Agency (Shin Bet)

Preface

We are pleased to offer you, our readers, the third issue of the journal Intelligence - in Theory and in Practice. It is the fruit of many efforts of the authors, who agreed to invest their time and energy in order to enhance the community's knowledge in the big-data field. The big-data phenomenon poses a range of new challenges for the intelligence community - some of them in the form of dangers, most in the form of opportunities. All agree, however, that it is an important growth engine for the intelligence community, to which the community needs to be properly oriented.

The concept of big data reflects a reality in which the quantities of data are so large and so varied that drawing insights from them cannot be done by focusing separately on each data bit. Instead a new approach is required that combines different aspects and gleans insights automatically from the relevant information. On the one hand, the phenomenon is generated by the huge growth in the quantities of data, primarily due to the expanded use of the internet and the augmented reach of the social networks, which together have completely altered human life and, particularly, the world of information, and continue to impel revolutions in this domain at a rapid pace. On the other, the technological quantum leaps enable the storage of the immense data in a way that allows its rapid retrieval and the extraction of insights from it that are not evident at first glance.

The challenges that this phenomenon poses are not unique to intelligence, of course. Indeed, civilian bodies in the business sector are at the forefront of finding ways to cope with big data, which involve efficient storage and the creation of tools for very rapidly drawing insights from it, while updating it in real time. This issue of the journal explores, among other things, the unique characteristics of the intelligence endeavor and the ways in which the intelligence community needs to learn from civilian entities that, for business purposes, deal with a much larger range of information.

A few key questions arise from this issue of the journal. One concerns the extent of the changes needed in how we manage the intelligence endeavor amid the transition to the big-data world. Is a "revolution in intelligence affairs" required, or is this a second-order change that calls for adaptation but does not challenge the basic concepts and architecture by which the intelligence community is built, the attributes of the intelligence profession in general, or the nature of the relations between the intelligence community and the civilian environment? The answer that emerges from this issue is that big data indeed requires a wide-scale revolution and that the first signs of this revolution are already evident in the field.

Another question concerns whether the big-data bases are indeed available to

researchers in general and to intelligence researchers in particular, or whether special expertise is required to delve into and use these bases. Here it will emerge that the greater the awareness of the need to maintain cyber security, the more difficult it will be to delve into the bases.

A further question is that of resources. Can intelligence organizations in a small country like Israel mobilize the resources needed to meet the big-data challenges, and what is the pattern of relations between the intelligence community and the giant companies that lead this field in the world? Clearly the solution that the American intelligence community has chosen - opening a special cloud computing service together with Amazon, for which it has paid $600 million - is not available to the Israeli intelligence community. What can and should be the Israeli equivalent?

And finally, ethical questions arise that Edward Snowden brought to public awareness, and that require the intelligence community to maintain the balance between its preventive work on the one hand and individual rights, particularly the right to privacy, on the other.

The present issue highlights the uniqueness of the intelligence community's approach to the big-data phenomenon and offers different perspectives on its implications for intelligence. Some of these perspectives come from within the community, others from people in other fields who deal with the big-data domain, and in the business world are at the forefront in doing so.

The overall impression that arises is that, despite the great occupation with the issue and the adaptations carried out by the different intelligence organizations in recent years, there is still a need for a community strategy to address the complex and revolutionary challenge, which has far-reaching implications both in terms of approach and of resources.

The issue is divided into parts according to subject area, with each part illuminating the big-data phenomenon from a different angle.

The introduction is by Col. Y., who serves in Military Intelligence (Aman). His article presents the results of an inquiry, which he led, into the changes that the digital era entails for the intelligence enterprise. Among the findings of this inquiry are the need for a digital transformation of intelligence organizations that deal with knowledge development and the need to "change the diskette" of intelligence personnel who have been trained to compartmentalize and now must adopt an approach of sharing as the default option.

The first part deals with the basic elements of the big-data phenomenon and the implications for intelligence. It is opened by Lieut. Col. Ts. of Aman, who surveys the technological trends in the new information world and, in the context of their ramifications for intelligence, points to the new professions that are needed to deal

with the concomitant opportunities, such as "information scientists" and "information civilians". The article emphasizes the need for a dramatic change in the volume of resources to be invested in the future in the intelligence organizations that deal with information.

In an interview we conducted with Dr. Yoelle Maarek, Amazon's international vice-president for research, she describes the ways in which the global information companies meet the challenges that the big-data phenomenon poses, some of which can indeed offer inspiration to the intelligence communities while others do not suit them because of these communities' hermetic nature and relatively small size. Among other things, she points to the need for research personnel to be involved in devising the system that filters the sea of data so that it does not turn into a "trash heap".

Dr. Haim Assa, who is an architect of systems based on artificial intelligence, describes the challenges of retrieval and fusion of data in the present era and how machines can be taught in a way that improves their intelligence capabilities. Part 1 ends with an article by Eran Baron, chief technology officer (CTO) at the Infinidat data storage company, who presents the characteristics of data storage in the civilian world and addresses the question of what intelligence needs to learn from the civilian world in this context - and, in particular, what it should not learn.

The second part focuses on social networks. Its first article is by Maj. A., who sets forth the principles of social network analysis (SNA) along with its advantages, among them the ability to analyze millions of data elements in a short time as a network. M., deputy director of the intelligence school of the Israeli Security Agency, then describes the attributes of the new networked world. Using the metaphor of Berlin divided by a wall, he demonstrates the potential possessed by new, social-media-based, intelligence analytical devices for preventive-intelligence purposes. Maj. D. of Aman highlights the importance of understanding the limitations of the social networks as a source of intelligence. Certain structural distortions stem from their users' unwillingness to share information, and some users also share false information.

The third part examines the current state of the research field from several standpoints. Dr. Shai Hershkovitz, a former Aman official and currently an expert in intelligence theory who also deals with the use of new technologies for intelligence purposes, discusses the idea of utilizing the wisdom of the masses for the benefit of intelligence research. He also proposes the establishment of a new intelligence discipline to be called CROSINT (crowdsourced intelligence). The third part continues with two articles - one by the researchers Dr. Carmit Valensi of the Institute for National Security Studies and Dr. Keren Sasson of the Davis Institute, the other by Maj. A. of Aman - that explore how to make use of new tools, such as capabilities

for the statistics of words and for the analysis of databases, to improve intelligence research.

The fourth part discusses the conceptual and structural implications of the current era. The opening article is by O., a civilian in Aman, who considers the significance of the data infrastructure for the intelligence endeavor as well as the possibilities that stem from it (including semantic retrieval, task-specific organization, etc).. Against this backdrop, he proposes substantial changes in intelligence activity. Brig. Gen. Itai Brun, former head of the Research Division of Aman, examines three different research approaches - the educational, the systemic, and the scientific, presents each one's advantages as well as its orientation to information and to big data, and recommends focusing on the scientific approach. The article that closes this part is by D. P., until recently a senior director in the Israeli Security Agency. He discusses the need for jointness in the field of intelligence force buildup, emphasizes the importance of a concept as a basis for technological and operational development in intelligence organizations, and points to the processes that are needed to implement and develop this organizational idea, some of which require a joint mobilization extending from the leadership of the organization to the operational units.

Finally, we look at developments in the academic world. Dr. Avner Barnea, a former senior official in the Israeli Security Agency and current research fellow at the Center for National Security Studies at the University of Haifa, takes a look at the preventive-intelligence endeavor and how it has changed in the face of the big-data phenomenon.

Acknowledgments

First, it is a pleasant duty for us to thank all of the authors. Our special gratitude goes to Nadav Argaman, head of the Israeli Security Agency, who wrote the foreword for the issue, and to Lieut. Col. Ts. of Aman, who patiently helped us delve into this complex subject and also wrote one of the articles in the issue. We are grateful as well to the head of Aman and the head of the Mossad, who are encouraging their personnel to write for this journal.

Thanks are also due to the security officials - in Aman and in the Prime Minister's Office - who approved the publication of the articles.

Finally, we wish to thank the leadership of the Israeli Intelligence Heritage and Commemoration Center - the chairman, Brig. Gen. (res). Dr. Tzvi Shtauber; the director-general, Brig. Gen. (res). Dudu Tzur; the head of the Intelligence and Terrorism Information center, Col. (res). Reuven Erlich; and the deputy director-general, Hanan Mazor. Their support and dedication make the publication of this journal possible.

Heartfelt gratitude to the Boksenbaum Neta Fund for its support for the institute.

Thanks and great appreciation are also due to the language editor, Nili Gerber, the graphic designer, Zeev Eldar, and the translator, David Hornik. This is also an opportunity to once again express our deep sorrow at the sudden passing of Effi Melzer, the publisher of the issue. Drawing on his broad knowledge and long-standing experience, Effi helped us greatly and spared no effort to promote and give shape to this journal.

We wish you, the readers, enjoyable and valuable reading and remind you that we wait to hear from you - whether proposals for articles or responses to this issue.

Yossi Kuperwasser and David Siman-Tov
Glilot, October 2018

The Journey of Clarifying the Concept and Use of Intelligence and Operational Superiority in the Digital Era

Col. Y. - serves in Military Intelligence (Aman)

In Essence

If one were to summarize the concept as briefly as possible, one would say: to apply the potential of the digital era to the systemic challenges that intelligence now faces. Alternatively: a different concept of intelligence-operational superiority, based on the understanding that the information explosion and the ability to strive to know "everything about everyone" makes possible an updated and different intelligence and operational response.

Over the past several decades (since the military intelligence bodies were instituted in the modern state), the way in which the different intelligence bodies have been built has been intended to enable responses to complex questions about enemies and rivals. The way to give a response was to create specific accessibility to relevant (and intimate) places, attain important data bits, and use them to piece together the "intelligence puzzle". The aim of all this activity is to provide the intelligence picture (a clarification of reality); to point, on its basis, to possible future scenarios; and to offer recommendations about the ways in which this reality could be affected.[1] The dream of every intelligence worker was to be a "fly on the wall" in the office of the intelligence object. The search for the most relevant and critical nodes gave birth to vital-information marking, the notion of requisite items, and to the collection gaps. Over the years, time after time, we were compelled t which this reality could be affected o prioritize, decide what was most important and urgent, and act accordingly.

The cyber era made us realize that the whole world is interconnected. This means that, at least theoretically, one can go anywhere. By utilizing the cyber dimension, one can provide a response to any vital information, **even the most complex.**

We defined the intelligence endeavor in the Information Age as a different approach to intelligence (and operational) superiority, because the Google company's

1 There are, of course, other ways in which proper intelligence activity has been described over the years. Here, though, we have sufficed with a relatively well-accepted description. In any case, the change that we will describe below, in the context of the new approach of the Information Age, is different from every other way in which intelligence has been described over the years. The issue of pointing to vital information is found in all the different approaches.

approach to the information explosion is different from that of its predecessors. As we were getting to work, a friend showed me a joke that was circulating on the network: "Where is the best place to hide a corpse? On page 2 of Google, where it will never be found". This joke captures the truth of the matter quite simply. As intelligence personnel, we grew accustomed over the years to reading hundreds and thousands of items so as to find a piece of a puzzle in one of them and try to connect it to another puzzle piece in another item, and thus try to construct the full picture. In actuality, most of the items do not necessarily contribute in any particular way to understanding the whole. Hence the task of the research officer is to separate the wheat from the chaff and find the important data item within the mountains of items.

Google has a different approach. When I ask Google a specific question that interests me (and have not "just" woken up in the morning and started reading through the column of news that arrived during the night), I am not prepared to go to page 2 of Google and read the title of one of the answers that appear there. I expect of Google that the answer will appear on the first page and in one of the first titles. If I don't get an answer I do not despair, and certainly do not say to myself, "OK, Google doesn't know". I say to myself, "Google knows everything but my question was not asked well enough". Never has Google asked me to prioritize one population group over another; it presumes, under the aegis of the Information Age, to know everything about everyone (even if that means about seven billion people…).

In our view, this, in a nutshell, is the different approach of intelligence in the Information Age. Superiority does not stem from one or another information item but from the information explosion itself and the ability to ask about whatever interests me. When the information is truly infinite, then clearly one cannot aim to read all the items and there is no need to reset the columns. The approach is different; one can and must wander among the infinite information. According to this approach, the answers are already to be found in the existing information; one just has to know how to wander through it optimally and ask the questions that interest the intelligence worker.

Keeping in mind the limitations of the metaphor, we can liken the situation to the global breakthroughs that occurred around the cracking of the Enigma. Even before that time there were geniuses who deciphered codes, but the British and the Americans knew that breaking the Enigma would require 20,000 people who worked for 20 years and invented the supercomputer, which changed the course of human history. This is our era when it comes to information, its use, and its significance. Hence the intelligence researcher continues to be very relevant (the relevance and importance have even grown), but the conversation between man and machine is changing.

Meanwhile the attitude toward the specialists' ownership of the information is also changing. For years intelligence bodies avoided transferring information to operational entities out of concern that they would make improper or irresponsible use of it. Within the intelligence community, too, the collection units did not transfer most of the raw information to the research units - supposedly for reasons of compartmentalization but also for practical and doctrinal reasons. According to the approach used in the past, if one provides access to raw visual information about a complex decoding to everyone, they may make improper use of it and jeopardize sources. Making SIGINT information available to researchers was viewed the same way. In the Information Age, however, there is no ownership of information; it belongs to everyone.

For comparison's sake, no one avoids making infinite medical information available on the internet out of fear that I may make improper use of it if, God forbid, I need it for treatment of one of my children. The information belongs to everyone, and everyone can ask whatever he wants and decide whether and when to turn to an expert or make decisions by himself.

The Boundaries of the Discussion

This article is based on intensive work being done by the IDF's intelligence corps in recent years in response to the opportunity and the obligation to change and to be changed in tandem with the dizzying global development occurring in the digital era. Because the information explosion is infinite, its impact on the intelligence approach is extensive. This article does not presume to give a full portrayal of intelligence in the digital era. It is derived mainly from a number of experiences, and we think there is a basis here for a broader conversation about the role of intelligence in the digital era and about the ways to realize the ever-more-ambitious visions.

Crises as Opportunities and as Goads to Transformation

In his book The Structure of Scientific Revolutions, Thomas Kuhn maintains that the great revolutions of humanity grew from crises. He similarly describes the phenomenon of paradigm shift: a prior paradigm gradually fills up with holes like a sieve, becoming less relevant by the day (but continuing to exist), while alternative paradigms develop beside it until the previous paradigm is (almost) expunged and new paradigms remain. Kuhn views the interim period as a "crisis period" that usually emerges as such only in retrospect. At the moment, intelligence appears to be in a period of transition or of a crisis of that kind.

A large part of the intelligence tasks still receive an effective and relevant response in the traditional approaches. A large part of the tremendous achievements of Israeli

intelligence in recent years stemmed from locating the vital bits of information, leading to the gaining of intimate access that enabled impressive intelligence and operational breakthroughs. But there are also other examples where, in the midst of the Information Age, we had to adopt a different approach to operational intelligence superiority.

One of the major examples is the phenomenon of "terror by inspiration", which takes the form of lone-wolf terror. When the potential terrorist (who sometimes did not know himself, even a day before the attack, that he was going to be the attacker) got up one morning, decided to take a weapon in the form of the family vehicle or the knife from the kitchen, and went out to perpetrate an attack, traditional intelligence stood helpless. How could this be warned of beforehand? What place or node could be sought as part of the response? Who can be prioritized as an essential piece of information to be monitored in lieu of someone else?

And indeed, time after time, we have found ourselves without a relevant and adequate response to lone-wolf terror. The crisis was so severe that we found ourselves going from funeral to funeral, from terror attack to terror attack, while each time saying to ourselves in retrospect, "Wait a minute, there was some sort of indication here that, if we'd paid attention to it, maybe the attack could have been prevented". As time went on and the intelligence entities continued to be irrelevant, the sense of crisis intensified; we realized that intelligence superiority was indeed under challenge. We understood that, when it came to terror by inspiration perpetrated by lone wolves, the way we had done intelligence work over the past decades was insufficient.

...we found ourselves going from funeral to funeral, from terror attack to terror attack, while each time saying to ourselves in retrospect, "Wait a minute, there was some sort of indication here that, if we'd paid attention to it, maybe the attack could have been prevented"

There are other significant challenges to which traditional intelligence has had trouble providing a relevant response in recent years. An example is the terror tunnels, and there are others. When one considers such challenges in terms of the age of information and intelligence, other, more relevant directions emerge.

Hence, when it comes to the digital transformation, there is no need to fear clarifying the situation and pointing to the current crises; without a crisis there will be no revolution. Until we formulate and define for ourselves (each organization separately, thus also each suborganization or unit, or even department or section) what crisis experiences our organization has failed to provide a response to, because

it still has not conducted internal processes of adopting practices and capabilities that accord with the Information Age, the organization or unit will not be able to carry out such an important and challenging task.

A digital transformation of organizations - learning from the processes that organizations have undergone in recent years. Over the past several years many intelligence organizations have tried to undergo a "digital transformation". Many have failed. This article will look briefly at some main lessons that, in our view, offer a key explanation for their failure.

One cannot build a second and third floor without a first floor, foundations, and sewage facilities. The first step - the most complicated and important, indeed the secret of success - involves the "sewage system". Until we deal with information, order it, and organize the relevant-information endeavor, we cannot begin the task. The aim is not to appoint someone who will manage the information and be responsible for preserving it; these are infinite quantities of data that must be put in the right places. The different databases must be connected, must be fused, it must be decided in which server which type of material will be preserved, and so on. This is a task that begins with mapping (what exists in the world, in equivalent organizations, and among us in our own organization) and continues with organizing the relevant server farm and crafting the organization's information strategy (which type of information is preserved where, in what configuration, and so on). As noted, the importance of this step cannot be exaggerated. If one can invest in only one issue in the Information Age, it should be the sewage system of the information.

Another way to convey the message is to note that in the data-science era, the data is even more important, complex, and challenging than the science.

The Theory of Continuity

In the Information Age, the capabilities to create the necessary manipulations and the intelligence investigation are based on continuity - continuity between the kinds of material, the kinds of information, and the different intelligence entities. Data science is neither a "branch" of SIGINT nor an extension of VISINT nor an updated way to do research; it too requires and is conditional on continuity. Furthermore, according to this approach, it may be that even when there is 80% continuity in the information, it will be worth 0% of the information, at least regarding some of the numerous issues. This is because the real power lies in the ability to ask, and to clarify the continuity as a whole.

For example, we have to structure intelligence so that we can ask questions in the following mode: a person who spoke to someone in Iran and lived 100 meters from a mosque was observed, during the past hour, driving north and last night did not sleep

at home. This question pertains to various kinds of material, and what is needed is the ability to manage and clarify investigations of this kind.

Developing the Knowledge and Building Up the Intelligence Force within the Context

"Context" is the magic word. The secret of success (or nonsuccess) lies in the ability to build up power in the relevant context. Much of the intelligence entities' failures in trying to build themselves up for the Information Age stemmed from the attempt to develop the knowledge and the questions machine **in a way that was generic and lacked a context**. Some also tried to take an external capability from industry and integrate it into the intelligence entity, and failed.[2] Recent years have taught us that, particularly in the era of the information explosion, one must build up the intelligence force, the "machines", and the information facilities so as to fit a particular context.

Hence it is necessary to "break" the dichotomous division between force-buildup bodies and force-activating bodies. Each time one builds an updated big-data capability in a particular context, it is possible and necessary to learn to expand it to other fields of activity, but the building must be done within a context. One instance of this was the establishment of departments in the Google style within different units. Although these departments did not try to replace the force-buildup bodies, it was only thanks to them that the giants of force buildup obtained a context.

Another concern is the need of every technological body to build the product entities for itself. One cannot expect that an intelligence body will describe a system and a technological body will build it. The technological bodies need a function that will manage the relationship between them and the force-activating bodies, as well as the force-buildup bodies that are within the force-activating bodies; this will enable the creation of quality products.

The Importance of the Question

Throughout the years the Jewish tradition encouraged the right, the responsibility, and the duty to ask. Four questions have been with us since the Passover Haggadah. Intelligence, too, focuses on the responsibility to ask the right question, which makes it possible to pinpoint a precise, vital bit of information and

2 We do not claim that one should not work with industry (it is in fact obligatory), but it is not possible to take something from industry and cut-and-paste it into the organization. Moreover, industry cannot define the organization's information strategy, information architecture, and so on. So while capabilities from industry must be utilized, it must be done in accordance with the information task that the specific organization is conducting.

clarify a complex reality - providing a major compass for intelligence activity.[3] According to the traditional approach, any successful question can lead to a relevant accessibility and thereby expose the secrets of the other side. If it is a question for which there is no answer in the information, deep insights can enable us to identify the adversary's logic, thoughts, and so on. In the era of the information explosion, however, one must assume that there is no question that for which there is no answer in the information. One needs to know how to probe the data for the right item, to construct questions that can contend with the information load, and to understand that when an answer is not obtained we must assume that we have asked the wrong question. (If the "sewage system" of the information has been properly organized, the main thing we have to do is improve our ability to ask relevant questions).

In addition, one can relate to a question that was asked by someone as a potential feature. If I have probed the information about a certain matter and the information was understood in a specific way, this can reveal something new about the information. Each month Aman alone produces hundreds of thousands of relevant features, and if we know how to study them we can upgrade the power of the machine to impressive capabilities of machine learning.

In Praise of Ontology

Ontology was born long before the digital era. It is the theory of "what is", of existence, and it poses questions about the different entities that exist in the world and the network of connections between them. Ontology deals with things that are shared, with what causes them to exist, and with what causes them to be connected. In the Information Age, clarifying, defining, and conceptualizing ontology is of huge importance. The information includes different entities that have a network of connections between them. A main precondition, a key to success, is to define the ontology of the information as a basis for guiding the organization's digital transformation.

For the defense establishment, the basic ontology concerns the network of relations and links between the person and the place. The entities we deal with are people and places that have a set of links between them. In the world of targets, for example, one seeks first and foremost the most precise location for marking, and after that whoever is at that location (whether there are civilians there, who is hiding there, etc).. When it comes to preventing terror, as in other domains of intelligence, the point of departure is often the person (who may want to perpetrate a terror attack), and after that we will try to find out where he is. Ontology, of course, involves additional and complex

3 For example, Aman chief Herzi Halevi, in his opening conversation for the post in 2014, reiterated that the condition for quality intelligence is the ability to point to the relevant and optimal question.

levels of relations and links between the entities (different objects, additional people, other places, fields of activity, etc).. There are basic and permanent strata of ontology and there is also dynamic ontology. Without a precise definition of the ontology, one cannot craft a strategy and an architecture for information.

The World Belongs to Those Who Share

People have always sought to connect information with information. For example, the margins of the pages of the Gemara and the Mishnah offer references that connect with other places in the text. The link constitutes knowledge in itself. The internet era has developed and refined this approach. The ability to wander among the pages of the internet and connect between one information item and another has enhanced the understanding that the "links" constitute knowledge in itself. In the digital era, collaboration in information, in knowledge, and in operative activity based on data constitutes a condition for success. The deeper the collaboration, the greater the ability to provide a better intelligence response. When it comes to information and knowledge, the challenge of collaboration is, unfortunately, not simple. Many people and organizations believe they must safeguard the information they have solely for their own or the organization's benefit. In the digital era, the world belongs to those who manage to overcome the challenge of collaboration; those who share with others, and whom others share with. One must invest in the ability to share and be shared with. This is first and foremost a systemic-conceptual challenge, but it also entails a complexity of architecture that enables the sharing of information. The consent of fellow organizations is required, and so is an architecture that makes it possible to implement the agreements regarding information in a simple fashion. Both aspects are, however, complex and daunting.

The Power of the Internet

Over the years we got used to the fact that the big secrets were to be found in isolated offices. In the cyber era as well, the prevailing notion is that the major secrets are harbored in the internal networks. Accordingly, in most cases, the more intimate, internal, and difficult to obtain the information is, the higher its classification and usually its relevance as well.

For us in the Information Age, the greatest challenge is not the ability to create intimate accessibility of one kind or another but the ability to exhaust the relevant information within the infinite information in general and on the internet in particular. It appears that the different intelligence organizations still have not achieved this revolution and certainly have not internalized it. The greatest effort is still invested in creating additional and intimate accessibility, while we assert that whoever wins

the competition to exhaust the existing information on the internet on behalf of his own organization's intelligence challenges will be a step and a half ahead of the other organizations. One of the secrets of success in the coming years is precisely the ability to use the power of the internet to the benefit of the security intelligence entities.

The Smart Digital Space

One of the visionary and revolutionary possibilities opened by the revolution of intelligence in the digital era is the "smart space", which has to do with the connection between the place and the person (intentionally in that order). The Information Age has fostered a revolution in the ability to reach and live within a defined spatial cell. The term "smart space" corresponds with "smart home" and "smart city". An era in which one can imagine a fusion between visual information and networked information in a given spatial cell entails a potential for another kind of intelligence-guided warfare along borders, for dealing with a warning, and, in general, for providing a different response to operational needs in a geographic context.

The "smart space" begins with defining the space. The point of departure is the choice of the spatial cell we optimally want to reach (it could be across the border, a place where a suspicious activity is being conducted…). Within the specific spatial cell one must create an ability to fuse the different sensors in the context of the operational problem that has been defined. For this spatial cell it is also necessary to organize the "sewage system" of the information in the context of that same location, to organize task-specific sensors, and to attune an intelligence and operational entity to utilize the information of the "smart space" to achieve an improved and, particularly, more precise operational response.

Getting to Work

Intelligence work in the digital era is intense, exciting, challenging, and very complex. It requires examining basic assumptions, renouncing old paradigms, and adopting new ones. It opens doors that did not exist at all and enables dreaming of new things. There is something very special about taking new paths that no one has ever trod. Though the metaphor is inexact, it is like someone who embarks on a new route, sometimes gets wounded and even loses blood (because the route is unpaved and full of various kinds of thorns), and sometimes comes to a precipice, or a cul-de-sac, and has to go all the way back. But usually this involves exciting breakthroughs, the paving of new paths. Moreover, often the entry into the information world makes it possible, at least in the early stages, to pick low-hanging fruits. Suddenly one can contend with a great many questions that seemed to be suspended without a response.

The subsequent stages are usually more complicated.

We need to make two additional points. First, this endeavor, as mentioned, is thorny and challenging and requires a large quantity of resources. One cannot achieve a digital transformation for an organization by making a few minor adjustments. Required, instead, is an internal and massive adjustment of resources, not merely complying with external standards for the revolution. (When someone goes to a psychologist, the fact that he pays 400 shekels for an hour is of value; it requires him to assume part of the responsibility for the treatment's success). In this case as well, organizations must internalize the fact that an organization that wants to survive must make investments of its internal resources to succeed. As a rule of thumb, we believe that at least one-quarter of the organization's resources should be channeled into this transformation. That means precious resources - quality personnel, the time and attention of senior managers, and so on.

A further explanation is that a change of this sort affects the entire organization. Traditional research changes in the digital era, and so does the generation of targets and of deterrence. Indeed such change begins in a specific entity, spurs revolutions around it, and during the process the whole organization changes.

Organizational Structures That Support a Faster Digital Transformation in Intelligence Practice

"Don't talk to me about the 'what', talk to me about the 'how'".

That sentence is the reverse of how we grew up over the years. We think the question "How does one construct the intelligence entities so as to realize the vision of the Information Age?" is of critical importance. In many regards the "how" is more important and more consequential than the "what". Furthermore, assimilating the concept of intelligence superiority in the digital era requires a turnabout and change of the second and third degree. It does not just involve setting up an "information section" or "Google department" or "data science field". The organization as a whole must undergo a transformation. The processing bodies change, the operations rooms assume a different form, and so on.

This section points to the relatively small transformation-driving entities, those that are built into the force-activating entities and have to correspond and operate together with the force-buildup entities. In this context we focus on three main outputs:

Information and the professionalization of information. Each body must have a specific domain whose activity is information. This entity will connect between the information tasks. Such a domain will be responsible for continuously organizing the

information in the relevant context, bringing more information into the organization, and connecting with and being connected with the databases of other organizations. Could it be that in the future we will not need the information domain in each body because all of it will be part of the larger information endeavor? To me it seems that we will always need such domains. In any case, in the coming years each organization that wants to survive and to succeed in its digital transformation will have to build for itself a specific domain that deals with the "sewage system" of the information, with building a foundation floor for the information, and with liaison with the information tasks.

Another trend that is developing, and for which the information field ought to assume responsibility, is the professionalization of information. Every basic intelligence division (section, small department, etc). needs to have an information intelligence officer of its own. This person will have expertise in the different databases and serve as the organization's information expert.

The use of information (mining and manipulating the information). The second entity is the one that specializes in the use of the information. While each intelligence worker must know how to probe the data by himself, also needed are experts who are able to investigate huge databases, specialize in constructing complex queries, and know how to develop the ability to pose complex queries. The information-mining intelligence officers need to have one foot among the technological specialists of data science and one foot among the intelligence researchers in all the entities for deep collaboration.

Sensors and specific sources. In the Information Age all sensors are an integral part of the information explosion. Seemingly there is no need for task-specific sensors; each sensor creates information items that are supposed to be part of the infinite information endeavor. Nevertheless, task-specific sensors have a relevance, particularly those that "dwell" in a location and are able to collect and compile information that is found there, information that, without those sensors, would be lost. In many regards the purpose of these sensors is the opposite of those designed to create accessibility; these sensors are designed to ensure that existing information does not go to waste.

Furthermore, specific sensors play a unique role in the transformation of the organization. A sensor gives a tangible sense of a new capability in a location, thus creating energy in connection to its establishment. Deploying specific sensors is of great importance when it comes to exporting the digital revolution. When sensors are deployed, one must insist on an enabling architecture. What this means is that all the sensors are connected to the information tasks (in real time) and fused with the other components of the information.

Four on four - a management method that suits the digital transformation of organizations. The "four on four" method that we have implemented appears to us a winning method. The size and complexity of the revolution also pose a challenge to the management approach that is employed. One of the concerns is that we will dream too big and be left only with dreams and presentations. On the other hand, there is an apprehension that, in order to be sufficiently practical, we will carry out measures that are too small and local and cannot enable a wide-scale revolution. Thus we arrived at the management method of four on four: four years, four months, four weeks, and four days.

It is hugely important, and takes courage, to dream that we have the ability to change the security concept and the way in which it is implemented. The Information Age entails a tremendous potential for massive revolutions. In the same breath we have to sum up what concrete steps we can and must take within four days; what processes of the information, its organization, composition, and so on, can be advanced in the next four weeks; and what measures we must and can advance in a time frame of four months.

One of the concerns is that we will dream too big and be left only with dream. On the other hand, there is an apprehension that, in order to be sufficiently practical, we will carry out measures that are too small and local and cannot enable a wide-scale revolution

Each year we reexamined all the stages: what we completed in the previous four days and could and had to do in the next four days; how we would progress on the tasks that were set for the next four weeks; what was the updated status of the outputs we had committed ourselves to complete in four weeks; as well as what new things we had learned about the updating of dreams for another four years. The pace of change in the digital era requires a management method of this kind. Moreover, in an era in which information multiplies itself every moment, the main way to learn and carry out revolutions is through friction. This method makes it possible and necessary to encounter friction all the way to the big revolution.

Furthermore, this method allows one to celebrate the small successes along the path, those that give the strength to continue the challenging work. One cannot tarry too long at the stage of forming one's dreams. One has to demonstrate operational outputs very quickly through intelligence and operational practice, a proven contribution. Thus already in the first stages there is a need to select significant issues that can be addressed differently with the capabilities developed as part of the

digital transformation; this can be advanced with task-specific intelligence research teams, with specific operations, while in any case ensuring that these are based on the information revolution and will mandate learning in the course of the operational friction. These successes will yield the energy (internal and external) to continue the challenging work.

Instead of a Conclusion – a View to the Future

This article has pointed to the revolution in the functioning of intelligence. This revolution is rooted in the broader, global revolution that we are living in - the digital era, the information revolution. We elaborated on why, in our view, this situation entails a different concept of intelligence and operational superiority. While the traditional methods have not vanished and apparently will continue to be relevant for many more years, the information explosion enables and requires a different kind of response as well. Such a response is not based on accessibility but on the data-science challenge, and particularly on utilizing and wandering among the information, on devising manipulations of it, and so on.

In certain regards we have tried to show that in the Information Age there is a new role for approaches in the style of the old kibbutzim - give what you can and get what you need. In recent years many have addressed the need to achieve intelligence superiority and to build an intelligence infrastructure that enables operational superiority (relative to our enemies, of course). It seems to us that in the Information Age, the quality and pace of the digital transformation are what will determine the advantage (or, God forbid, the opposite) between us and our adversaries and enemies, both in intelligence and operational terms.

It seems to us that in the Information Age, the quality and pace of the digital transformation are what will determine the advantage (or, God forbid, the opposite) between us and our adversaries and enemies, both in intelligence and operational terms

This article has not addressed two revolutions that are a product of the information revolution: in the perceptual and information-war domain, as well as the profound realization that in the digital era information is (also) the reality. Although the issue of perception has occupied people and affected information since the dawn of humanity, the digital era makes it possible, through information, not only to clarify reality but also to shape it.

On a personal note, the intelligence task in the digital era is the most fascinating, complex, and exciting endeavor I have ever known. It requires me to scrutinize basic

assumptions, to part from some of the assumptions I grew up on, to embrace new ones. When we succeeded, there were many instances where it directly helped save human lives. When we failed and were not precise or quick enough, the outcome was often funerals. We had commanders who demanded and enabled us to embark on the work, even when in many regards it was a journey into the unknown. Even in the difficult moments (and there were many of those), such commanders were mainly concerned that we should keep our heads up - not giving up on the stringent investigation, while in the same breath mustering strength and confidence and continuing the work. Intelligence in the digital era could not have developed and flowered without the commanders in Aman - an organization that encourages dreaming and enables implementation. It is a pleasant duty to thank the commanders, the colleagues, and the comrades without whom nothing would have been achieved.

The intelligence endeavor in the digital era was also a forging of new partnerships and friendships; of working with Units of Aman; with departments of the teleprocessing branch; with entities and individuals of the Israeli Security Agency, the Israel Police, and COGAT (the Coordinator of Government Activities in the Territories); with the Research and Development Agency of the Defense Ministry, and so on. Also taking part in the work were civilian companies and new friends I met along the way without whom nothing would have been achieved.

We have embarked. This is just the beginning of the journey, and in the terminology of Thomas Kuhn's aforementioned book The Structure of Scientific Revolutions, we are the transition generation, which entails confronting numerous and complex challenges. The competition between us and our enemies is a competition of learning. It seems to me that in the coming years, the competition of learning will take place primarily in the digital domain.

Intelligence Derivatives of the World of Big Data

Lieut. Col. Ts. – serves in Aman

Introduction

Recent years have seen far-reaching developments in the field of information processing, particularly in the civilian market. These developments are called by different names, including the "big-data revolution" and "data science". They affect the daily lives of human beings all over the world at the personal and the public level, and they have a dramatic potential to influence the way in which intelligence is produced. The aim of this article is to survey the opportunities that this change creates for the intelligence community, the lessons that have been learned so far in Aman, and the challenges that still stand before us.

A Change in the World

Over the past decade the information field has experienced two global trends that are linked to each other. On the one hand, a change has occurred in how the world uses information-producing technology, which has fostered an exponential growth in the quantity of information. On the other hand, as a response to that phenomenon, several important technological developments have occurred in the field of information analysis that have made it possible to analyze very large quantities of information at great speed. As a result of these advances, a new industry has emerged that is developing new methods and devices, new positions in organizations, and special training in the information-processing field.

Over the past decade the information field has experienced two global trends - a change in how the world uses information-producing technology as well as technological developments in the information-analysis field, which allow the analysis of very large quantities of information at great speed

Alongside the production of this information, the main factors behind the change are the use of smartphones and of computer-

ization that can be worn on the person, as well as the Internet of Things (IOT) trend,[4] which has seen an enormous growth in the number of components of networked computerization that are found in homes ("smart homes") and in cities ("smart cities"). On the national level, too, there are large-scale projects in which very great quantities of information are collected, such as research projects in physics or the search for personally tailored medications through genetic mapping.

Beyond the[5] growth in the number of components that produce information, a change has also occurred in how people use these components, particularly in the social-networks domain, a consequence of which is that people provide more and more information about themselves (intentionally or unintentionally).

In the information-analysis field one can point to several trends:

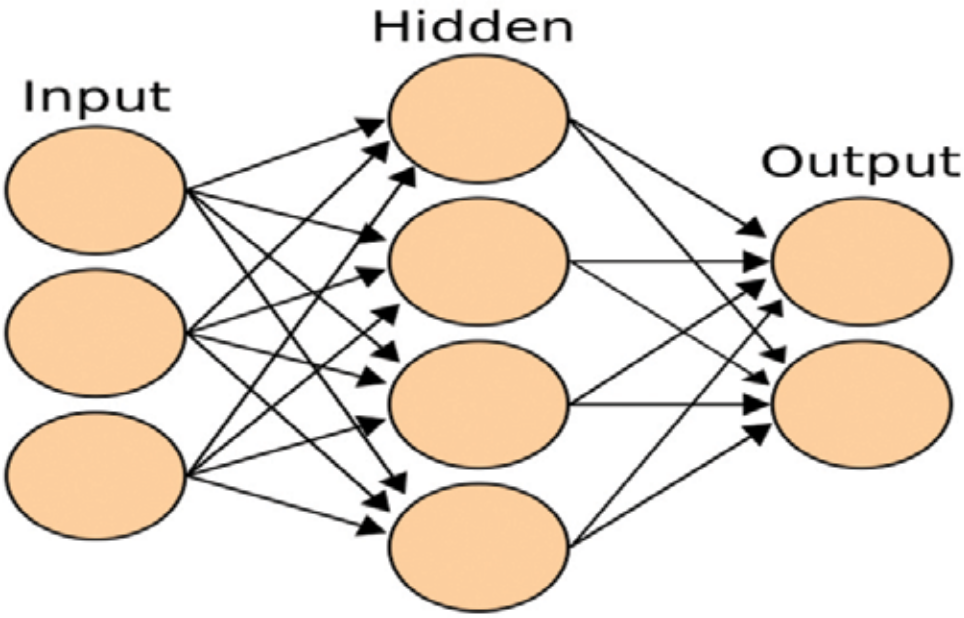
- **Decentralized processing technologies:** The maintenance of huge databases by the great technological giants (Amazon, Twitter, Facebook, Google) has led to the promotion of new technologies for dealing with information in decentralized fashion. The key aspect of these technologies is that they bring the decentralized processing to the information, instead of bringing the information to the processing units. Such computing forms the basis of the Hadoop family of technologies, which originally developed from technologies that were developed by Google and were opened for the use of the whole industry. The Hadoop project is an open[6] code project that defines a software framework for decentralized application and processes large quantities of data.
- **The algorithmics of machine learning and deep learning:** Machine learning is not a new concept, but in the past it required large quantities of manually labeled information for training the machine as well as expensive computational power. The algorithmics were also relatively limited. The comparable computational paradigm known as "artificial neural networks", which simulates the computational activity of neural networks in the human brain and enables rapid, equivalent machine learning, was also considered to have a limited potential that had exhausted itself.

4 It is hard to locate the source of the term; it appears in a 1999 presentation by Kevin Ashton who was then with Procter & Gamble.

5 "White House Special" with D. J. Patil, U.S. chief data scientist, partially derivative podcast, December 12, 2016.

6 Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung, "The Google File System", 2003; Jeffrey Dean and Sanjay Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters". 2004.

Fig. 1: A basic structure of a neural network.
(source: GFDL, Wikimedia commons)



The decline in the costs of calculation has decreased the cost of training; the increase in the quantity of information has facilitated training; and in the algorithmics field a breakthrough has occurred in research on the artificial neural networks with an eye to the use of deep neural networks. The main virtue of the new technology is that it enables machine learning from examples, without human beings defining in advance the differentiating patterns to seek in the information, in a way that is similar to human learning.[7]

- Commercial investment in technologies for handling unstructured information: Unstructured information (audio, picture, video) is information that is difficult to index, retrieve, and search. Algorithmics for audio and picture processing is a field in which much has been invested in academia and in applications of relatively niche-specific technologies (satellite pictures, voice laboratories). The shift in this field lies in the transition from research conducted in academia and in security organizations to research conducted by the computerization-technology giants (Google, Facebook, Microsoft, Apple, Amazon, IBM). Their investment is directed at new problems that stem from the existence of huge masses of

7 G. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets", 2006.

pictures and videos and from the developing industry of electronic accessories.[8] The enormous commercial incentive, the competition between the organizations, and the use of a new algorithmics have led to substantial improvements in the algorithmics of finding items in pictures and videos and in the algorithmics of audio processing. The algorithmics of identifying a speaker in an audio and of spotting a keyword in an audio is already at a commercial level, and for the first time the practical possibility of Speech to Text is being discussed openly. Face identification in pictures has also become an available commercial technology. Regarding all of these technologies there is still some distance between the public hype and their degree of practical development, particularly when it comes to problems that interest intelligence organizations, but this distance is shrinking.

- **Hardware acceleration technologies:** Graphics technology, which has served in the past to improve the quality of the picture in computer games, was found to be very well suited to the running of parallel big-data algorithms. The big-data field has become a prime commercial target for companies that produce such hardware.

Opportunities for Intelligence Organizations

There are two families of opportunities that enable the big-data revolution in intelligence organizations. One is the family of collection opportunities; today an increased quantity of information in fields that are relevant to intelligence questions exists in databases and in networked computerization systems in a way that is more or less accessible. People's willingness to provide a great deal of information about themselves in the social networks and in smartphone applications likewise facilitates information collection in a way that was not possible in the past, including information about people who are trying to conceal themselves. A second family of opportunities stems from the availability of commercial technologies for intelligence uses. If in the past the intelligence entities were at the technological forefront, and were the ones (along with academia) to deal with difficult research challenges such as processing a picture, processing an au-

Great opportunities for intelligence spawned by the big-data revolution: collection in many domains, along with technologies that are available for analyzing the abundant information

8 The Apple company produces Siri, Amazon produces Alexa, Microsoft produces Cortana, and Google produces Google Assisant.

dio, or researching immense databases, today the intelligence organizations can ride a wave that is led by the commercial industry and particularly the computerization giants.

The Big-Data Challenge in the Intelligence Organizations

Before looking at the unique challenges that the big-data revolution poses for the intelligence organizations, I will set forth several concepts that are relevant to the intelligence world and that describe the big-data phenomenon. It is commonly claimed that the big-data paradigm differs from previous information paradigms in three attributes, which are known as the 3 Vs: the volume of information, the variety of information, and the velocity with which the information needs to be handled.[9]

These concepts point to the difference between the challenges facing commercial industry and the challenges facing intelligence. Regarding the volumes of information, there is no intelligence body that can compete with the quantities of information in the Facebook or Google network. Also when it comes to the velocity of handling information, the commercial challenge is greater than the intelligence challenge; on most of the intelligence questions about information one is likely to receive an answer in time constants ranging from minutes to hours, and in certain domains in time spans of seconds. Even in the latter case, this involves what is known in the industry as "near real time" rather than "real time". which is expressed in time constants of less than a second. In the commercial world, determining which advertisement a user will receive when he browses the site, or the return on an answer to a question on the Google search engine, requires an operation in time constants of less than a second.

An exception is the phenomenon of variety. Unlike commercial organizations, for which the information they collect is often "theirs" so that they can standardize it, and which usually focus on information of a certain kind or of a limited number of kinds (credit-card transactions, tweets on Twitter, etc)., intelligence organizations harbor information that has been generated by a wide variety of sources, over which these organizations have no influence since they are not under their control. When it comes to variety, intelligence organizations are at the high end of the scale, and this poses a huge challenge for them.

The challenges that confront the intelligence organizations in general, and Aman in particular, are the following:

9 The first use of this distinction was apparently made in a report by the Gartner company. Since then some have added other criteria, and one can find references to Veracity (credibility of the information), Value (business advantage), Variability (change in the meaning of the information), and Visualization (the way of presenting the information). The original report by the Gartner (then META) company: Laney Douglas, "3D Data Management: Controlling Data Volume, Velocity and Variety", Gartner, 2001.

- **Coping with volumes of information**: Various publications speak of the exponential growth whereby the amount of information accumulated in the world doubles itself each year. In recent years Aman's measurement has indeed shown such growth, sometimes at an even higher rate. Part of the growth is explained by the increased success of collection, some of it by the increase in the quantity of information held by the enemy and about the enemy. This growth makes it necessary to establish an "intelligence information enterprise" based on commercial big-data technologies. Doing so, however, is hindered by the fact that the intelligence organizational budget has almost remained the same even as the volumes of storage grow. True, the costs of storage in the world are rapidly declining, but at a lower rate than the rate of increase in the quantity of information. And, more important, the limitations of military procurement and of classification require the use of computer hardware specially purchased at prices that sometimes do not reflect the global decline in prices. Consequently the intelligence-information enterprise is in a bind where it has to stop entering certain intelligence materials into storage, despite the huge resources and the risk that were invested in obtaining them, or reduce the duration of their storage, which diminishes the ability to ask numerous intelligence questions that are based on learning over time. The choice that is almost always made is to reduce the duration of the storage.
- **Coping with the variety of information:** As noted, this problem is not unique to intelligence bodies, but it is severe among them compared to the civilian world; intelligence must cope with an enormous variety of materials that are unorganized and unstandardized, contain duplications and gaps, and have changing degrees of reliability. This problem requires the building of devices and methods along with substantial investment in manpower for standardizing and organizing information.
- **Coping with contents that are unique to intelligence questions:** Although the civilian industry giants make large investments in the field of analyzing natural language, this investment is directed at spheres where the profit potential is the largest. Most of the investment goes into languages and dialects that are spoken by large, technologically advanced populations. Even the Arabic language, which is spoken by many people, is not at the top of the industry's priorities, and in the case of a spoken language like Arabic that has a large number of dialects, this problem is even graver. Beyond the issue of language, in other fields as well intelligence focuses on questions different from those that the industry addresses. Intelligence looks for different terms, a different vocabulary, and has different fields of interest than the civilian world.
- **The wisdom of the masses versus the intelligence "scoop":** The industry gi-

ants direct their efforts at issues that interest the general public. For example, Google uses the information for which browsers search to rank results for other users. The use of such "wisdom of the masses" in intelligence, however, is limited, since intelligence often seeks precisely the unique, "hiding" bits of information. Items that someone has already viewed are usually less, not more interesting. The intelligence community is too small and too divided into issues of interest to constitute a "mass" in the sense of the commercial algorithmics of the "wisdom of the masses". The intelligence organizations must also find solutions to questions for which the answer is a statistical anomaly or a singular "scoop". In the intelligence context, this is a sobering fact compared to elations over the notion of the social network and the wisdom of the masses.[10] Whereas the logic of commercial activity is to find the widespread common denominators and deduce from them the behavior of the crowd, intelligence, apart from the benefit it can gain from this approach, also searches for the anomalous and the unique.

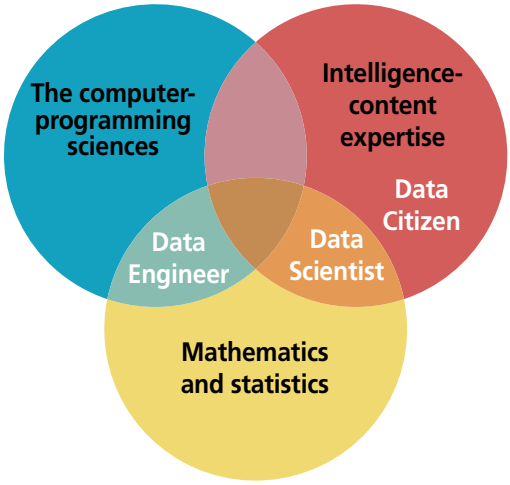
- **Learning of human physical phenomena and a learning competition:** Alongside dealing with physical phenomena (face identification, license plates, voice signatures, text font format), it is very difficult to enable a machine to learn human phenomena, which have an element of change on the time axis, or even to engage in a "learning competition" involving, for instance, behavior patterns of terror cells and individual terrorists or the combat doctrine of an enemy organization. The enemy that is perceptible changes its modes of activity, and the machine, which has to learn from examples, cannot necessarily do so. This difficulty intensifies in the transition from routine to emergency, since an enemy whose behavior is different in routine and in emergency can cause mechanized learning in routine situations to be ineffective in emergency situations; or, no less problematically, the learning in war could be too slow to affect the conflict's outcomes.
- **A structural challenge:** The intelligence community, including Aman, was formed organizationally before the digital era, in a world in which SIGINT, VISINT, HUMINT, and open-source information were disciplines that were foreign to each other. In the present era, in which most of the information collected is digital, this distinction has been blurred, and the structure of the intelligence community creates duplications and inefficiency. At the moment there exist several big-data bases built with different technologies that do not enable digital fusion and cross-checking with each other, so that intelligence suffers. What is needed is a new form of organization that concentrates the collected information in a

10 Calvin Andrus, "The Wiki and the Blog: Toward a Complex Adaptive Intelligence Community", Studies in Intelligence 49, 3 (September 2005).

single architecture and perhaps even in a single "information endeavor". It is very difficult to make such a change in the intelligence community in the absence of a structural change. Such change is likely to encounter harsh opposition since it entails breaking the traditional disciplinary identities. Besides the filtration, research, and assessment organizations, a new form of organization is needed because one cannot answer intelligence questions in the big-data era in the same serial fashion in which intelligence was generated in the past.[11]

- **The human challenge:** Although this challenge is not unique to the intelligence community, it is worth addressing. Change is also necessary in the technological intelligence organizations, which require specialists in a variety of new professions - from data engineers, whose task is to order and standardize the material, clean it, and enter it into databases in a way that enables retrieval, to data scientists, who employ different algorithms on the data to provide a response to the difficult intelligence questions. The qualifications required for these new professions combine professional intelligence knowledge (domain experts), mathematical and statistical knowledge, and programming knowledge. New qualifications are also required for the research and analysis organizations. Most of the future researchers need to be "data citizens", that is, possessing "data literacy or awareness". Here the intelligence organizations can benefit from the fact that these professions are currently being established in civilian life in general and in academia in particular.

Fig. 2: The big-data era requires human expertise.



11 See O. O., "Joint Investigation Teams as a Response to the Big-Data Era: The Test of Practice", Intelligence–in Practice 1 (December 2016); and in the same issue, Lt. Col. N, "An Intelligence Knowledge Community as an Operative Mechanism That Provides Strategic and Systemic Flexibility to Aman".

- **The technological change and the cloud challenge:** Most of the new technologies that have emerged in the big-data field make their way primarily to the public computerization clouds. The public-computerization-cloud market is a relatively centralized one in which a small number of huge, non-Israeli companies operate.[12] The intelligence communities, which deal with secret information, have severe limitations in making use of public clouds, and this is particularly true of the Israeli community. At the same time, it is now impossible to remain innovative in the big-data field without being connected to the open-code community and to the public computerization clouds. This challenge mandates a unique defensive technological response, and may also require a political effort. At present the large computerization clouds exist at a large number of places in the world. These clouds, which were constructed with a focus on the civilian market, also extend the services to the governmental world. The change, however, is slow and requires attention at the political level in contexts of functional continuity for emergency situations, cyber defense, and other political issues.

- **An ethical challenge:** Big data raises many ethical questions even in the civilian environment. For intelligence and security organizations, these questions take on a different hue. Effective action by intelligence organizations facilitates and requires the use of big data to penetrate privacy on a wide scale. Moreover, because intelligence organizations influence military operations, a transition from military activity in which the decision is entirely human to an act that is based on a machine - whether a machine that decides, recommends, or provides information relevant to a decision - requires a high awareness of the ethical aspects of the machine's activity, deliberate measures to ensure the quality of the information, and improved mechanisms for ensuring quality.

Summing Up: The Change That Is Required from the Intelligence Standpoint

Aman has been in a process of entering the big-data era for a number of years. The process has brought some impressive intelligence successes while posing not a few difficulties. In the coming years, Aman, if it is to exhaust the potential of the big-data revolution, needs to make several changes. In general these changes need to be based on three main insights:

12 At the time this article was written (2017), the market was controlled by AWS of the Amazon company, and after it came Azure of Microsoft and GCP of Google in that order, alongside a large number of clouds with a smaller market share. In recent years the market has been centralizing and (relatively) "small" companies have been leaving it.

1. Exhausting the potential requires a transition from local activity "on the run" with limited resources to a political, community, and military posture - for example, as part of the IDF's multiyear program.
2. The change is also paradigmatic within the intelligence organizations, and mandates recognizing the fact that the big-data field is a core field along with the traditional ones. This realization has far-reaching implications: organizational, administrative, and regarding resources.
3. In the big-data field, the intelligence community cannot and should not compete with the giant organizations that lead the field globally. Instead it should focus on the niche areas that are required for success and are not promoted by the industry, while using devices, technologies, and methods that are developed in the industry.

From Local Activity to a Global Program

At the lowest level of coping with the big-data world, there is a need for very large computerization resources for storing and processing data and for training learning machines. So far the budget in the intelligence organizations has been determined for the computerization issue, in the best case, in terms of linear growth, and often even less than that. Computerization has been dealt with similarly to other areas in an organization where storage is required and that do not grow exponentially from year to year. In light of the exponential growth in the quantity of data, the pressure on the information endeavor mandates deviating from the conceptual framework that views the equipment for Aman's computerization as a regular expense, like expenses for construction or food. Computerization equipment is very expensive in Aman terms but less expensive in IDF, community, and national terms. Hence it needs to be understood that it is a core IDF and community capability, to be funded accordingly.

When one looks beyond the quantitative aspect to the use of civilian big-data technologies, most of which are being developed in the civilian world, with an emphasis on the public clouds, it seems that both the IDF and the community form of organization may be too small, creating a need for political organization and effort. Adapting the intelligence and IDF community to make use of the public clouds requires coordinated activity and defensive, legal, budgetary, and technological preparedness. It requires a change in the procurement processes since in the public clouds payment is according to use, and the crafting of a legal policy that will determine which data can be in which cloud and in what way. The use of public computerization clouds entails a change in their physical location or in their defensive envelope. That entails very large expenses, which the large cloud organizations will be prepared to take upon themselves only if they understand that it is a worthwhile deal. Hence action is

necessary at the state level and not solely at the IDF level. For example, the American intelligence community has worked out a collaboration with the Amazon company in establishing a classified computerization cloud at an overall cost of $600 million. One can also[13] imagine that in certain fields that interest intelligence communities of different countries, there will be collaborations that allow the saving of resources as well as a focusing of effort on the information endeavor.

From a Marginal Field to a Core Field for Aman

The information domain, or IT as it is still sometimes called, is perceived as secondary to the core fields for Aman: collection, operation, and research. It is not seen as central to the intelligence agenda. For example, the infinite opportunities and the "scoop"-oriented vital-information marking have been the guiding factor for the different collection units, while the information endeavor's influence on them has been relatively marginal. However, a new perception of the information endeavor and its importance, as described in this article, entails a significant influence on the intelligence project both in the collection and research fields. For purposes of training a machine, one sometimes needs to use information that formerly was considered unimportant for collection. This does not mean that the information endeavor will be in the lead of the collection disciplines; such a change in the organization is neither expected nor desirable. There is a need, however, for a change in the balance of power, and for an increase of the information endeavor's influence on collection.

The information field must change from a marginal one to a core one in the intelligence enterprise, both in the collection and research domains. Resources must be invested in "machine training" even if those resources ostensibly lack direct intelligence value

The intelligence organizations' perception of the information issue as minor has caused the structure of the assessment and research entities to ignore this issue and its importance completely. The successes in using big-data technologies in recent years have stemmed mainly from the "changing of the diskette" of commanders who have grasped the revolutionary potential of the big-data field, devoted significant manpower resources to the issue, and altered the structure of their units to adapt them to the new era. Still, this change has been sporadic rather than systemic.

A third aspect of the consequentiality of the information endeavor in the past is

13 http://www.businessinsider.com/amazon-web-services-launches-secret-region-2017-11.

that the endeavor operated when a source arrived as a trigger, or in response to a problem that intelligence users presented to it. No deep thought was involved, nor any form of organization that served the information endeavor as such. The machine worked "in the service of the person" and not the other way around. Today we see the beginnings of a change in Aman as more and more people operate "in the service of the machine" - in the tasks of labeling information, training the machine, and building connected information and knowledge infrastructures in the different arenas, in relatively low proximity to a particular source or a specific intelligence question. The understanding that just as the information endeavor serves the researchers and the collection units, so these should serve the information endeavor, is developing in Aman but is still far from internalized.

From the Technology Front to Climbing on the Shoulders of Giants

The big-data field was born and promoted by giant companies such as Facebook, Google, and Amazon. Some of the technologies in the field are promoted today by large open-code projects. The realization that there is no way in which intelligence organizations can compete with the efforts of the civilian industry requires intelligence to open up more to the industry, to learn to use civilian technology, and to beware trying to progress on the basis of technology that differs from the civilian technology. The traditionally hermetic nature of intelligence, for reasons of secrecy but also from an unwillingness to rely on external development, could create e an obstacle for it in the big-data field. Aman must find solutions and tools that will enable its technological units to work with the industry and enjoy its fruits. That entails finding ways to connect intelligence personnel to the civilian environment and finding an appropriate defensive envelope.

The Big-Data Revolution from the Standpoint of the Giant Organizations

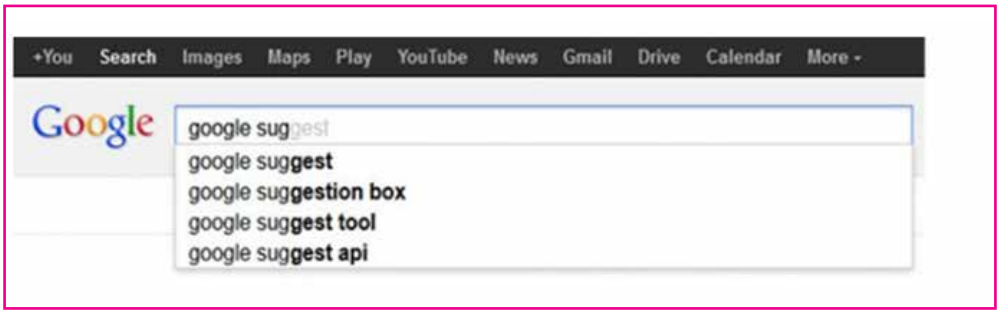
An interview with Dr. Yoelle Maarek,
vice-president of research, Amazon

Interviewers: David Siman-Tov and Lt. Col. Ts.

Question: Dr. Maarek, what do you think of the term "big-data revolution".
Answer: "Big data" is a marketing term. In the past they used the term[14] "data mining", and that's still the preferred term in professional and academic circles. "Big data", however, deals with connecting huge quantities of data to enormous computerization resources. Making connections of that kind is certainly a revolution.

Question: In what sense is it a revolution?
Answer: When we talk about big data, we need to talk about both sides of the phenomenon, the positive and the negative. The positive side is the big opportunities we now have.

In general, the search engines in the world have failed for three main reasons: either their ranking algorithm isn't good, or the content that the user is searching simply isn't in the base, or the content is there but the user doesn't know how to translate his need for information into an effective search query. In an attempt to solve the third problem, when I was at IBM my group and I developed algorithms to help with questions on PalmPilot, but we used the documents themselves to produce the suggested queries, and the queries didn't look natural. Later, when I joined Google, I discovered that there's a prototype for helping with the query, which produces a static group of documents that runs on labs.google.com. The beauty of this device was that



14 The source of the term is apparently John Mashey, chief scientist of the Silicon Graphics company during the 1990s.

it served as a basis for real queries of real users. I pounced on the opportunity, and the challenge was to change this functionality from static to dynamic so that it could run on all the queries in the Google search engine and be updated all the time. After two years of work with the development team in Haifa, we launched Google Suggest. This project requires enormous quantities of data (millions of users) and parallel computing platforms (MapReduce).

Today, in Amazon, we're taking the challenge one step further. Our mission is to do what's best for our customers, and we have to process a great many signals in order to understand them and supply their needs. That's the reason we need cloud services such as AWS (Amazon Web Services) that not only support parallel processing but also promote processing services for machine learning. In the organization that I'm part of, Alexa Shopping, we use big data to turn the idea of artificial intelligence into reality. We enable the customers to ask about products and to buy them by speaking to an Echo machine. This changes the[15] paradigm of human communication with tools, and it motivates us to expand the boundaries of science. In this space that disrupts reality, we wouldn't be able to progress without using huge amounts of data and powerful computing platforms, but many content experts as well, so as to really understand the customers and start drawing the right conclusions from there.

Question: Is this something that intelligence organizations can do?

Answer: The examples I've mentioned point to the limitations that security and intelligence organizations face when they try to use similar techniques. Governmental organizations don't have millions of users. They try to compensate for that by asking users to provide explicit feedback, using machine-learning techniques, and using domain experts to label information for the algorithm. But this is challenging because[16] machine learning is less successful when it operates on small amounts of data, and intelligence organizations need to invent new approaches, or connect to external resources, to compensate for the relatively low quantities of internal information [that come from users].

Question: And what is the negative side of big data?

Answer: The negative side of big data is of course "big garbage". Sophisticated algorithms, especially algorithms for deep learning, show excellent results, but it is very hard to understand the operation of the machine and to explain it. The algorithm

15 The smart speaker/microphone of Amazon, known as Echo, which is linked to the Alexa service.

16 In a certain sense this is a renewal of an old practice but for different reasons. In the past users would mark keywords in texts because the search engines would index only those keywords. Today specialists need to mark keywords so as to help the machine-learning algorithm.

becomes a kind of "black box" and we have to count on it playing its role well. Lately there's been public concern about the dangers of artificial intelligence, but artificial intelligence is dangerous only when you use stupid algorithms and stupid scientists. I believe in a careful approach where we thoroughly check the algorithm and understand why we accept results of one kind or another. This is what we mean by "interpretable". It's irresponsible for a scientist to say the reason he accepts certain results is because "that's what the machine decided". Each step has to be monitored so that the analysts (content specialists) can verify the results. Of course, this is even more important for security and intelligence organizations, which make life-and-death decisions with algorithms. There is still much room for activity by human beings, especially because positive examples can change over time. Analysts have to adapt themselves to this new[17] era - to understand the concept of features, and how their actions contribute to the operation of the learning machine. For this purpose intelligence organizations have to construct a spectrum of training. Researchers and analysts need to know how to write code, and programmers need to understand the business field they operate in. On the most basic level, intelligence organizations at least need to produce heterogeneous communities of business experts and technology experts in order to bridge the gap.

Question: Are there other ways in which intelligence organizations differ from civilian organizations in the big-data context?

Answer: Intelligence organizations confront a more difficult challenge than civilian commercial organizations both because they're closed and because they're often limited in their ability to use trial-and-error approaches. They are very sensitive to the results of recall and not only to precision. So their problem is more complicated.[18]

Unlike electronic accessories, which many people are willing to use even though

17 Positive examples are examples that the machine uses in order to learn. The reference here is to cases in which there is a change in the phenomenom itself, so that renewed learning has to be created.

18 The term "recall" describes the extent to which a search engine finds all the possible correct results. The term "precision" describes the extent to which each result that the search engine returns is indeed correct. These are the two common parameters for comparing search engines, and there is an inverse relation between them (the more one improves precision the more one detracts from recall, and vice versa) but not a linear one. For example, the Google search engine is very precise and people often receive an answer to their question in the first results that it returns, but there is much less emphasis on recall; it may be that many results very revelant to the user do not show up at all in the search results and the user does not know about them. In cases where missing an important item could cost human life, making a choice based on recall/precision measures is risky.

they still aren't perfect, thereby providing usable information that allows ongoing improvement of the product, intelligence users are less patient.

Intelligence organizations face another considerable difficulty. Because they are closed, they're limited in their ability to use cloud infrastructures and public software services. Amazon, for example, is developing the electronic accessory Alexa as a platform and enabling other organizations to build specific knowledge fields for Alexa ("skills") above the platform. This also helps these organizations, which don't have to develop a speech-identification application by themselves from scratch; it also helps Amazon add other applications for its customers. Fewer and fewer software organizations in the world build their services for "closed" organizations, and intelligence organizations could find themselves sealed off and lacking access to the wealth of software services and resources that exist in the public space.

Question: In your opinion, what will be the next big breakthrough?

Answer: I believe that voice-run machines are the next revolution. In the future all the machines will be voice-run. I'll leave it to the intelligence organizations to analyze the opportunities that await them on the issue.

Advanced Data Retrieval in the Big-Data Era

Dr. Haim Assa
architect of artificial-intelligence-based systems

The Sources of the Data

Intelligence is received from collection sources of various kinds. These sources are distinguished from each other by different sorts of access to information - visual, SIGINT, human, open-source, and so on - but each category also includes a great variety of subsidiary sources that are differentiated by how the information is obtained. Visual intelligence, for example, can be obtained from cameras that produce pictures and videos from airborne platforms such as satellites or aircraft with or without a pilot; from cameras emplaced on vehicles, balloons, or at ground observation posts; from manned observation posts in which the lookouts write what they see in texts, and so on.

Usually collection is done from predefined "entities" - particular people or systems, such as missile batteries of different kinds or various command and control systems. Collection can be performed on several "wavelengths" - that is, by a number of collection methods with different wavelengths such as "textual" and SIGINT of various kinds.

Often added to the data is the interpretation of the person responsible for obtaining it and for its initial processing, even through translating or summing up the raw data. In addition, "intelligence assessments" by researchers and intelligence officers of the data that is streamed by the collection personnel are a kind of intelligence entity that has a unique status and takes different forms (free text, text in a regular format, presentations, videos, maps, etc)..

One of the important developments globally in recent years regarding the processing of intelligence information is the realization that the products of the collection personnel need to be integrated. A SIGINT source provides data of a different kind than optical data, and information written in a text by an intelligence officer and disseminated on various networks is yet another kind. Each intelligence worker presents the information from his own perspective, which usually differs from that of his fellow. SIGINT, for example, does not identify the platform from which the signals are broadcast, which could be an aircraft, a car, or a building, while optical-collection personnel are capable of doing so but do not have the SIGINT information whose source is in the entity that they film. Because a common major factor for everyone is time, precision in defining the point in time at which the different data items are

received is critical. Hence the integration or fusion of these components of the information is likewise critical and hugely valuable when it comes to understanding the overall intelligence picture.

As far back as 20 years ago, a markup language was developed that posits a set of laws, which, in turn, create a kind of "markup" of words and documents that enables a machine (a computer) and a person to read them and understand them, and also enables the transfer of data between a machine of one kind and a machine of another kind. This is a protocol that is characterized by the field on which one wants to focus when transferring data between different machines, such as the field of sports, economics and business, and so on.

An example is achieving integration between textual and electronic information. Textual information is seen as unstructured - a free text that was written by an analyst, lookout, or intelligence officer. Digital information is structured, meaning that each particle of it can be stored in an organized, permanent fashion - with regard to time, frequency, location of the source, and so on. Text, however, is not organized, and each person writes as he wishes. Forging integration between the two kinds of data requires a "structuring" process. That means each sentence is exhausted to its core and that core can be retrieved in an orderly manner, thus achieving the structuring that can be combined with the digital information with the addition of two elements: time and place.

Digital data - unlike textual data - is structured. That is, each of its particles can be stored permanently

Every text is built from sentences. Each sentence has a triad of elements that forms the core of human language: subject - verb - object. Today those elements can be retrieved with advanced technologies built on a basic analyzing capability (Analyzer) and the analysis of parts of speech (Parser). Integrating or fusing two kinds of structured "information" is already an easier task, facilitated, among other things, by semantic databases and by techniques based on semantic networks (see below). Here we encounter the notion of big data, since we need to store immense quantities of information of the different kinds in databases of a new kind that have an enormous capacity and, moreover, an indexing capability from a new world.

Indexing entails that each particle of information that enters the database has an address where it "exists", and knowledge of that address enables one to retrieve it with a query that is determined by a researcher or an automatic system. The aim is to use the relevant information particle in accordance with "time and place" in order to fuse it with an information particle of another kind at the same point in time and in the same place.

Until recently the capability to quickly retrieve and fuse all the relevant information from the huge database was limited. The technological solutions for achieving capabilities within the framework known as big data were created only in recent years, and even these solutions were partial. The great recent breakthroughs have involved improving the indexing capability as distinct from the storage capability - for example, the Elastic Search method or a device developed by the Attivio company.

But this does not mean the challenges created by these requirements have ended; how does one "fuse" SIGINT information with optical or textual information? What does it mean to fuse? How can an automatic machine (i.e., a software) know how to connect one kind of information component to another, and how can an automatic machine infer what an intelligence researcher is supposed to infer, if at all? Before attempting to clarify how an automatic machine is meant to carry out the fusion, I will emphasize that only an automatic machine can cope with the vast oceans of data that are stored at fantastic speed in the abovementioned databases.

Fusion

Fusion means obtaining information that is produced by integrating at least two information items about the same object (people, places, territories, structures, etc).. Fusion between collection products that are "electronic signals" is based on time and place. If, for example, a person is perceived by a COMINT device at a certain place and at a specific point in time, while some sort of radar detects the car in which he is sitting or driving at the same place and the same time, it can be determined by a calculation, not too complicated, that the person who spoke on the cellular phone is the same one who is traveling in this specific car. When the fusion is between a content that is presented in texts on the one hand and electronic, signal-based information on the other, the fusion becomes more complicated and requires a "common language".

Over the past 20 years standard "bridges" have developed that facilitate "understanding" between different computerized systems; for example, XML is a standard from which many derivatives have been generated. The common language, however, is not sufficient. It also requires a definition of the field of interest in which the "conversation" between two different systems is conducted.

Ontology

A common language entails defining the field of interest in which the fusion exists. For example, if the field we are dealing with is maritime, the assortment of relevant entities in the sea differs from the assortment of entities we are familiar with in the aerial space. In the sea we deal with entities such as ships, submarines, maritime conditions, fish, currents, and so on, while in the aerial space we deal with planes,

helicopters, drones, birds, and so on. Sometimes geography is also part of defining a certain field, since in one particular area - a land area - the relevant entities are wadis, mountains, and caves, and in another space the relevant entities are desert water sources, oases, and so on. "Creating an ontology" means defining the relevant entities and activities for an issue on which we want to perform fusion, or gain insights from the emergence of patterns whose details, while they indeed exist somewhere in the gigantic databases, are like a needle in a haystack.

It is clear from the foregoing that the issue known as big data does not solely concern the storing of large quantities of information. It also requires indexing to facilitate data retrieval and a search capability for the sea of data that allows us to identify the different behavior patterns of entities or the development of physical, chemical, or other processes. The details pointing to the existence of such patterns are in the databases, but they are few and sometimes do not appear exactly as we want them to. And most difficult of all is to connect a datum that has been identified with another datum that has been identified, and to construct from these a "more complex product" that is referred to as a pattern. This difficulty also has to do with the mathematical aspects of identifying a pattern and with aspects of retrieval from very large databases. The difficulty reaches its peak (even becoming a crisis) when we deal with databases that are based on semantic graphs, as we will see below.

Patterns

A pattern is, for example, a process in which Ms. Cohen, who is a mother of two, leaves her house every morning along with her two children, puts them in her car, drives to the kindergarten of the younger child along a regular route and for a duration of five minutes, and when she arrives, gets out of the car, opens the door for her younger son and helps him get out, kisses his forehead, and the boy runs to the kindergarten and disappears. Ms. Cohen gets back in her car and continues to the school, drops off her older son in the same way, then continues to the Afarsek café at the outskirts of the city. The data on Ms. Cohen are amassed from different collection sources and entered into a single database, using the common language described above. The sources are textual: reporting by a lookout who saw Ms. Cohen in different places; Ms. Cohen's cell-phone conversations, which indicate her location at different points in time along her car route; the knowledge provided about the location of the kindergarten, the school, and the Afarsek café; and data about the woman herself. These data are defined as traits of the respective entities (i.e., Ms. Cohen, the two children, the school, the kindergarten, and the café are entities).

The difficulty lies mainly in retrieving all the relevant data, in connecting them to each other, and particularly in the fact that the data that are stored are "imprecise".

That is, they are not precisely the "signal or word" that we seek. Indeed we do not know Ms. Cohen's driving pattern. We know a few details beforehand, but not the recurring route that is represented partially along with billions of other data about similar cars and similar women to Ms. Cohen.

There are "imprecisions", or data that do not directly and immediately indicate the existence of Ms. Cohen's pattern. For example, a lookout has identified the Subaru that she usually drives to the kindergarten, but he has no evidence that she is the one who drove the car. A cellular sensor that Ms. Cohen activated was recorded at a distance of a hundred meters from the school as she was leaving it, and another lookout says he saw the Subaru two minutes after the moment at which she talked on the cell phone. The data of a radar that monitors the town for the municipality show that a "certain" car indeed passed the place where a cell phone was pinpointed. A rapid check by the software shows that this car's route indeed went past the school and the kindergarten, and the point in time at which the lookout saw the car is "approximately" the same point in time that the radar indicates.

Using all these data, the system must make a decision that it is very possible that this is Ms. Cohen. But this still is not a pattern. A pattern is a process that repeats itself. That is, the system needs to identify repetitions of this process of Ms. Cohen's. And, indeed, the next day data were obtained that were similar to the previous data - but Ms. Cohen's cell phone did not operate. It is still hard to say that a pattern has emerged, but for several days these sensors have kept "filling" the database.

The difficulty lies in the fact that other cars pass along the same route, other cars stop beside the kindergarten and the school, very many cars are observed on their way to the café, and so on. The same pertains to the cell-phone conversations; huge numbers of mothers are talking on their cell phones at this hour, and the question is how the system can identify the pattern of Ms. Cohen. The difficulty is greatest when we do not know that Ms. Cohen is the entity performing this pattern. We know about the "vague" existence of the pattern, but we do not know who it is that is enacting it. The technology for analyzing the information used to identify patterns in the large databases has developed greatly in recent years, and it allows us to contend with these questions - indeed not easily, but in very advanced entities the technology achieves a certain success.

A Learning Machine

One of the important requirements for intelligence of any kind is "prediction". Identifying patterns of activity in very large databases (with trillions of entities) is a big step forward, but making a prediction on the basis of the data in the databases is an even bigger one.

A prediction is based on study of the processes and events that have occurred in history, or on a pattern that recurs a significant number of times. A learning machine is based on two elements: training and classification. When a new text reaches a learning machine, the machine is supposed to decide which category (subject of interest) the text belongs in: sports? science? politics? and so on. "Learning" means that the learning machine "trains" on "examples". The examples are supposed to represent what the machine "looks for". In the textual world, for instance, there need to be a few dozen texts that represent a subject of interest. The machine will train on these texts and generate a "feature vector" that represents this category.

A "feature vector" is a set of words or combinations of words that represent the issue in question. For example, words such as "judge", "ball", "fans", "players", and so on indicate a text connected to a game such as soccer or basketball. To distinguish between the two games, the training must involve learning from texts about either basketball or soccer. The learning process will produce a vector based on anchors (words and combinations of words) that are characteristic of soccer or basketball. For example, a penalty kick occurs only in soccer, while basketball has other characteristic features, such as timeouts or foul shots, that do not occur in soccer.

In general, when a new textual document reaches a learning machine, the machine will "read" it, retrieve the relevant set of anchors from it, and compare it to the set of anchors (feature vector) that represents the category. The same occurs when identifying pictures, processes, or patterns. Identifying a process means being able to come up with a prediction. For example, in the case of Ms. Cohen, who takes her children to kindergarten and to school every morning, the system can specify her "next" step at each stage of the process that she performs. When Ms. Cohen leaves her house, the system can assess at a high level of probability that she will arrive at the café after a certain amount of time.

When a new textual document reaches a learning machine, the machine will "read" it, retrieve the relevant set of anchors from it, and compare it to the set of anchors that represents the category. The same occurs when identifying pictures, processes, or patterns

Although the example of Ms. Cohen is simple, it incorporates many other highly complex processes. For example, the system has determined that each time snow falls in Korea and the Japanese stock market falls, and a new president is elected in the United States, North Korea conducts a new missile test after a year (that is, a year after the U.S. president is elected). In order to identify this pattern the system must

investigate hundreds of billions of data, perhaps trillions, and reach a conclusion that this pattern indeed recurs and is indeed a pattern. The next time it snows in Korea and the Japanese stock market declines, and if it is only a few months after the election of an American president, the system will be able to predict at a high probability that the North Koreans will perform a missile-launch test. It will also be able to predict that after the snow in South Korea and the election of the American president in the preceding months, North Korea may carry out a missile test, but at a lower probability than in the previous case - but still at a high-enough probability to call for alertness.

The great difficulty in the above-described procedure lies in identifying this pattern (snow in Korea, etc). amid the immense sea of information particles, a pattern of a certain kind that recurs sufficiently even in different variations of "names of entities" or "names of events". Prediction itself, once a pattern is identified, is not especially complicated. The revolution lies in identifying these patterns among the mountains of data that need not be related to each other logically or in any other fashion.

Semantic Web

Big data is a very general subject. It has two main components: the manner of storing enormous quantities of data in a database, and the ability to retrieve the data rapidly and connect them with each other so as to derive an insight or an inference. One way to do so is to connect data that are still within the databases through uniform representation of the data. For example, entities are represented in a uniform way, actions are represented in a uniform way, and so on. Defining such "representations" is the essence of "ontology" in its engineering sense. In other words, a kind of editor (which has been developed by many companies) makes it possible to represent entities, actions between them, and their properties on a single graph. This capability tries to produce the WEB 3 technology, which is based on semantic databases or what are called semantic graphs, which, in turn, make it possible to create links between entities according to their common properties and according to their links with other entities (linkification), and also according to many other components insofar as the user can specify them. The semantic graphs constitute a great revolution in the world of information processing, but the systems that have been developed still have difficulty contending with vast quantities of information (up to 50 billion entities). But this issue, too, is on the verge of very creative solutions, including some by Israeli companies.

Conclusion

Big data is a concept entailing two components: a storage capacity that includes the indexing technology, which makes possible the rapid retrieval of any information

particle from huge databases; and the ability to perform manipulations rapidly among trillions of data (at the threshold of real time) and thus reveal patterns and links, thereby generating inferences that a human being cannot make because of the tremendous quantity of data. The field of inferences is still in its infancy, but already today there is considerable development in the technological sphere that focuses on algorithms and prediction processes based on learning machines and "deep learning" - a refinement of a learning machine that uses networks of neurons to extract information from several strata of information (as distinct from a classic learning machine). The deep-learning technology makes possible important breakthroughs on an issue such as analysis of a picture. Likewise, analytical processes based on Bayesian[19] algorithms, networked and dynamic, allow "decision-making" by robots even in the absence of historical information. The reinforcement learning-machine technology was developed to augment this capability. That is, the system "corrects" the components of learning as - again - a robotic system that lacks historical knowledge is activated.

All these capabilities are intended to draw inferences from the sea of data we live in. I believe that these capabilities will be further improved and refined, and that capabilities of a new kind will be created both for civilian and military uses. One can already predict the next advance in the cybernetic domain. If so far this domain has disseminated information, whether in the social networks or in systems such as Google and Yandex, in the not too distant future we will also encounter the dissemination of "knowledge". It will be knowledge that is relevant to everyone and derived by integrating information from several sources, using deep-learning techniques and other modules according to the needs and interests of each one of us. These interests will be identified by a global network (such as Facebook), which - with a push - will provide relevant "knowledge" according to our needs.

19 Bayesian algorithms are algorithms that weight both the assumption on which they are based and the inverse assumption.

What the Intelligence World Should Learn from the Civilian World and What to Avoid When It Comes to Storing Big Data

Eran Baron – CTO of Infinidat[20]

Introduction

In his outstanding article in this issue, "Intelligence By-Products of the World of Big Data", Lt. Col. Ts. addresses many operational and intelligence concerns and notes that efforts are "hindered by the fact that the intelligence organizational budget has almost remained the same [in recent years]". The present article analyzes the topic from the standpoint of infrastructure (the main component of the cost in these projects) and considers some trends in the big-data world that have been adopted by commercial actors, and their compatibility with the intelligence sector. The two worlds perhaps share challenges that are similar in their computational nature, but they are fundamentally distinct in terms of organization and of the budget available to them for infrastructures. Whereas a commercial organization like Amazon is willing to invest another $100 million in infrastructure that it believes will return its cost within a year and generate a profit for many years, it is more difficult to quantify the return from an intelligence system.

In his article Ts. notes the advantages of leveraging infrastructures from the open-source world and the civilian world in general (to the extent that it is possible) while avoiding the development of equivalent tools that will not be updated, maintained, or improved to the same extent over time. Whether or not the intelligence world chooses to adopt this approach, decentralized infrastructures create challenges in the infrastructure domain that can be analyzed separately from a particular software device, and this article deals with that topic.

Between the Civilian World and the Intelligence World

Large civilian companies such as Amazon, Google, and Facebook have considerations that differ from those of the intelligence world and enjoy greater room for maneuver. Hence the devices that they develop for big data and release to the world,

20 Eren Baron, CTO of the Infinidat company, has over 12 years of experience in the software world and is focused on the storage world. Infinidat was founded by Moshe Yannai and helps its customers exhaust the full potential of their information. The company's software-based architecture meets the conflicting requirements for larger, faster, and less expensive information storage.

as well as the devices that are developed in the open-source community but receive most of their code from these companies, are developed according to the same basic assumptions. Below are some examples of differences between these companies' considerations and those of the intelligence world:

Fig. 3: A comparison between the big-data considerations of the civilian world and of the intelligence world.

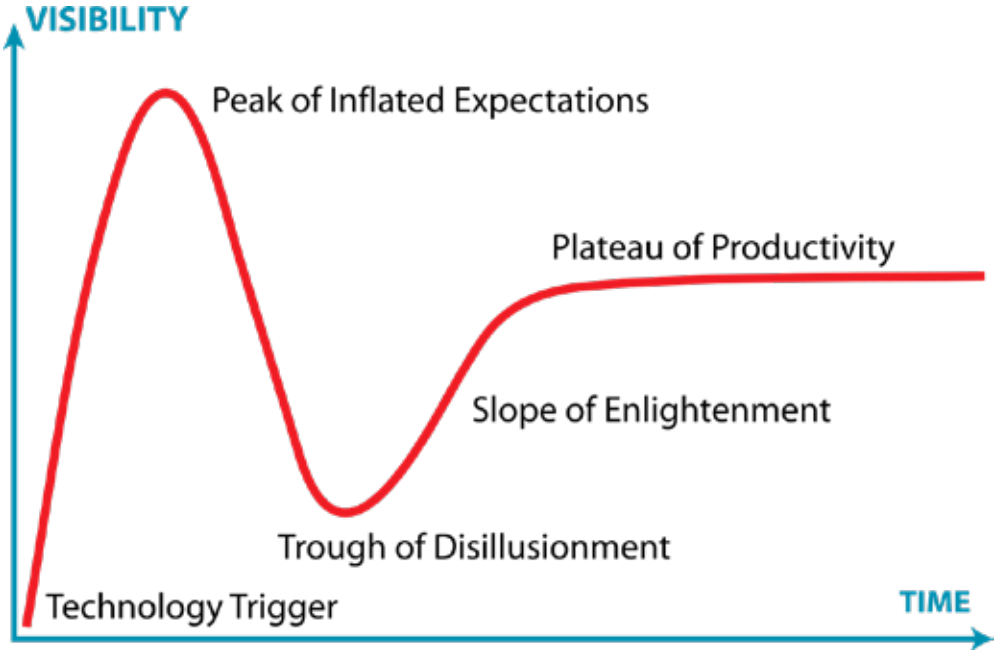
The challenge	The civilian solution	The difference from the Israeli intelligence world
Lowering hardware costs	Purchasing White Box servers from sources in the East	No way to purchase with U.S. military aid from foreign suppliers who do not have representative offices in Israel
Because cheap servers are not reliable, the data must be duplicated many times in a way that wastes more electricity and space	Building computer rooms in cold countries so that there is no need for refrigeration	Restricting the databases' location to Israel
Inexpensive hardware requires troubleshooting at the lowest level such as driver services	Setting up designated teams to write the drivers	A unique expertise that entails lengthy processes
Complex systems require high-quality maintenance	Preserving manpower through high economic incentives so as to preserve organizational knowledge	The civilian world can offer higher incentives than the military world; many programmers want to open a company of their own

"Don't Throw Out the Baby with the Bath Water"

In order to appreciate the challenge of adopting new technologies properly, one needs to understand the way in which the technological community adopts a new technology - which involves what is known as the hype cycle:

- A new solution appears in the market that seemingly answers a need.
- (Too) rapid adoption of the solution (usually without sufficiently researching its compatibility with the need).
- Sobering up - those for whom the solution is not suitable abandon it (some do so because a misguided building of infrastructures has led to cost inflation).
- Enlightenment - more and more ways to use the new solution are proposed, and the number of their customers grows accordingly.
- Productivity plateau - the solution reaches maturity both in terms of capabilities and of being adopted.

Fig. 4: The curve of the hype cycle of new technologies (Wikipedia).



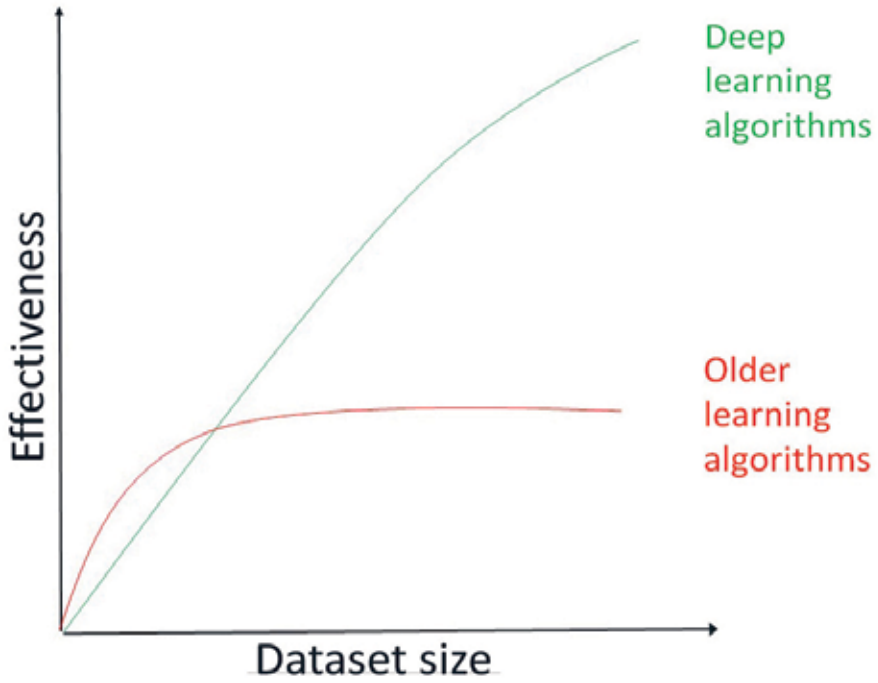
(Wikipedia, Author Jeremy Kemp)

The rapid adoption is a consequence of need. Usually, though, no one ascertains at the adoption stage whether all the basic assumptions of the solution are compatible with the environment. Ignoring such incompatibilities sometimes leads to technological or economic failure. Economic failure, which occurs when the technology is too expensive, prompts many customers to abandon the technology completely and look for the next wave, instead of reexamining those basic assumptions that they ignored at the beginning. Sometimes a reconsideration of those assumptions can completely alter the economic model and turn failure into success.

What Drives the Rise in Costs?

Much of the budget needed for these projects is a consequence of the large quantity of information that is involved. With the old machine-learning devices there was a clear boundary for the quality of the product, and at a certain point there was no reason for an infinite input. With modern deep-learning-based devices the quality of the product improves as the quantity of the input grows:

Fig. 5: Comparing quality of the product between old algorithms and deep learning.



Because the architecture of most of the big-data solutions is decentralized, one of its basic assumptions was always that the data would be preserved within the servers. While this basic assumption stemmed mainly from the need for performances, and from the belief that traditional storage arrays could not meet the performance demands of a big-data environment (which was indeed true at the time), it also created great waste when every bit of information had to be preserved numerous times (and hence also rising costs).

Nowadays new storage architectures again make it possible to separate the information from the server without harming performances while significantly lowering costs. Why do the abovementioned cloud giants not do this? For the same reason already mentioned: their belief that everything can be diffused in a software that runs on a cheap hardware, with not inconsiderable investment in developing an open code at the "low" levels of communication with hardware components.

On Open Costs and Unseen Costs

One of the difficulties in calculating the overall cost of big-data projects concerns the second and third circles of the costs. The decentralized architecture requires recovery from any failure (of a disk, server, etc). by sending information over the network, which raises the bandwidth needed for the network and hence inflates the cost of the communication item. For the most part these projects are priced according to the assumption that the computerization and storage are the main costs. In practice, however, projects of this kind have shown a recurrent pattern:

Fig. 6: Calculating the costs of a big-data project.

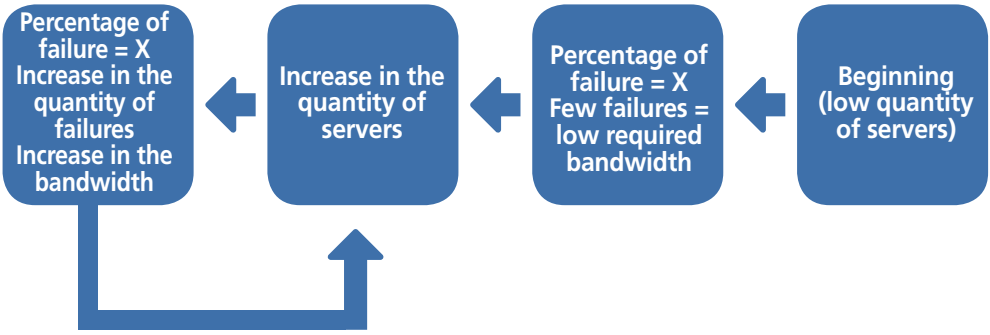
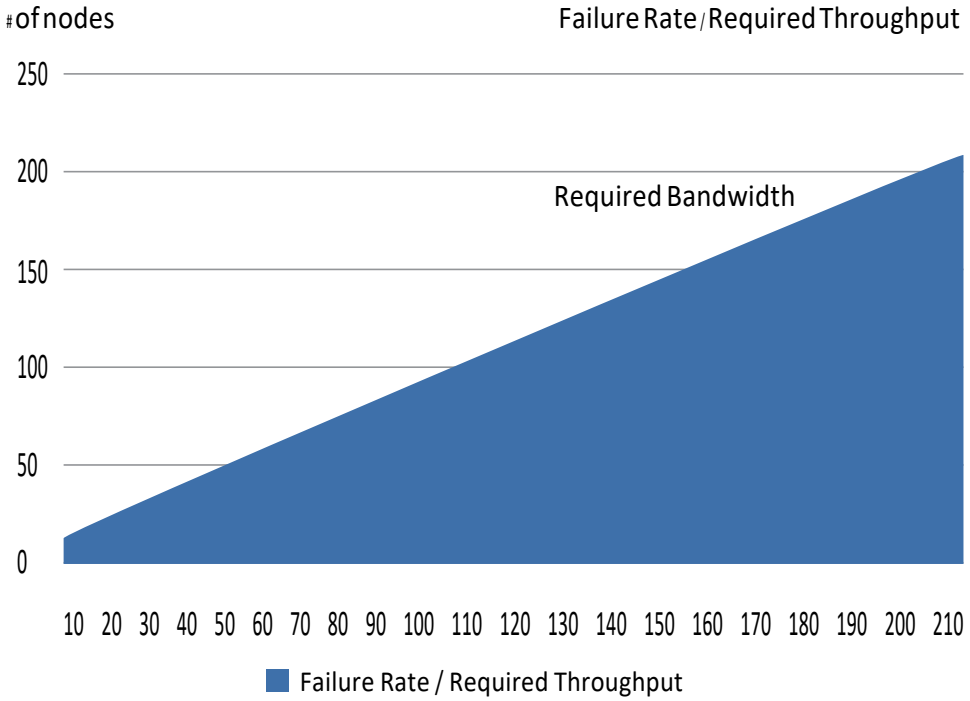


Fig. 7: Failure rate vs. information needs.

Over time, this behavior appears as follows:



In light of these problems, the American Sprint Corporation decided to transfer its Elastic Search solution from an integral architecture (information preserved in the servers) to a separated architecture (information preserved separately in a storage system).[21]

Conclusion

The world of Israeli intelligence does not enjoy the same budgetary freedom and the same abilities to invest in developing code that the cloud giants enjoy. The Israeli intelligence community must employ different considerations regarding these solutions when it comes to its operational needs, adapting the infrastructures (the main cost component) to its reality. Passing over that stage can create a mistaken picture at the command levels of a system with needs that are not sustainable over time, and lead to the rejection of solutions that are vital to the efforts of the corps.

21 One can hear about their challenges and about the improvements in terms of costs and performances that the transition achieved for them in their lecture at Elasticon: http://www.elastic.co/elasticon/conf/2017/sf/it-as-the-transmission-of-the-sprint-business-engine

Analyzing Network Intelligence in the Big-Data Era

Maj A. – serves in Aman

What Is a Network and What Is Its Contribution to Research?

Through observing the behavior of networks, network research allows a deeper understanding of the role of the research objects and of the common phenomena on the network while identifying "centers of gravity" (main influential factors).

Fig. 8: A semantic network compared to a flow network.

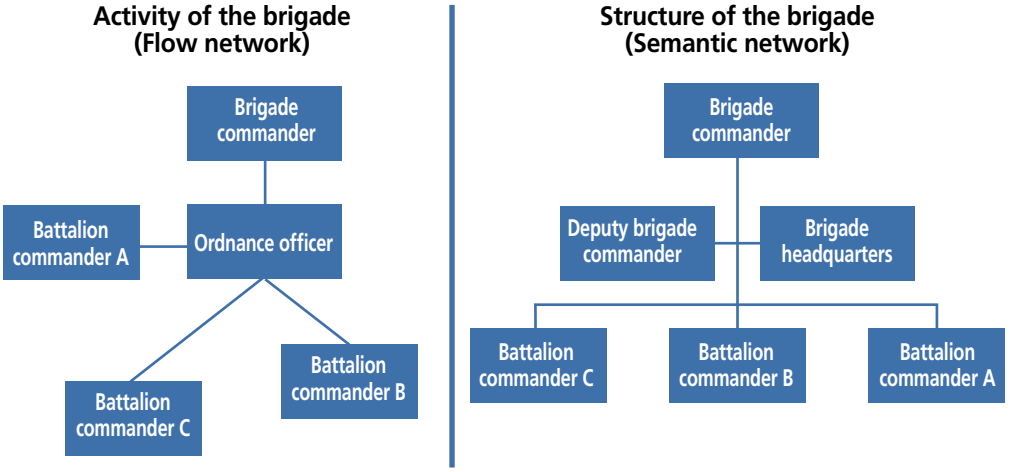


Illustration: Ostensibly the key actor in the brigade is the brigade commander (semantic network), but in practice, when the ammunition runs out, it turns out in the flow network (the connections created between the components of the brigade) the ordnance officer is the center of gravity.

This article deals with networks. For purposes of a common language, a *network* is all the information that is comprised of *nodes* and *edges*. The node is a point in the network and the edge is the line that connects between it and other points.

A classic case is that of social networks (Facebook, Twitter, etc)., including telephony and other networks; but links between computer servers, passes in a soccer

game, or an interaction between proteins can be characterized as a network as well. Organizational research is almost inevitably network research, since an organization can be characterized by the interactions that occur in it. The network is essentially a tool of expression that allows one to define the nature of the interactions and the relations between the elements of the system.

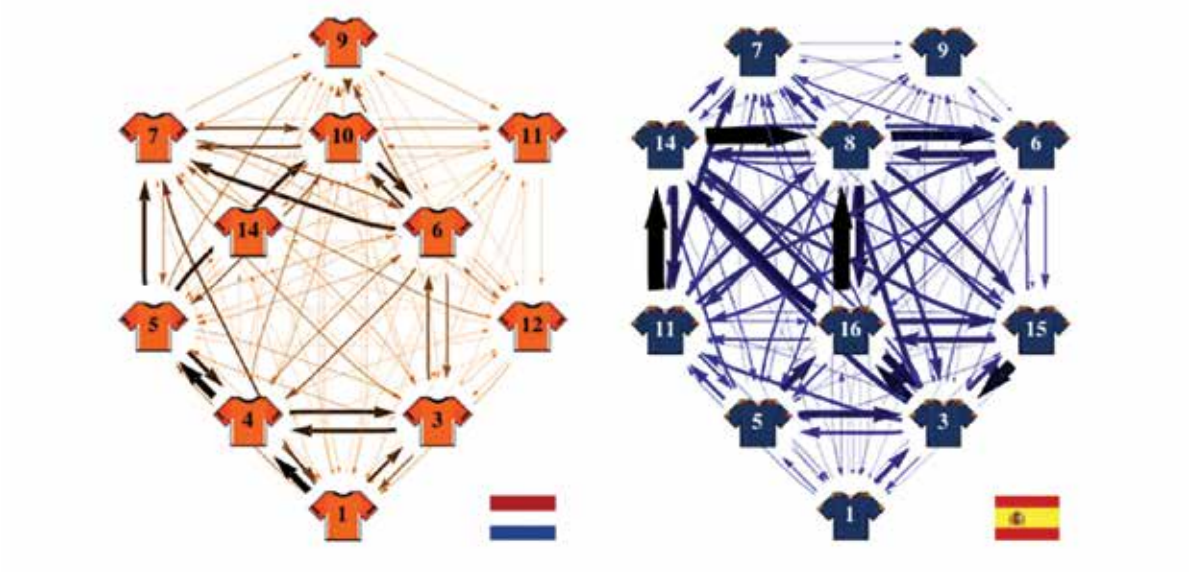
How does social-network analysis (SNA) differ from the analyses that have been performed so far in intelligence? In the absence of a big-data concept or capability, most of the research activity consisted of link analysis - that is, research on links that specifically investigates anchors and the nodes around them. However, SNA tries to look at the network as broadly as possible and, based on an algorithm, to identify the (data-driven) centers of gravity without relying on a researcher's intuition. Such an approach facilitates a "view from afar" of the research object and a broadening of the research object if necessary (communities of people rather than separate individuals or organizations). Network analysis also makes it possible to arrive at insights that are not held even by the research object itself. We are not always aware of the bottlenecks in the processes we are conducting, formal or informal, and we do not know the centers of gravity of an organization of which we are members at any given moment.

SNA tries to look at the network as broadly as possible and, based on an algorithm, to identify the (data-driven) centers of gravity without relying on a researcher's intuition

Big data fostered a significant advance in network research. On the one hand, the influx of information challenged the researcher's ability to understand the relationships and see a broad picture with the traditional tools; on the other, the computing and algorithmic capabilities that were developed make it possible to analyze enormous quantities of information and offer valuable insights. The structuring of the information as a network enables us to produce insights that our human limitations would not permit. The human imagination is limited and has difficulty containing a complex system. Analyzing the system as a network allows the intelligence officer to "tell a story about the data" - that is, to describe phenomena in a context and in real time.

In traditional analysis, the intelligence officer assumes a scenario and then seeks it in the data. Network analysis, however, offers the opposite capability: to arrange the data as a network and to understand the system by which the researcher is freed of the need to make assumptions in order to understand the network. Thus network analysis hews to data that make it relatively easy to refute or substantiate the intelligence picture, as opposed to a thesis based on intuition.

Fig. 9: A "network analysis" of a soccer game.



A network analysis of a Spain-Netherlands soccer game through documentation of the passes (the thickness of the line represents the quantity of the passes). One can specify the main players (centers of gravity) and which side (right/left) was more dominant (or in the IDF parlance, which actions were possible and which actions were chosen).

Hubs and the 20/80 Rule

Although large networks represent very complex systems, that does not mean they do not operate according to laws. As the research on network theory developed, it turned out that the world of the networks is not random as was commonly thought, and the understanding of these laws facilitates effective research.

Law 1: The power law

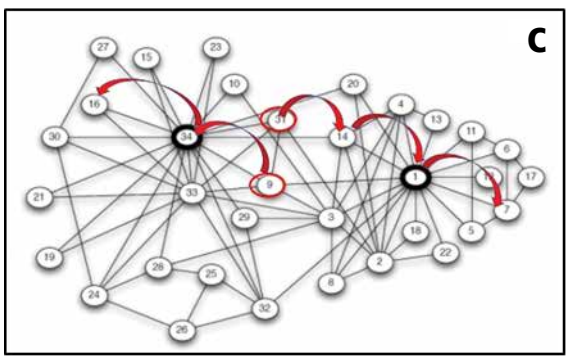
It appears that a few very "popular" nodes "control the network as a whole". Three examples of such nodes on the internet are sites such as Google. In a network of people, these are people who know everyone and everyone knows them. Such connectors are known as "hubs". Hubs are found in almost every network in the world; a low number of people are responsible for a high percentage of the conversations that enter and exit, and so on. Indeed one can posit here Pareto's 80/20 rule regarding the distri-

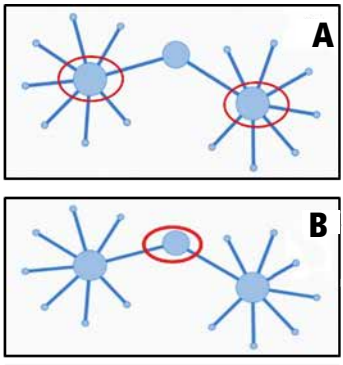
bution of the links: 80% of the links belong to 20% of the nodes in the network (i.e., to the hubs). In the research this phenomenon is called the "power law" or "Zipf's law". By locating these nodes one can identify the centers of gravity of the network, and by comprehending these centers of gravity one can comprehend almost all of the network. Locating these nodes requires the use of SNA algorithmics.

Social Network Analysis - SNA

SNA is the world of algorithmics that makes it possible to construct and investigate the network. There are tens if not hundreds of algorithms for identifying hubs in the network. The main ones, which are in wide use, are:

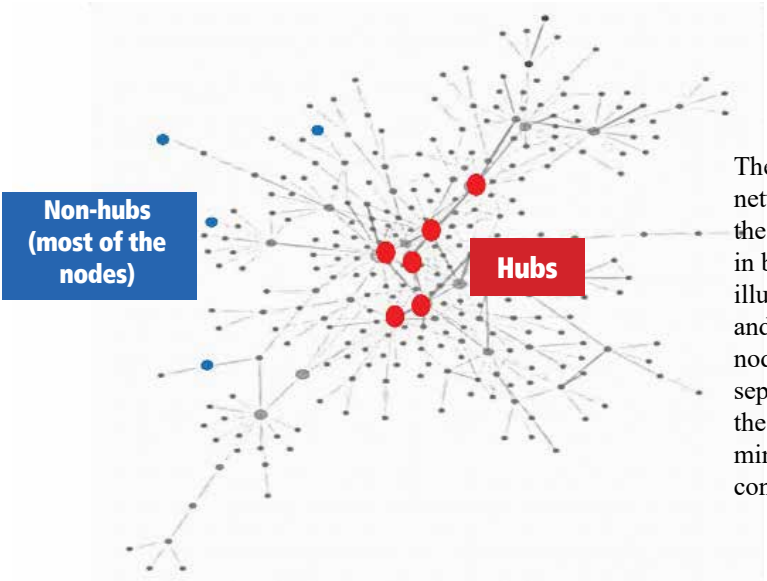
- Degree centrality: Measures the quantity of connections of the node. The logic is that the more the node is connected to entities, the more central it is. A possible example in an organization: the entity that sets the agenda in the organization (operations rooms, bureaus).
- Betweenness centrality: Measures the quantity of the shortest routes in the network that pass through the node. The logic is that the bridging role makes the node central. A possible example in an organization: the entity that connects isolated areas, draws the organizational periphery inward, the source of the entry of new ideas and creativity.
- Closeness centrality: Measures the node's distance from the other nodes in the network. The logic is that the node that is in the "center" is central even if it is not connected to numerous entities. A possible example in an organization: the location makes the actor dominant because of its relative closeness to other actors.





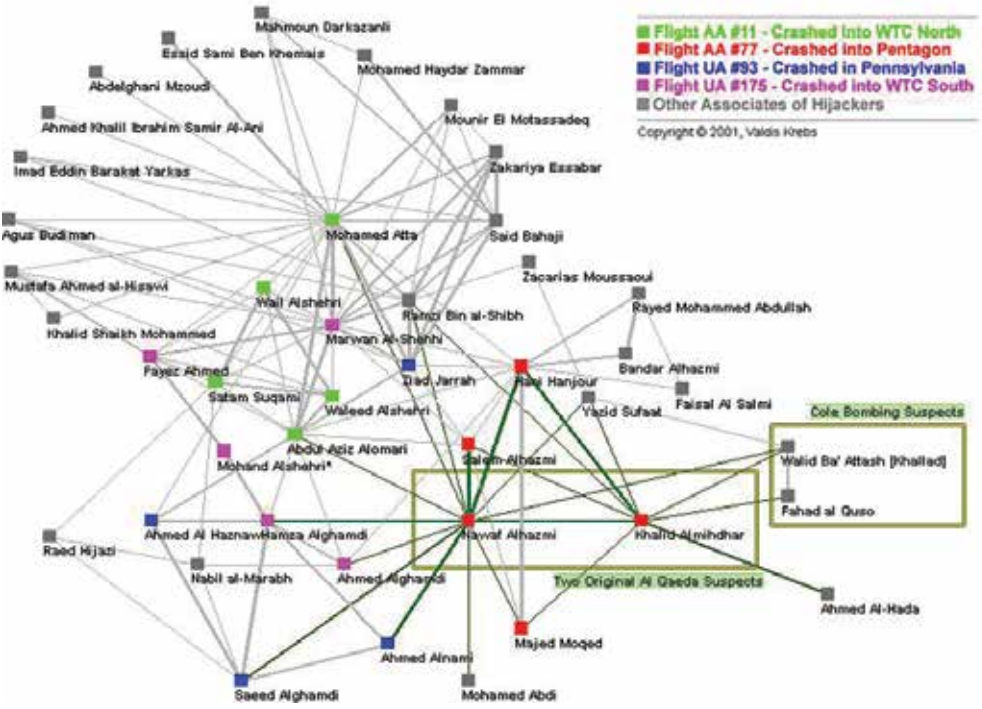
In the above examples: The red circle indicates the most significant measures of centrality: (a) degree, (b) betweenness, (c) closeness. The red arrows illustrate the distance between the central nodes (node 31 and node 9) and the rest of the network.

Fig. 10: Network phenomena.



The main elements of the network are in red and the rest of the nodes are in blue. The example also illustrates the power law and shows that most of the nodes in the network have separate connections while the hubs (which constitute a minority) have most of the connections.

Fig. 11: Network research of the terror network of Al Qaeda in the attacks of September 11, 2001.



When researching an organization by means of network analysis, it is not enough to scrutinize the central actors because errors could occur. An example is American research that was done on the Twitter network in Egypt to identify the main elements behind the "Arab Spring" revolution of 2011. The most central elements were found to be Justin Bieber and Katy Perry, whose contribution to popular music is well known but not necessarily to understanding the power factors behind the events...

To identify the centers of gravity that interest us, the researchers, we will need Law 2 of the network: *congregating networks.*

A closer look at the network's internal structure reveals that the network is not random but built of clusters. These clusters have a logic of homophilia (the tendency to connect with someone similar to oneself), but, at the same time, profuse internal interactions. Such a division of an organization into communities can reveal the organization's "real" structure when it is under scrutiny.

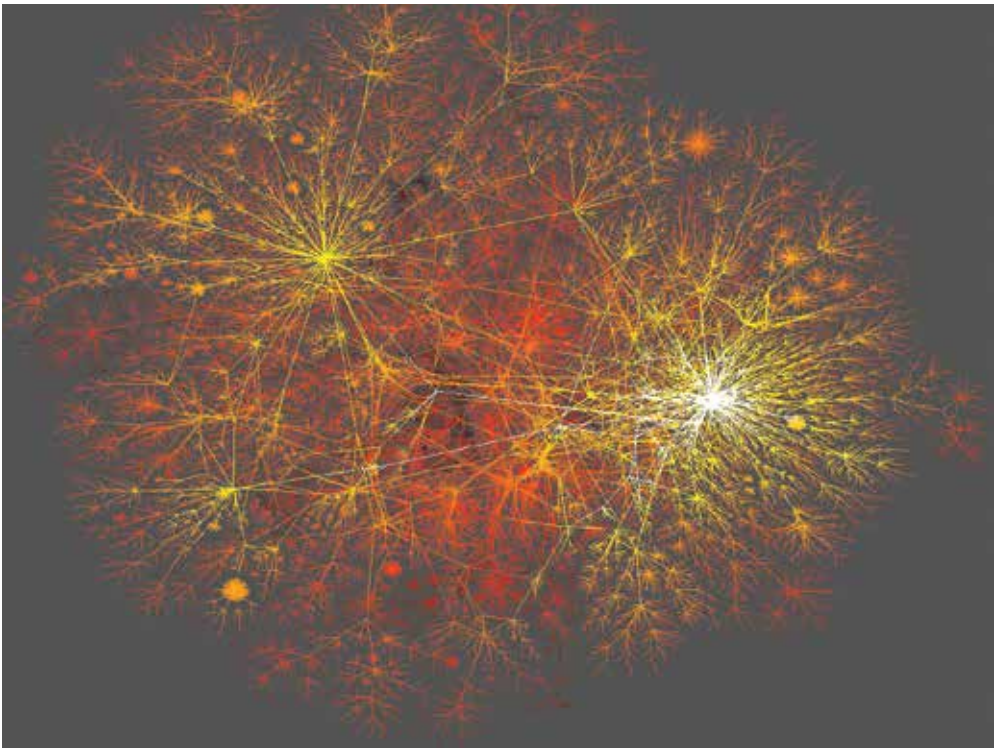
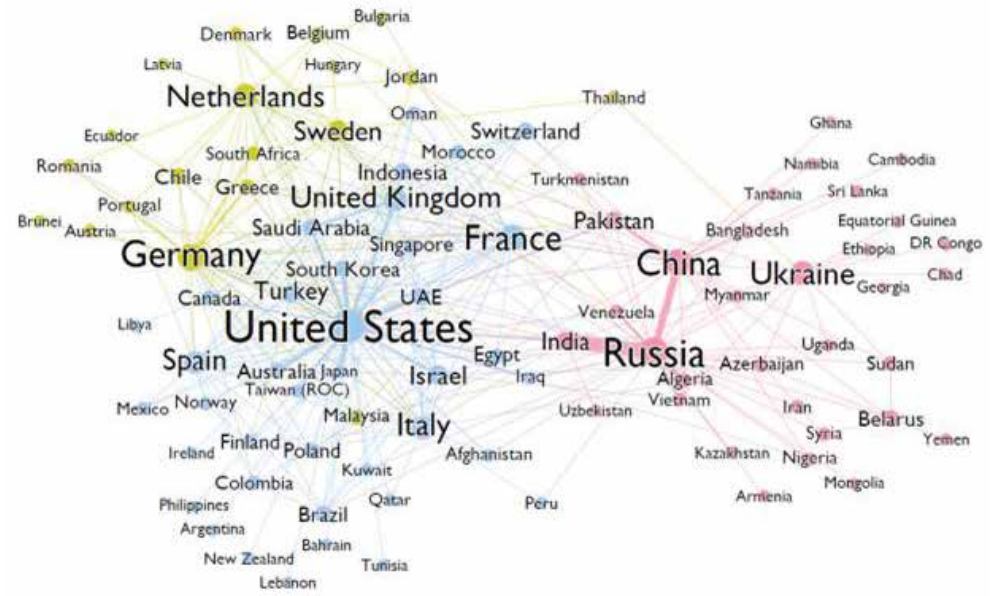


Illustration of an internet network in which one can spot the branching-out from the core of the network, which creates branches and sub-branches. Each such branch is a community that contains subcommunities.

Mapping the communities and finding the key actors in the communities of interest makes it possible to locate the organizational centers of gravity. In the example of Tahrir Square, the main actors that drove the revolution would likely be found in the community we would brand as the "Tahrir community", and the pop singers would likely be in a different community that we would brand as "music lovers". In this domain as well there are tens if not hundreds of algorithms with different logics for the grouping of the communities.

Today there is no "scientific" definition for the question "What is a community?" The most common definition is that a community is a group of nodes in a network for which the quantity of edges between them is denser than the quantity of edges for other nodes in the network.

Fig. 12: Network research on arms deals.



SNA research that is conducted on arms deals .The illustration shows the trading communities and the centers of gravity in each community .This research ,for example ,reveals three central communities :the Eastern one) Russia/China ,(the European one) Germany ,Britain ,(and the American community ,to which Israel belongs.

Dynamic Research

SNA research can be conducted on a specific period of time ("snapshot"). Although this is now the widespread method, in recent years network research has dealt increasingly with dynamic analysis of the network over time. This field also offers a basis for detecting changes (or in the intelligence language, alerts and warnings).

What Are the Relevant Tools for Detecting Changes?

In-depth network research indicates that in many networks there is no stability at the individual level. Nodes appear and disappear and it is hard to base a routine on them. However, communities in the network tend to be more stable (even if the nodes within them tend to change). For example, it would be difficult to characterize the routine of a commander effectively, but the "battalion community" is relatively stable. In addition, any significant change in an organization/network will probably cause an irregularity for more than a single individual. That is, identifying an irregularity of a single individual does not necessarily indicate an event. A change in the behavior of the community, however, does constitute an event (a "telltale sign") that requires attention.

Thus, for identifying an irregularity in particular and for intelligence in general, the advantages of SNA are:

- An ability to analyze macro areas (a community, a network) and not just micro areas (the separate individual, a node).
- Observing "through the eyes of the data" with no need to deal in hypotheses about scenarios, which naturally are biased or limited by the human imagination. The point of departure is that the past does not necessarily tell us the future. Hence:
  - Unlike with other methods, there is no need to make prior assumptions about the data.
  - There is no need to "train" (certainly not substantially) the algorithm about the data.

To sum up, employing SNA (algorithmics of the network) enables intelligence to identify centers of gravity and communities of interest. The big-data revolution offers an infrastructure and an algorithmics for analyzing millions of data in a short time, so that one does not necessarily have to mark anchors of interest that are based on the traditional research and begin the research with them (small data by a method of scenarios). One can, instead, analyze the entire network all at once and remain cognizant of it over time.

Therefore one can discover:

- Who are the dominant factors in real time (with no need for prior knowledge) for purposes of targeting or surveillance.
- What new and unfamiliar factors have "suddenly popped up".
- How the enemy organization "really" operates (not according to the official structure tree but according to the interaction in the network), so that one can identify a possible course of action, a situation picture of the enemy, and so on.

"Angels in the Skies of Berlin": New Intelligence Questions in a World Steeped in Data

M. – deputy director of the School for Intelligence of the Israeli Security Agency

A cornerstone of modern intelligence thinking is the notion of a *modus operandi*. Intelligence organizations invest many efforts and resources in researching the modus operandi of enemies, in identifying it and characterizing it. The main goal is to achieve an information-collection capability regarding the enemy's activity and to monitor that activity in a way that makes it possible to detect telltale signs of elements of this modus operandi, and thereby to identify the enemy's intentions and counteract them.

"Ask What, Don't Ask Why"

SIGINT, which has developed over the past hundred years, also operated according to this intelligence conceptual framework. Intelligence organizations monitored and collected information about the communication activity of their adversaries, analyzed and investigated how their adversaries' modus operandi revealed itself in the communication signatures of their activity, and devised mechanisms to identify the telltale signs within the communication activity. With the development of the communication world over the past 15 years, intelligence organizations have had to deal with a growing quantity of signals from numerous and diverse implements and with different kinds of digital signatures. Intelligence organizations have managed to adopt and develop

technologies to analyze links and characterize communication behavior, thereby expanding their ability to identify and monitor their adversaries' activity and discern telltale signs of it. Thus the main trend is still to conduct researches and collect data that make it possible to depict the adversary's modus operandi, detect its telltale signs, reanalyze it, and so on, thus investing great organizational and intellectual energy in specifying the modus operandi of the foe.

The intelligence world can be likened to the commercial and advertising world. The classic intelligence approach resembles the advertising methods of research and focus in the second half of the 20th century, when huge resources - financial, organizational, and intellectual - were invested in market research and analysis of the preferences and tastes of potential consumers, with the aim of devising effective advertising that would lead to consumption.

In addition to the changes that have occurred in the technology and paradigms for analyzing human activity, the intelligence organizations' adversaries have undergone a change as well. If in the past the intelligence bodies dealt with adversaries that operated as organizations, whether of states or terror and crime organizations, in recent years intelligence organizations have shifted much of their attention to networked and nonhierarchical adversaries. These entities do not forge connections on a basis of predetermined organizational structures, and they generate activity through inspiration and the spreading of messages rather than direct guidance. This is a "flat world" of adversaries, and it sometimes requires dealing with individuals who constitute threat and risk factors in themselves and operate independently. These changes in the nature of adversaries require intelligence organizations to alter their basic approaches and, instead of seeking the adversaries' modus operandi, to look for telltale signs of other kinds. Such signs could be a change in behavior, a change in appearance, a rise or decline in the volume of activity, the forging of new connections, contact with hubs of networks, and so on. Intelligence agencies have been forced to change their approach to collection, collecting very numerous data while deploying relevant sensors to gather and make use of them, and thus also changing the kinds of questions they ask about the information that is collected.

The big-data world was born from the convergence of several technological developments: an enhanced ability to produce and collect a large quantity of information with powerful and diverse sensors, an increase in the volume of data storage and in its miniaturization (alongside its transfer to clouds), a greater quantity of information in the world because of the growing use of technologies that generate digital signatures in daily human activity, and an upsurge in the power of computation that makes it possible to deal with enormous quantities of information and to analyze them in a short time span and simultaneously.

The big-data world fosters a change in patterns of thinking that is not only quantitative but also qualitative and paradigm-changing. This is a world that generates question, researches, and business opportunities of a new order. In their book Big Data, Mayer-Schönberger and Cukier describe the main paradigmatic changes that the big-data world produces among the prevailing modes of thought.[22] A key principle that they posit is "Ask what, don't ask why" - meaning that, in the world of big data, there is no reason and no need to attempt to research and depict the research object's model of activity; what works instead is to use data-based prediction through algorithms that identify correlations and not necessarily dependencies. In other words, even if we cannot explain the research object's model of activity and cannot prove that a certain phenomenon stems from its milieu, it is enough for the algorithm to reveal a correlation between the two phenomena so that we can make effective use of this relationship.

The giants of online retailing with Amazon at the forefront, as well as all the arenas of online commerce, use the enormous information they have about their customers' purchases to come up with sales pitches targeting consumer behavior, even in places where the customer himself is not aware of the connection between his purchases, the products he has viewed, the time he has spent at each product page, the extent of his interest in the pictures or features of the product, and so on, and the sales pitches that the site gives him. Sometimes we can explain the connection and the correlation between different consumer behaviors. For example, a person who starts taking an interest in car seats at the newborn size will probably be interested in buying diapers for a baby and is worth targeting with advertisements for products of that kind.

Intelligence Questions in the Big-Data Era

Today's intelligence organizations make extensive, thorough, and effective use of big-data technologies and methods in several fields of interest and regarding several major questions. They adopt the methods for data collection, analysis, and use to arrive at focused information on the adversary's actions and intentions, thereby "finding a needle in a haystack". Intelligence organizations identify both threats and opportunities posed by their adversaries in the cyber domain, and in this domain there is naturally much use of methods from the big-data world. Among other things, the organizations use such methods in analyzing an immense traffic of data, in trying to detect anomalies that indicate the potential for a cyber-attack, in scrutinizing network behavior patterns that may offer telltale signs of cyber activity, and in identifying adversaries' weak points.

22 Viktor Mayer-Schönberger and Kenneth Cukier, Big Data: A Revolution That Will Transform How We Live, Work, and Think (Boston and New York, 2013).

Beyond the cyber domain itself as an arena where organizations and their adversaries contend with and collect information on each other, cyber is also a far-reaching realm for collecting huge amounts of information about the activity of individuals and groups and the forms of communication between them. Many intelligence organizations deal with this realm, using similar methods to those used in the business-civilian world for commercial and marketing purposes. If business organizations want to know the extent of interest in a certain product, or to identify a need that can be satisfied through focused marketing, intelligence organizations monitor network activity and discourse to identify the extent of interest and activity in a field that constitutes a potential threat. Intelligence organizations also look for opportunities related to their areas of concern, and within the sea of information they also try to identify those individuals and groups who hold a potential for threat and damage. In addition, they look for opportunities to improve their own collection and attack capabilities.

Intelligence organizations are naturally interested in questions that also occupy business organizations, only the focus of questions being different. A business organization asks itself, "What does the public think of my product?" "What is my product's status vis-à-vis competing products?" "How can I identify a new and emerging need and how can I enter the market at the point where my product will answer the need?" Business organizations seek answers to these questions using methods of big-data analysis, from analyzing the sales and consumption data in their and their partners' databases to monitoring the network discourse to using techniques of opinion mining, brand monitoring, and sentiment analysis. Many companies provide such services to companies and corporations for a wide range of products.

The difference between the business world and the intelligence world: whereas the focus of interest of a corporation or a company will be the product they are producing or the idea they are trying to market, the intelligence organization will be interested in the ideological or psychological motive

Intelligence organizations also make use of such methods and ask questions of a similar kind, the only difference lying in the focus of interest. Whereas the focus of interest for a corporation or a company is the product they are producing or the idea they are trying to market, an intelligence organization will be interested in the ideological/religious/psychological motif that could lead to terror activity, in certain weapons, or in any other issue within its field of concern.

Intelligence Questions of a New Kind

At the same time, there is a considerable area in which intelligence organizations have a built-in advantage over business corporations and companies, an advantage that can generate new and creative intelligence and research opportunities, allow intelligence organizations to ask new research questions of a different kind, and broaden the picture of their world along with the basis of information that will enable the decision-makers to manage risks and conduct better decision-making processes. In an effort to stimulate thinking about new intelligence questions that can be answered with big-data tools and methods, I will describe an imaginary reality that combines a historical sociopolitical dimension with a contemporary technological dimension.

Let us put ourselves in a fictitious reality-setting. We are in the year 1983, in the midst of the Cold War, a geopolitical situation polarized between the Eastern bloc led by the Soviet Union and the West led by the United States. We are in a Berlin that is divided between east and west. The eastern part of the city belongs to the German Democratic[23] Republic or DDR (East Germany), while the western part is an enclave belonging to the Federal Republic of Germany or BRD (West Germany). At the heart of the city is a wall whose main passageway between the east and the west is Checkpoint Charlie. In the two sides of the city, two opposing intelligence organizations operate. On the western side, among others, is the BfV, the domestic intelligence service that deals with threats of ideological subversion (such as neo-Nazism and communism) and with thwarting espionage by the Eastern bloc. On the other side of the city is the Stasi, which is both a foreign intelligence agency and a domestic intelligence and security agency responsible for monitoring and spying on the population.

In our imaginary reality, the technology and the nature of communication belong to the year 2018. The citizens and residents of the two countries make use of smart cellular phones, and the large majority of them have email accounts and active profiles in the different social networks. These citizens and residents use devices that are connected to the internet and produce different digital signatures, such as smart TVs, smart watches, fitness wristbands, smart cars, digital and biometric identification tags, and so on. In this fictitious reality, both of the intelligence organizations have unlimited access to all the data and the signals generated in the city of Berlin and in the border areas between the two countries.

Now, having described the imaginary reality, we will put ourselves in the shoes of the research personnel of the two opposing intelligence organizations and try to think about the new questions we can ask about our research objects. Such questions

23 The mood of the period can be felt by reading some books of different kinds and watching some recent movies and TV series. Recommended, among others, are David Young's Stasi Child (Minotaur Books, 2017), the movie Das Leben der Anderen, and the series Deutschland 83.

are intended to enrich our intelligence understanding, broaden our capabilities, and thereby fulfill these organizations' mission.

The head of the Stasi's Department for Protecting the Country against Imperialist Propaganda and Subversion has assembled his workers and asked for research on the leisure culture of the residents of East Berlin, categorized by age and employment; in particular, he wants a more precise analysis of the leisure activity of government employees. The Stasi's algorithmics personnel have gone to work and constructed a research that analyzes the characteristics of this leisure activity using methods of big-data analysis. This entails analyzing all the data on the location of the cellular phones that belong to or serve the residents of the eastern part of the city, while segmenting the length of the days of the week and the hours of the day so as to reveal the graph of change between hours spent at work and hours and days of leisure activity. The analysis includes a comparative inquiry into the locations of cellular phones and the activity of their owners on the social networks, with analysis of the contents they display on the networks. That, in turn, includes pictures, posts, and texts, distinguishing between words that typify leisure activity and words that typify domestic activity.

Another analysis by the Stasi's research division deals with the correlation between the citizens' social contacts - that is, with which people they are in contact - and the hours at which they conduct their communications with these people. The aim is to find out whether citizens tend to spend leisure time together with their family members, with their friends from work, or perhaps with other sorts of people who distance them from their friends from work.

In light of the findings of the Stasi's research, the State Security Service has made changes in its surveillance methods. It was decided to recruit new informers in unfamiliar leisure locations that turned out to be popular during the weekends among citizens of the eastern city aged 25-35. At the same time, an all-clear signal was heard in the agency's corridors when it turned out that the state workers tend to stay in their homes even during leisure hours and are not swept into the sorts of leisure activity that are uncontrolled by the state. The research department arrived at this conclusion because the state workers' smartphones tended to remain in proximity to other devices connected to the accounts of the phones' owners such as their smart TVs and PCs.

On the other side of the wall, the head of Berlin's BfV department asked his algorithmists to build a data-based intelligence picture that would analyze all the East Germans who had succeeded or tried to cross the border during the past three years. The department chief requested that the research provide an answer to two questions: First, is defection from east to west a way in which East German intelligence infiltrates spies into the west? Second, can potential candidates for defection be identified beforehand so as to encourage them to defect and build platforms that will enable

them to reach the west?

The research conclusions of the West German service were that the media consumption of defectors and of those who tried to defect included much more browsing of Western sites, as seen in an ascending graph of such consumption, while they were less exposed to official material of the East German regime. The research personnel also discovered a point on the time axis of media consumption at which the candidate for defection begins to show geographical proximities to the wall and to the border with West Germany. Based on the findings of this research, the technology personnel of the West German service devised a search mechanism for all the digital media consumption in the east that could identify users who produced a similar graph; on that basis they created an alert for the axes' convergence with the appropriate point, along with an automatic mechanism for sending a text message to all who appeared in this alert. The message was conveyed through clandestine media in an effort to help potential defectors. Meanwhile each potential defector's contacts with East German intelligence were surveyed so as to verify that he was not an East German spy intended to infiltrate the West.

What Can We Learn from This?

So much for our imaginary world; each reader can take his intelligence world and try to apply these notions to contemporary reality. In my view, one only needs to use imagination and intelligence creativity in order to construct questions of a new kind, of a kind that have not been asked until today, questions to which the big-data world enables us to provide valuable answers that can enrich knowledge about adversaries and increase intelligence agencies' ability to offer tools to the decision-makers. Such tools can help the decision-makers both manage the intelligence establishment and manage risks, while assessing threats and opportunities in a way that better reflects the reality.

The key to generating relevant researches with big-data methods and tools in the intelligence world is the awareness of the possibility of asking new questions; the understanding that big data not only create a quantitative difference that allows answering old questions with new tools, but also a new reality in which one can ask completely new questions, the response to which will give intelligence personnel a more accurate picture of the adversary and the environment in which he operates. I suggest that the way to use this key is by forging new collaborations between those engaged in classical intelligence research and those dealing with big data, algorithms, and analysis in areas of information mining, with mutual inspiration generating collaborative ideas.

The Social Networks: What Do They Tell and What Do They Hide?

Maj. D. – serves in Aman

Background

The upheaval that took place in the Middle East at the beginning of the decade highlighted the role of the Arab population as an important political actor in the region. The events proved that this population can act politically even in countries with an authoritarian regime; it can spark revolutions and protests and influence decision-making processes of the political echelon. This has fostered the realization that social research is an essential aspect of understanding and formulating the intelligence picture, and that one cannot grasp the complex reality without insights into society and its attributes (the ethnic-religious component, the economic situation, sentiments toward the regime, etc).

As pointed out by Brig. Gen. Itai Brun, former head of the Research Division of Aman:

Another research challenge that the "upheaval" posed was the need to better understand the populations of the Middle East... The "upheaval" showed the increased weight of the populations in the streets and squares, but primarily in the awareness of the decision-making leaders; the populations that took to the streets in 2011 and brought about the removal of the leaders in some countries impelled the leaders in other countries to undertake economic and social reforms.[24]

The realization of the importance of population research raised questions about how to go about studying the population in countries with a nondemocratic regime and whether the traditional research methods - primarily opinion surveys[25] - are suitable for monitoring public opinion in the various countries. The study and monitoring of publications of research institutes that track public opinion in the Arab world indicated that as recently as 2017, opinion surveys were the most common tool for observing and monitoring public-opinion trends all over the world including the Middle Eastern countries.[26]

---

24 Itai Brun, "Clarifying Reality in an Era of Transformations and Changes". Intelligence and Terrorism Information Center, 2015, 33-34 (Hebrew).

25 In this article the term "opinion surveys" refers to surveys whose aim is to set forth an intelligence picture of public opinion at the time the survey is conducted. The reference is not to election polls, which are a subspecialization of the survey profession and are aimed at predicting future behavior.

26 For example, the Arab Youth Survey, the Arab Barometer, the World Value Survey, the Arab Index, and others.

Although the great advantage of opinion surveys is their ability to accurately represent the entire population that is the research object, this approach is not without limitations. A principal one stems from the data-collection method. In opinion surveys the rationale for data collection is the asking of questions. The researcher asks the questions and the interviewees respond only to what was asked according to how they understood it. This method could lead to biases related to the wording of the question, how it is presented, and its location in the questionnaire, and to biases stemming from the respondents' "social forethought", from the interactions between the interviewer and the respondents, and so on.

In light of these methodological limitations, recent years have seen increased attempts to conduct research on opinions, perceptions, and moods of populations by analyzing the discourse on the social networks. More and more claims are heard that studies analyzing the discourse on these networks are better than other research methods (such as opinion surveys) at representing the perceptions, opinions, attitudes, and moods of entire populations of countries. One major reason for the growth of this kind of research is the large quantity of information that can be analyzed. In this article I will focus on what can and cannot be concluded from studies of the discourse in the social networks, with an emphasis on the Arab world.

Researches That Analyze the Discourse in the Social Networks

Studies focusing on the discourse in the social networks are based on analysis of the very large quantities of information to be found in these networks. The researches are conducted in different ways: using off-the-shelf softwares to monitor the discourse in these networks, or independently through "free" browsing of the internet both as a "passive" observer and actively with "involvement" in the research domain. Unlike opinion surveys, the rationale for the collection method in researches of the discourse in the social networks is observation - a more "anthropological" method in which the researcher, for the most part, is an observer of the research domain and does not intervene in it by asking the research subjects questions. Ostensibly this method is more neutral and entails fewer biases, thus enabling a more authentic discourse.

The drawing of conclusions about what takes place in the social networks ignores factors that impair the authenticity of the discourse and could lead to mistaken notions

However, drawing conclusions about entire populations from such studies is problematic because it ignores many factors that can impair the authenticity of the discourse in the social networks and hence also the abil-

ity to derive insights and conclusions about public opinion from it. Currently the researchers who analyze the discourse in these networks face two main challenges. The first concerns the characteristics of the users of these networks; the second concerns the authenticity and neutrality of the public discourse in them. Not infrequently these challenges go hand in hand and raise serious questions about whether the findings of these studies accurately represent opinions and perceptions and whether they could lead to biased and mistaken notions.

Contending with these challenges requires examining the characteristics of the browsers of the social networks and considering how representative these characteristics are of the population as a whole. First the term "representative" needs to be clarified. A "representative study" is one that seeks to represent opinions of large populations and thereby draw conclusions about such populations. It must, therefore, be constructed from a sample that is identical in its characteristics (the background variables) to those of the population being researched. Otherwise the study is not representative and one cannot draw general or comprehensive conclusions from it about the population it is supposed to represent.

Are the Characteristics of the Social-Network Browsers Indeed Representative of the Population as a Whole?

Studies of the participation of populations in the social networks show that the characteristics of the users of these networks are not identical to those of the population as a whole. This finding is not dependent on the country or on the nature of its regime. For example, a study at Oxford University[27] found that Twitter users in Britain and the United states are younger and better off socioeconomically than internet users as a whole, who are also younger and of higher socioeconomic status than the general population. In other words, Twitter users are not representative of all internet users and internet users are not representative of the entire population. The author of the article also notes that Twitter is more popular among social elites. Based on the penetration data, which are relatively low,[28] for this network in the Arab world, we can assume that the situation in Arab countries[29] at least is not essentially different. Specifically, a 2016 survey of 22 Arab countries found that, among the social-network browsers, 81% lived in cities, 64% were under

27 G. Blank, "The digital divide among Twitter users and its implications for social science". Social Science Computer Review (2016).

28 Eleven million users in the Arab world as of 2017, according to the Arab Social Media Report, 7th ed.: "Social Media and the Internet of Things: Towards Data-Driven Policymaking in the Arab World: Potential, Limits and Concerns". Mohammed bin Rashid School of Government, 2017.

29 Ibid.

30 years old, 67% were men, and 80% had academic educations.

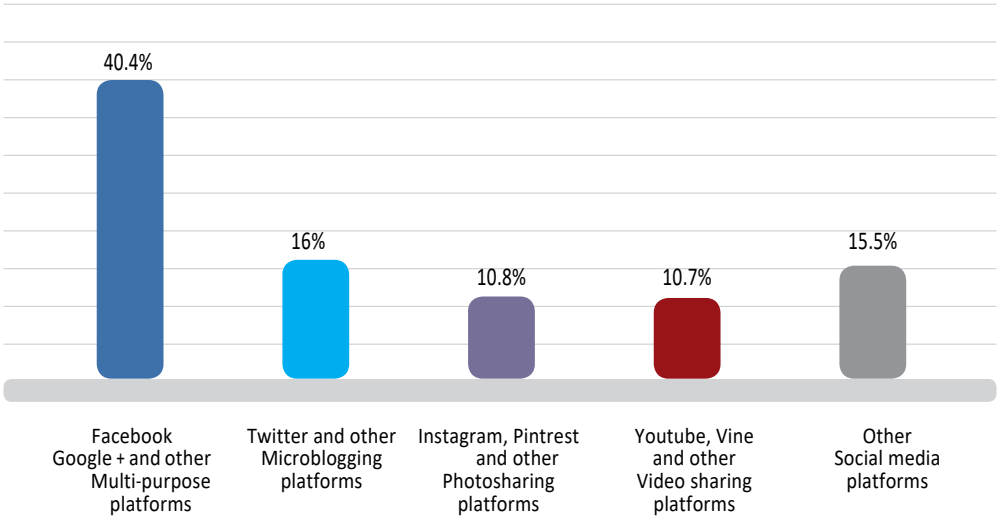
The above data make clear that the characteristics of the network browsers are not identical, or even similar, to those of the general population. What this means for research is that we cannot claim that the moods, opinions, and perceptions manifested in the public discourse in the social networks indeed represent the public opinion of populations as a whole. Moreover, we do not know on which measures the social-network browsers differ from the general population. Thus, in contrast, for example, to opinion surveys in which statistical manipulation allows one to correlate the weight of a certain group in the sample with its weight in the population so that the sample will be representative, when analyzing the discourse in the social networks we do not have enough information about the users' characteristics to be able to do so.

Another factor that hampers the ability to construct a "representative sample" of the discourse in the social networks is that some of the users, out of concern for their privacy, choose to provide fictitious details in their social-network accounts

Another factor that hampers the ability to construct a "represe
tative sample" of the discourse in the social networks is that some of the users, out of concern for their privacy, choose to provide fictitious details in their social-network accounts. A report on the social media in the Arab world, for example, found that 15% of the browsers provide false information in the social networks. Most commonly lied about is the name; 61% of those who provide false information do so about their name. More relevant to the question of representativeness, however, is that 40% of the browsers who gave false information said they lied about their age while 35%
lied about their geographic location. In addition, a third of the browsers turn off the device for pinpointing their geographical location when they use social media. Naturally, and consistently with the degree of penetration, a majority of the browsers who provide false information do so in the Facebook, Google, and Twitter networks. What these findings entail is that, regarding age and geographical distribution in a country, we are unable to ascertain that the characteristics of the sample are identical to those of the general population, and hence we cannot assume that the research is representative.

A third variable that also affects the representational validity of the discourse in the social networks is the fact that in each network many of the browsers have several accounts. For example, a survey conducted by the Mohammed bin Rashid School of

Fig. 13: Where false information is reported in the social networks.



Government in Dubai found that 46% of the social-network users in the Arab world have a large number of accounts on at least one social-media platform. The fact that a considerable portion of the browsers have several accounts on a single social-media platform causes an overrepresentation of their opinions compared to the opinions of those with only one account. In other words, the discourse is likely to be biased toward their opinions and hence does not necessarily represent the discourse in "reality". Similarly, an article posted on the blog of the Oxford University website[30] noted that 40% of those with a Twitter account have never tweeted and 15% of the users are responsible for 85% of the tweets. Although the study deals with the American population, presumably the phenomenon is not unique to the United States.

Another factor that impairs the representational capacity of the public discourse in the social network Facebook, which has the highest penetration in the Arab world, is that the company restricts the ability to monitor the discourse; only public accounts can actually be tracked. That is, users whose account is protected by privacy settings are not exposed to the monitoring capability. Clearly this greatly reduces the ability to represent extensive populations using social-network-based research. Furthermore, experience indicates that in many cases public accounts belong to state communication bodies and to actors that want to influence the public discourse such as politicians, cultural figures, and so on.

All these findings underline the fact that research on the social-network discourse

30 "Did you consider Twitter's (lack of) representativeness before doing that predictive study?"

does not constitute research that optimally represents public opinion for all of a country's population.

The Degree of Authenticity and Neutrality of the Social-Network Discourse

The second challenge that significantly affects the ability to draw conclusions from research on the social-network discourse concerns the degree of authenticity of the opinions and perceptions that are manifested in it. As noted earlier, the rationale in such research is that a free and authentic discourse is conducted in these networks and that they constitute a platform that enables such a discourse. However, recent data present a totally different picture. The report "Freedom on the NET" of the Freedom House research institute,[31] published in November 2017, offers a gloomy appraisal of the degree of freedom, anonymity, and authenticity in the social-network discourse. The report says that the involvement of regimes, and attempts at influencing the online discourse, have grown more and more widespread and sophisticated, the aim being to suppress opposition voices and promote an antidemocratic order.

According to the report's authors, such intervention is carried out mostly through manipulations such as creating fake news, using bots and trolls, breaking into accounts so as to monitor citizens, blocking sites, and so on. However, in addition to such technological manipulations, the report notes that physical attacks on regime opponents, to the point of imprisonment and even murder, are increasingly common. When it comes to the freedom to browse the internet in Middle Eastern countries, the report presents a picture where in most of these countries such freedom is limited or nonexistent.

According to a report for 2017, "Media Use in the Middle East",[32] 41% of the browsers in the countries that were examined[33] (compared to 36% in 2013) said they were concerned that the authorities were monitoring their activity on the internet. Likewise, 20% of the respondents (30% in Jordan) said they had changed their behavior in the networks out of fear of surveillance by the authorities. The common strategies were changing privacy settings, writing less critically or less frequently, and having contact with fewer friends.

The report published by the Mohammed bin Rashid School of Government in Dubai also found that 36% of internet browsers in the Arab world choose not to express negative opinions about the authorities in the social media. In addition, only 29% feel

31 "Freedom on the Net 2017: Manipulating Social Media to Undermine Democracy". Freedom House.

32 Northwestern University in Qatar.

33 The countries chosen were Lebanon, Tunisia, Saudi Arabia, the United Arab Emirates, and Jordan.

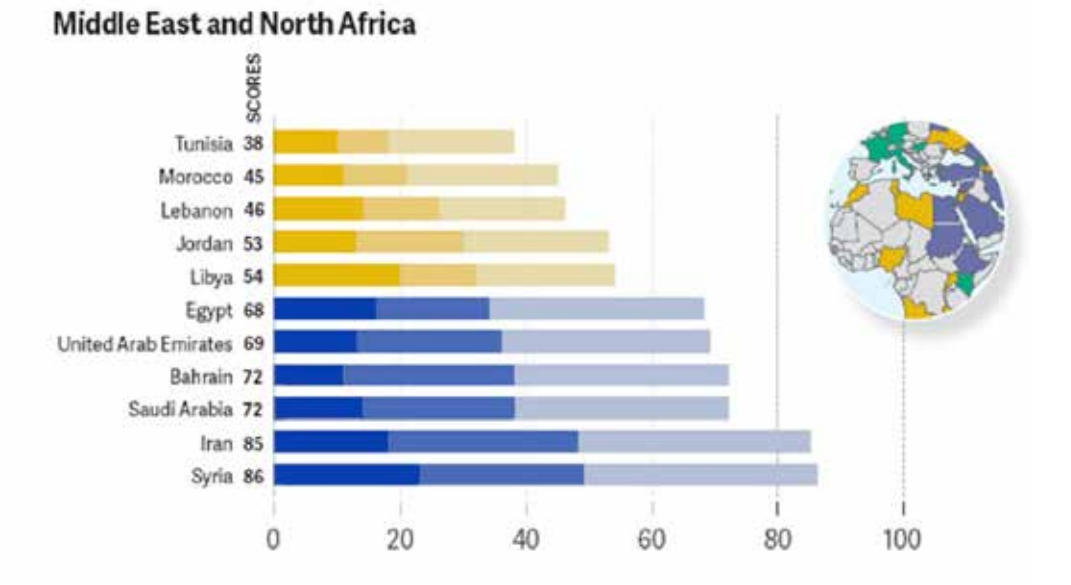


Fig. 14:
Blue indicates the lack of freedom to browse the internet, yellow indicates partial freedom, and green indicates freedom to browse. The higher the grade, the lower the freedom to browse the internet; a grade of 100 means a total lack of freedom and a grade of 0 means total freedom.

that they can freely and candidly express their negative opinions about the ruling establishment; one-quarter of the respondents said they employed self-censorship when expressing such opinions; 21% said they chose not to express themselves negatively; and 11% said that they hinted at their opinions and did not present them directly. Moreover, 11% said they used sarcasm in the social networks. These data highlight how problematic it is to rely on the social media as a source for the analysis of public opinion, public sentiment, and societal trends. The authenticity of the public discourse in the social networks is also compromised by the phenomenon of writing "for someone". In the Arab world, as in the world in general, there is a widespread phenomenon of employing browsers on the network to promote agendas and specific perceptions. The expression of opinions by these browsers biases the discourse such that it does not represent the "reality". What these data basically indicate is that the authenticity of the discourse on the internet in general, and in the social networks in particular, diminishes over time.

Conclusion

Based on the foregoing, I maintain that findings of studies of the public discourse in the social networks cannot be regarded as fully representative of populations of countries and indeed can lead to biased and mistaken interpretations and intelligence analysis.

Does this mean we should stop using studies of the social-network discourse to gauge perceptions and moods of various populations? The answer is unequivocal: no! We must keep developing the ability to conduct studies of the social networks and exploiting the advantages of such studies. In my view, their main advantage lies in the enormous quantity of information that they provide. Hence it is accurate to refer to such research as "qualitative research in great quantities" (quantities that generate quality). Such studies can greatly enhance the researchers' ability to understand the subjects of interest and to identify the foci, main ideas, and context of the discourse in question. Still another advantage of such researches is that they provide a "close-up" view of the power factors and public-opinion shapers ("influencers"). Researches of this kind must be used to detect and understand the forms of activity and degree of influence of public-opinion leaders in the social networks; such researches do not require us to represent the population as a whole but only the public-opinion leaders themselves. In addition, studies can be performed in order to observe or map the characteristics and patterns of use of the social-network browsers, or in order to identify populations and frame them in the social networks (for example, youth, students, Facebook groups). Furthermore, the great quantities of information and the technological and methodological developments enable researchers to analyze the structures of contacts between the members of the social networks (social-network analysis); this can greatly contribute to understanding the social structures, power relations, and influence holders among the social-network users.

Treating findings of studies of the public discourse in the social networks as fully representative of populations of countries can lead to biased and mistaken interpretations and intelligence analysis. Nevertheless, such studies should continue to be conducted so as to gauge perceptions and moods of various populations

In addition, the compromised authenticity of the social-network discourse requires researchers to seek the new or different platforms where authentic public discourse is indeed facilitated. Studies of public opinion in the social networks constitute a developing field that, on the one hand, does not lack methodological challenges, and

on the other, harbors many opportunities for research and intelligence. We must consider the use of this methodology critically, recognize its limitations, and exploit its advantages to produce a broader, more detailed, more in-depth intelligence picture of the reality in question. We must also combine these research methods with the classic research methods. Mixing methods rather than restricting them will facilitate understanding, validation, and the drawing of conclusions.

CROSINT (Crowdsourced Intelligence): Using the Wisdom of the Masses for Intelligence Purposes

Dr. Shai Hershkovitz[34]

"The Torah is only learned in a group". (Masechet Brachot 63:72)

Introduction

At the beginning of February 2017 the research and development agency of the American intelligence community (known as the Intelligence Advanced Research Projects Activity - IARPA) announced the launching of Project Create, the aim of which is to improve the community's analytical capabilities and its ability to convey the analytical products to the consumers of the community, and to its partners, by crowdsourcing. Improving[35] analytical capabilities and their theoretical formulation is a long-standing challenge of the American intelligence community, which has established different research frameworks under the heading "Structured Analytic Techniques" (SATs).[36] Although these techniques constitute an analytical "toolbox" for intelligence researchers, by nature they do not encourage the researchers to share and develop common knowledge. These techniques are more accurately defined as "recipes" for conducting research, and they are designed for the individual researcher or a small group of researchers.[37] Project Create is intended to address the limitations of the existing methodologies by developing technological capabilities that support analysis based on principles of sharing and crowdsourcing. The guiding assumption of the project is that content

34 Dr. Shai Hershkovitz is an expert in intelligence theory and research and a strategic consultant to governments and corporations in the world. Today he is head of research for the XPRIZE Foundation, an organization that works to promote futuristic technologies by combining crowdsourcing and competitions.

35 **"IARPA Launches 'CREATE' Program to Improve Reasoning Through Crowdsourcing",** Director of National Intelligence, **February 9, 2017**. https://www.dni.gov/index.php/newsroom/press-releases/item/1735-iarpa-launches-create-program-to-improve-reasoning-through-crowdsourcing

36 See, e.g., a study published by the RAND Corporation in 2016: Artner, Stephen, Girven, Richard S., and Bruce, James B., "Assessing the Value of Structured Analytic Techniques in the U.S. Intelligence Community", RAND Corporation, 2016. https://www.rand.org/content/dam/rand/pubs/research_reports/RR1400/RR1408/RAND_RR1408.pdf

37 For an example of such a theoretical document, see a document published by the CIA:
US Government, "A Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis", Central Intelligence Agency, March 2009. https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/Tradecraft%20Primer-apr09.pdf

experts from various fields, who are not necessarily intelligence workers and not members of the intelligence community, can develop knowledge that will challenge the intelligence researchers and thus forestall a raft of problems inherent to research, such as cognitive biases.

Crowdsourcing for purposes of analysis, research, policy formulation, and idea development (or ideation) has been used outside the intelligence community for over two decades. Over the past two or three years, amid a certain dwindling of the "fashionableness" of the crowdsourcing idea alongside the maturation of the field and the recognition of its limitations, a second generation of crowdsourcing has begun to develop. It is characterized by a combining of advanced technological tools such as artificial intelligence, big data, and second-order analysis of the behavior traits of the crowd (big knowledge) - all this alongside the ongoing development of the requisite knowledge on the motivation and incentivization of crowds.

Crowdsourcing seeks to produce a mix of efficiency and central control, similarly to traditional approaches to knowledge development (research) and strategic planning, along with the benefits that lie in the democratization and decentralization of innovation and creativity

The aim of this article is to suggest the potential of crowdsourcing in the intelligence context to the Israeli reader. The article will survey the use of the crowdsourcing for purposes of collection, processing, and intelligence research; discuss the advantages and disadvantages of crowdsourcing in these contexts; and outline the field's future directions of development in intelligence-practice contexts.

Crowdsourcing: Background

Crowdsourcing refers to a wide variety of situations in which ideas, opinions, or researches are generated by a large group of people.[38] A more precise definition is that crowdsourcing is a model based on information technology (IT) that is designed for problem-solving and idea development and leverages decentralized knowledge existing among groups and individuals to create various resources for organizations.

38 Jeff Howe, "The Rise of Crowdsourcing", Wired Magazine, January 6, 2006. https://www.wired.com/2006/06/crowds and Jeff Howe (2006a), "Crowdsourcing: A Definition", Crowdsourcing: Tracking the Rise of the Amateur (weblog, June 2), URL (accessed November 24, 2006): http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html

In other words, crowdsourcing[39] as an approach to developing knowledge in tandem with information technology - seeks to produce a mix of efficiency and central control, similarly to traditional approaches to knowledge development (research) and strategic planning, along with the benefits that lie in the democratization and decentralization of innovation and creativity.[40]

As Prpic, Taeihagh, and Melton note, the research literature on the subject focuses on three main issues:[41]

- **Virtual labor markets (VLM):** These are information-technology-based markets in whose framework individuals and organizations can agree about performing any work in return for financial compensation.[42]
  The participants take it upon themselves to carry out micro-tasks in return for payment, such as translating documents, designing jobs, programming, filming, transcription, and so on. This mainly involves tasks that still cannot be done well, reliably, and rapidly with computerized means. Examples of these markets may be seen in Amazon's Mechanical Turk project, which involves marketings that millions of people participate in, usually in anonymous form.
- **Open collaboration:** This is activity in which organizations present problems to the general public, sometimes on specific platforms and sometimes through specific pages on social-media sites. The participants are invited to propose solutions and to respond to solutions that other participants have proposed.

39 Daren Brabham, "Crowdsourcing as a Model for Problem Solving: An Introduction and Cases", Convergence 14 (1): 75-90. http://journals.sagepub.com/doi/abs/10.1177/1354856507084420

40 Jeff Howe, Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business (New York: Crown Business, 2008). https://dl.acm.org/citation.cfm?id=1481457&preflayout=flat

41 Prpic, John, Taeihagh, Araz, and Melton, James, "The Fundamentals of Policy Crowdsourcing", Singapore Management University, September 2015. http://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=3117&context=soss_research; Vreede, Triparna D., Nguyen, Cuong, Vreede, Gert-Jan D., Boughzala, Imed, Oh, Onook, and Reiter-Palmon, Roni, "A Theoretical Model of User Engagement in Crowdsourcing", Springer, vol. 8224, pp. 94–109, October 2013 https://link.springer.com/chapter/10.1007/978-3-642-41347-6_8; Estellés-Arolas, Enrique, and González-Ladrón-de-Guevara, Fernando, "Towards an Integrated Crowdsourcing Definition", Sage Journals, Journal of Information Science, March 9, 2012. http://journals.sagepub.com/doi/abs/10.1177/0165551512437638?journalCode=jisb

42 Horton 2010, Horton & Chilton 2010; Irani & Silbernam 2013; Wolfson & Lease 2011. http://john-josephhorton.com/papers/online_labor_markets.pdf; Horton, John J., and Chilton, Lydia B., "The labor economics of paid crowdsourcing", ACM, 2010. https://dl.acm.org/citation.cfm?id=1807376; Irani, Lilly C., and Silberman, M., "Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk", ACM, 2013. https://escholarship.org/uc/item/10c125z3; Wolfson, Stephen M., and Lease, Matthew, "Look before you leap: Legal pitfalls of crowdsourcing", University of Texas, Proceedings of the ASIST Annual Meeting Journal, 2011. https://utexas.influuent.utsystem.edu/en/publications/look-before-you-leap-legal-pitfalls-of-crowdsourcing

The participants do[43] so on a volunteer basis and often do not receive financial recompense for their ideas, only gratitude.

- **Tournament-based collaboration (TBC)** or **idea competition**[44]
  refers to activities in which the organization places its problems on an information-technology-based platform, usually one provided by a third party or, in rarer cases, one that belongs to an organization - for example, Challenge.gov.
  These platforms attract distinctive[45] audiences with knowledge relevant to the kinds of problems that are raised, and the platforms are based on a competition that offers prizes; the winners are those among the crowd who have proposed the idea that the organization has chosen. When this technology is used for purposes of innovation, it is called "open innovation"[46] and involves an attempt to generate ideas alongside problem-solving, which, as noted, are two of the main applications of crowdsourcing.[47]

An additional aspect, or more precisely a subfield of crowdsourcing application, is "online communities of experts" or "online communities of practice". In such communities experts in a certain field collaborate with their colleagues, usually on analytical tasks (such as conducting studies or making predictions), or in response to ideas that were proposed by a general audience of participants.

---

43 Adi, Ana, Erickson, Kristofer, and Lilleker, Darren G., "Elite Tweets: Analyzing the Twitter Communication Patterns of Labour Party Peers in the House of Lords", University of Glasgow, Policy & Internet, Vol. 6, No. 1, 2014. http://eprints.gla.ac.uk/93862/1/Elite_tweets_Adi_Erickson_Lilleker.pdf; Crump, Jeremy, "What Are the Police Doing on Twitter? Social Media, the Police and the Public", Wiley Online Library, December 2011. http://onlinelibrary.wiley.com/doi/10.2202/1944-2866.1130/abstract; Small, Tamara A., "e-Government in the Age of Social Media: An Analysis of the Canadian Government's Use of Twitter", Wiley Online Library, December 2012. http://onlinelibrary.wiley.com/doi/10.1002/poi3.12/abstract

44 Blohm, Ivo, Leimeister, Jan M., and Krcmar, Helmut. "Does collaboration among participants lead to better ideas in IT-based idea competitions? An empirical investigation", 2011, Inderscience Enterprises Ltd., Int. J. Networking and Virtual Organisations, Vol. 9, No. 2.
https://pdfs.semanticscholar.org/4a8b/a79f7b514e8a9c37ef72c819fb49a2402126.pdf; Piller, Frank T., and Walcher, Dominik, "Toolkits for idea competitions: A novel method to integrate users in new product development". Wiley Online Library, May 25, 2006, http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9310.2006.00432.x/full; Schweitzer, Fiona M., Buchinger, Walter, Gassmann, Oliver, and Obrist, Marianna, "Crowdsourcing: Leveraging Innovation through Online Idea Competitions". Research-Technology Management Journal, Vol. 55, No. 3, December 28, 2015. https://www.tandfonline.com/doi/abs/10.5437/08956308X5503055

45 Afuah, Allan, and Tucci, Christopher L., "Crowdsourcing as a Solution to Distant Search". The Academy of Management Review, July 2012.
https://www.researchgate.net/publication/267027676_Crowdsourcing_As_A_Solution_To_Distant_Search

46 Sawhney, Mohanbir, Prandelli, Emanuela, and Verona, Gianmario, "The Power of Innomediation". MIT Sloan Management Review, January 15, 2013, https://sloanreview.mit.edu/article/the-power-of-innomediation

47 Morgan, John, and Wang, Richard, "Tournaments for Ideas". University of California, California Management Review, 2010. http://faculty.haas.berkeley.edu/lyons/Morgantournements.pdf

Crowdsourcing and Formulating Public Policy

The use of crowdsourcing for design, planning, and implementation of policy became widespread after it took hold in the business world. To a large extent, crowdsourcing's growing popularity stemmed from public disappointment over public-policy implementation and the expectation that the crowd could produce better results.[48] Political crises, the rising power of populist parties and forces, the weakening of the political center, and what seems to be an inability to predict results of election campaigns - all these led many to look askance at the traditional political establishment and seek new approaches, which sometimes incorporated traditional methods of making and implementing policy.[49]

Many projects the world over now try to involve the public in policy-planning processes at the different levels (state, municipal); in legislation and regulation; in learning about the public's desires and needs; in contending with crises; in direct communication with elected officials, state institutions, and the public itself; and even as an alternative to the survey method when trying to predict the results of election campaigns.[50]

Crowdsourcing plays a special role when it comes to formulating and implementing national security policy. The constant spread of information technology not only creates challenges for national security but also expands the spectrum of ways in which to contend with these threats. For those in charge of national security, information technology also offers a domain in which the public can be invited to participate and to contribute its abilities and knowledge to dealing with security challenges. The examples from recent years are numerous and involve two main aspects of national-security activity: collecting data - mainly for purposes of internal security - and finding solutions, primarily technological, to security challenges:

- Data collection: Crowdsourcing can be a powerful tool for data collection in real time, especially when it comes to crisis events. For example, during the riots in Britain in the summer of 2011, pictures of the rioters were uploaded to Flickr so as to help law enforcement identify them. Later British law enforcement launched a Twitter-based campaign to encourage the public to help identify rioters, whether by uploading pictures or videos or monitoring rioters' tweets.[51]

48 Lehdonvirta, Vili, and Bright, Jonathan, "Crowdsourcing for Public Policy and Government". Wiley Periodicals, September 3, 2015. http://onlinelibrary.wiley.com/doi/10.1002/poi3.103/pdf

49 Brabham, Daren C., "Crowdsourcing". MIT Press, 2013, http://wtf.tw/ref/brabham.pdf

50 Hui, Glenn, and Hayllar, Mark R., "Creating Public Value in E-Government: A Public-Private-Citizen Collaboration Framework in Web 2.0". Australian Journal of Public Administration, Vol. 69, Issue Supplements 1, March 2010. http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8500.2009.00662.x/pdf

51 This effort led to the arrest of 770 people, of whom 167 were put on trial. See Hui, Jennifer Y., "Crowdsourcing for National Security". RSIS, Nanyang Technological University, March 2015, https://www.rsis.edu.sg/wp-content/uploads/2015/03/PR150317_Crowdsourcing-for-National-Security.pdf

In another case in the United States, a Virtual Border Watch was set up to guard the border between Texas and Mexico. The enforcement personnel placed network-linked cameras along the border at locations known for infiltration and smuggling, and the public was invited to enter a special site and watch through the cameras in their free time. When identifying suspicious movements, people could easily contact a special operations room that dispatched forces to the place in question.[52]

- **Finding technological solutions to security challenges:** R&D bodies have been making use of crowdsourcing for years to find solutions (mainly technological ones) to security problems. DARPA, for example (the Defense Advanced Research Projects Agency of the U.S. Defense Department), has long promoted initiatives to use crowdsourcing for R&D purposes. In an outstanding example known as the Red Balloon Challenge, in 2009 the organization offered a prize of $40,000 to whoever could develop an effective system to identify the locations of thousands of balloons across the United States. The winner was an MIT group that used social-media-based crowdsourcing to locate the balloons. In another case, in 2011, DARPA challenged researchers to[53] develop an algorithm that could decipher documents that had been shredded. It created a site showing remnants of shredded[54] documents and invited the general public to try to decipher them, with a $50,000 prize for the winner. Nine thousand teams took part in the project and the winner was a group of programmers from San Francisco who solved the challenge in 33 days. Then, in 2013, the U.S. State Department launched a competition aimed at promoting technological innovation in weapons supervision. The winner, who received $10,000, proposed a certain technology that the U.S. administration adopted. The British government, too, has experience in crowdsourcing. In 2011 the staff of the national SIGINT organization Government Communications Headquarters (GVHQ) launched a competition called canyoucrackit. The aim was to identify code-cracking experts among the general public. The participants were challenged to break codes of well-encoded documents, and the GCHQ sought to recruit the winners to its ranks.[55]

52 Tewksbury, Doug, "Crowdsourcing Homeland Security: The Texas Virtual BorderWatch and Participatory Citizenship". Surveillance & Society 10(3/4) (2010): 249-262 http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.916.8346&rep=rep1&type=pdf

53 "MIT Red Balloon Team Wins Darpa Network Challenge". DARPA, December 5, 2009. https://web.archive.org/web/20101111082411/https://networkchallenge.darpa.mil/darpanetworkchallengewinner2009.pdf

54 "DARPA's Shedder Challenge". DARPA, 2011, http://archive.darpa.mil/shredderchallenge

55 "Behind the Code". GCHQ, https://www.canyoucrackit.co.uk

Crowdsourcing in Intelligence Contexts

Amid the growing popularity of crowdsourcing, including its applications in the world of intelligence and law enforcement, alongside profound changes in the intelligence profession fostered by the spread of new technologies, some wonder whether crowdsourcing is an intelligence discipline in itself like the five "traditional" disciplines (HUMINT, GEOINT, OSINT, SIGINT, and MASINT). The question is[56] whether crowdsourcing constitutes an extension of other disciplines, such as OSINT or HUMINT, or a discipline in its own right. Like Stottlemyre, we think it is a unique discipline that does not fully overlap with the other ones: it does not require secrecy (either in the method of collection or the kind of information obtained) and is not limited to "one on one" (an agent and an operator) as human intelligence functions in its classic form; and it does not exactly fit the principles of OSINT, which assume, among other things, a passivity of information collection (that is, open-source information collection that occurs without dependency on a collector, involving "waiting" for collection). The new discipline, which here we will call CROSINT (crowdsourced intelligence), combines the human dimension of HUMINT and the open dimension of OSINT, but also entails making broad (hence not secret) appeals to a large number of people and may produce both sensitive and unclassified information items; and it can produce second-order information items based on the aggregate power of the crowd or the analysis of their patterns of discourse.

The new discipline, which here we will call CROSINT, combines the human dimension and the open dimension, but also entails making broad appeals to a large number of people and may produce information items based on the aggregate power of the crowd

The research literature on the applications of crowdsourcing in the intelligence world - as well as the various initiatives in the field - can be divided into three main topics:

- **Information collection:** As noted earlier, the crowd can serve as an information source in itself, whether this involves "extracting" information that it has or using it to obtain information, including in real time. Another aspect in this context is information processing, primarily visual, by the crowd. For example, in mid[57] 2015 the National Geospatial-Intelligence Agency (NGA) launched, together with

56 Stottlemyre, Steven A. "HUMINT, OSINT, or Something New? Defining Crowdsourced Intelligence". International Journal of Intelligence and CounterIntelligence, Vol. 28, No. 3, May 15, 2015. http://www.tandfonline.com/doi/abs/10.1080/08850607.2015.992760

57 Ibid.

the DigitalGlobe company, an initiative to enable crowds to analyze geospatial data, particularly satellite photos, aerial photos by drones, and data obtained from mobile devices. The project, Hootenanny, offers a platform that is accessible to all, and participants can decipher existing photos and upload photos themselves.[58]

- **Knowledge development:** In intelligence contexts that involve the crowd, knowledge development largely overlaps with the use of crowdsourcing to generate new ideas (ideation) and solve complex problems. As far back as the late 1960s, the American security community made use of crowdsourcing to solve the complex problem of attempting to locate the missing submarine *Scorpion*. It had disappeared in the Atlantic Ocean in 1968, and the navy's efforts to locate it had come to naught. Hence the navy set up a large team of experts from various disciplines who collaborated and, together, succeeded to find the spot where the submarine had sunk. A more up-to-date example is the Wikistrat consulting and research company, which offers a virtual community of experts.[59] The company's community of analysts numbers over three thousand people, including specialists from various fields and backgrounds. The company conducts research for, among others, governmental bodies (for example, the Africa Command of the U.S. army), and in each project dozens of specialists collaborate in real time on an online platform. Wikistrat is actually a research body based on a crowd of experts (as opposed to a "general" crowd) that focuses primarily on geopolitical and strategic issues.
- **Prediction:** When it comes to the connection between crowdsourcing and intelligence, the most developed field is that of predicting trends and events. The goal is to overcome the obstacles inherent to human thinking about the future, whether at the individual, group, or organizational level, by forming a diverse crowd whose predictive ability - so it is claimed - is greater than that of an individual or a limited group of people. Below are some prominent examples:
  - The Delphi technique is a long-standing way to use groups to generate predictions. Questionnaires are sent to groups of experts, and the anonymous responses are collected and shared with the group members. The experts are allowed to change or adapt their responses in the next round of voting, and after a certain number of rounds, common predictions emerge that embody the

58 "NGA and DigitalGlobe open source toolkit to harness the power of collaborative mapping". National Geospatial-Intelligence Agency. June 22, 2015. https://www.nga.mil/MediaRoom/PressReleases/Pages/2015-16.aspx

59 Full disclosure: the author of the article served as the company's deputy-director general for strategy from 2014 to 2017. The company's website: http://www.wikistrat.com

opinion of the majority.[60]

- Prediction markets in the intelligence context were established at the beginning of the 2000s by the DARPA research body and the CIA. The market was known as Future Map and had two elements. The first was designed to give analysts in the different agencies a gamelike platform, based on gambles, for generating geopolitical predictions; the second was already open to the public for purposes of spawning predictions. Because of harsh public criticism of the fact that the government was offering ordinary people a chance to profit from negative events (such as the death of leaders or violent coups), the project was canceled in 2003.[61]
- A broader and still-active project, also financed by the American intelligence community, is the Good Judgment Project led by Philip Tetlock. Its aims are to identify cognitive and personality characteristics of "super-forecasters"[62] and to find out what groups of experts can contribute to improving predictive capabilities for geopolitical events.[63] After American intelligence, still stunned by its predictive failures regarding the invasion of Iraq, made a public appeal in 2006 for innovative proposals to improve the community's predictive ability,[64] Tetlock and his partners devised a prediction tournament.
- A further initiative that was recently launched under the leadership of IARPA, and with the participation of the HeroX crowdsourcing company, is called the Geopolitical Forecasting Challenge. It encourages participation in a forecasting competition that is conducted on the HeroX gamelike platform.[65]

Crowdsourcing: Possible Approaches for the Intelligence Community

The growing popularity of crowdsourcing in recent years, and the importing of related concepts from the business world into the intelligence world, have created great hopes that the use of crowd-based methodologies will help intelligence agencies overcome the challenges of collection, processing, assessment, and prediction both on the tactical (for example, countering terror attacks) and strategic levels (for example, predicting revolutions or election results). In intelligence-practice contexts, crowdsourcing helps

60 Yeh, Puong F., "Using Prediction Markets to Enhance US Intelligence Capabilities". Central Intelligence Agency, Vol. 50, No. 4 (2006). https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/csi-studies/studies/vol50no4/using-prediction-markets-to-enhance-us-intelligence-capabilities.html

61 https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/csi-studies/studies/vol50no4/using-prediction-markets-to-enhance-us-intelligence-capabilities.html

62 Super-forecasters are individuals whose predictive ability is more accurate than that of most of the population (including professionals in geopolitical fields). See, for example, Tetlock's book Superforecasting: The Art and Science of Prediction (Broadway Books, 2016).

63 https://www.gjopen.com

64 Siman-Tov, David, "Who Needs Super-Forecasters?" Intelligence - in Practice, no. 2, August 2017 (Hebrew).

65 https://herox.com/IARPAGFChallenge/guidelines

in overcoming three basic challenges: the uncontainable complexity of the operative and strategic environment, personal and group cognitive failures, and pathologies that stem from organizational structures and work processes.

- **The challenge of the environment:** The main attributes of the operative and strategic environment in which organizations function are great complexity and the rapid pace of the changes that occur in it. This has certain ramifications: first, individuals' and groups' cognitive ability to contain this complexity is limited; second, the pace of change often does not allow orderly processes of planning and decision-making, which of course require organizational attention and resource allocation, whereas rapid change necessitates an urgent and ongoing response; third, because of its complexity, it is difficult in the strategic environment to identify a direct link between an action and an outcome, and difficult to assess in advance what effect an action will have.
- **The challenge of thought:** Another set of challenges in strategic planning stems from group and individual cognitive limitations.[66] Human cognition is not a passive but an active process in which the individual and the group construct for themselves a version of reality based on assumptions and conceptions. This process, however, is biased by effects known as "perception distortions", and usually it is not aware of the basic assumptions and the conditions that affect the structuring process.[67]
- **The challenge of organization:** A last group of problems concerns organizational structures, procedures, and behaviors. Generally speaking, these problems can be grouped under the heading "problems in organizational communication". In the context of[68] planning and decision-making in the strategic environment, such problems result in a lack of information sharing and of knowledge development. This can create a strategic threat to organizations, since often different functions are fed by different information sources, all of which are needed to develop the strategic knowledge on whose basis plans are formulated and decisions are made.

The use of large communities makes it possible to overcome - in part - the problem of the environment's complexity and the challenges to thought, because the "collective brain" of the crowd is much more powerful and diverse than the

66 Heuer, Richards J., Jr., "Phycology of Intelligence Analysis". Central Intelligence Agency, 1999. https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/psychology-of-intelligence-analysis/PsychofIntelNew.pdf

67 Another seminal study in the field, mainly from the decision-makers' standpoint, is Jervis, Robert, Perception and Misperception in International Politics (Princeton, NJ: Princeton University Press, 1976).

68 I do not elaborate here on the communication channels (direct, indirect, mediated). For elaboration, see Samuel, Yitzhak, Organizations (Tel Aviv: Kinneret, Zmora-Bitan, 1996), 133-158 (Hebrew).

"brain" of the individual person. Large groups are able to collect large and diverse quantities of information, to process it relatively well and rapidly, and to analyze it from a variety of perspectives, thereby diminishing - though not precluding - perception distortions. Finally, with regard to organizational communication, the online nature and global distribution of these crowd communities, alongside the fact that their motives are generally not economic, enable complex intelligence processes to be carried out simply, efficiently, rapidly, and inexpensively.[69] In intelligence-practice contexts, the use of crowdsourcing can enable intelligence personnel to diversify their viewpoints when they scrutinize any collection or research object. The essence of crowdsourcing is to assemble numerous people from different backgrounds who have different personality, behavioral, and intellectual traits and come from a wide range of fields of expertise. Such variety facilitates a more complex analysis incorporating diverse viewpoints that mostly are unavailable to a small team of researchers, all within a relatively small time frame and at relatively low cost compared to the possible benefits of the work.[70] All that is even more true regarding virtual platforms on which hundreds and even thousands of people collaborate in real time. The fact[71] that a crowd-based discourse is not hierarchically or organizationally structured, and is instead decentralized, can foster interesting internal dynamics that also have an analytical value in themselves (for example, standpoints held by a certain kind of experts compared to another kind). And, finally, the connection between a closed group (such as the intelligence community) and an open group (such as a crowd community) can

The essence of crowdsourcing is to assemble numerous people from different backgrounds who have different personality, behavioral, and intellectual traits and come from a wide variety of fields of expertise. Such variety facilitates a more complex analysis incorporating diverse viewpoints that mostly are unavailable to a small team of researchers, all within a relatively small time frame and at relatively low cost compared to the possible benefits of the work

69 Hershkovitz, Shay and Shkolnikov, Alina, "Harnessing Collective Wisdom". IVEY Business Journal, September-October 2017. https://iveybusinessjournal.com/harnessing-collective-wisdom

70 Lobre-Lebraty, Katia and Lebraty, Jean-Fabrice, "Crowdsourcing: One Step Beyond". Wiley, August 2013. https://www.wiley.com/en-am/Crowdsourcing%3A+One+Step+Beyond-p-9781848214668

71 Gupta, Ravi and Brooks, Hugh, Using Social Media for Global Security (Wiley, 2013). https://books.google.co.in/books?id=Fm0uWf9EL7cC&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

redefine the boundaries of the intelligence community, enable the penetration of new ideas, and challenge the existing concepts.

At the same time, it would be mistaken to regard crowdsourcing as a total solution to all the collection and research challenges that intelligence organizations face, since this method has several liabilities. On the general level, studies have shown that for a certain kind of problems crowdsourcing is not necessarily effective and can lead to less accurate results than those produced by individuals or small groups of researchers. This stems, among other things, from the fact that crowd-based experiences tend to be "freer" from a methodological standpoint than the structured research methodologies commonly used in intelligence services. In addition, because of participants' difficulty in identifying and focusing on the accurate or valuable conceptions within a very wide range of analyses, and because of the frequent mixing of opinion and information and of general knowledge and accurate and proven empirical points, crowd-based processes may not meet the strict methodological standards of intelligence organizations.[72] Moreover, it can be problematic to rely on a broad group of participants, some of whom often are not experts and have unclear motives for participating.

And, finally, often the nature of the problem is not suitable for crowd-based analysis. Intelligence sometimes deals with issues that require obtaining intimate information, and the general public does not have a relative advantage in securing and processing such information. Sometimes dealing with a particular problem entails exposing a secret that the organization wants to protect, and in such cases open discussion among broad communities can pose a problem of information security. Indeed one of the main claims against the use of crowdsourcing by intelligence organizations is the inherent lack of secrecy. At least on the strategic level, however, the intelligence challenges are not secret (though the information on which assessments are based may be classified). Often, particularly when it comes to knowledge development, the task does not concern (secret) information but processes of thought, knowledge development, and knowledge itself. While intelligence organizations are indeed uneasy about publicizing topics of interest, the use of open platforms for the general public does not necessarily risk exposing sensitive information. When an intelligence organization addresses a question to the public, it does not have to reveal all the levels of the question, certainly not the classified ones. Moreover, compared to social networks, in which the personal profile is open to all, many collective platforms enable the concealment of participants' identity (for example, they can be anony-

72 Boulos, Maged N. K., Resch, Bernd, Crowley, David N., Breslin, John G., Sohn, Gunho, Burtner, Russ, Pike, William A., Jezierski, Eduardo, and Chuang, Kuo-Yu S., "Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: Trends, OGC standards and application examples". International Journal of Health Geographics, December 2011. https://ij-healthgeographics.biomedcentral.com/articles/10.1186/1476-072X-10-67

mous) or the building of compartmentalized domains restricted to a certain public, and usually limit access to the platforms by requiring a user name and password.[73]

Toward the Future: Directions and Development of Crowdsourcing for Intelligence Purposes and Others

Kelly Olson, a senior figure in the innovation field in the Obama administration, asserted that the incoming administration (namely, Trump's) needed to keep improving the applications of crowdsourcing and keep investing in R&D in this context. Although the new[74] administration's policy on this issue still is not clear, the crowdsourcing industry already appears to be huge: a report by the IBIS World research company, revised in March 2017, says that in 2012-2017 in the United States alone, the crowdsourcing market grew at a rate of 37%, its value is estimated at $6.5 billion, and it will probably keep growing at similar rates in the coming years. At the same time, it appears that the market[75] has reached a certain level of maturity or even saturation. With over three thousand companies in the world offering crowdsourcing-based services (over half of them in the United States), entrepreneurs and investors appear to be looking for the next generation of crowdsourcing applications. Hence the field is developing in new directions, including medicine and higher education - two realms that have taken some time to adopt it but are now doing so with great enthusiasm. Meanwhile there is an effort to find new mechanisms for incentivization that will motivate people participating in large communities to contribute actively to them. An additional focus is to develop capabilities for multidimensional graphic presentation, and in real time, of information items, insights, patterns, biases, and the blind spots of crowd-based discourse.

The field's main trend of development involves combining people and machines, or more precisely, the use of advanced technologies, particularly artificial intelligence, to analyze the discourse generated through the crowd, and the use of the crowd to improve artificial-intelligence capabilities in general and those of machine learning in particular.[76] These two endeavors can yield what we call "big knowledge" - the aggregate knowledge of communities and the second-order insights that can be derived

73 Olson, Kelly, "Federal agencies take citizen engagement to new level". GSA, December 12, 2016. https://gsablogs.gsa.gov/gsablog/2016/12/12/federal-agencies-take-citizen-engagement-to-new-level

74 "Crowdsourcing Service Providers: US Market Research Report". IBIS World, March 2017. https://www.ibisworld.com/industry-trends/specialized-market-research-reports/advisory-financial-services/outsourced-office-functions/crowdsourcing-service-providers.html

75 Ibid.

76 Hershkovitz, Shay, "The future of crowdsourcing: Integrating humans with machines". The Hill, March 20, 2017. http://thehill.com/blogs/pundits-blog/technology/324807-the-future-of-crowdsourcing-integrating-humans-with-machines

by analyzing the discourse (namely, "Who says what and why"), alongside the use of knowledge development by human communities to strengthen the knowledge-development capabilities of machines. Outside the intelligence world, some trends are already evident:

- The CrowdFlower company enables organizations to carry out different tasks by combining machine learning and human judgment, which allows the machine to improve its work by monitoring human behaviors, using human inputs, and learning from these behaviors.[77]

  The Artificial Intelligence for Disaster Response (AIDR) combines crowdsourcing with machine learning in real time to provide solutions when disasters occur.[78]

  Wirewax offers an interface that combines artificial intelligence and crowdsourcing to identify patterns in pictures and videos. The aim of using people (i.e., a crowd) is to help machines learn to process pictures and videos more accurately by learning human behavior patterns.[79]
- The Debategraph company offers a cloud-based service that helps knowledge communities present arguments in graphic or textual form, raises questions, supplies information, and assesses the level of analysis of members of the community, using tools based on artificial intelligence and the automatic visualization of big data.[80]

In the Intelligence Contexts, the Next Generation of Crowdsourcing Is Already Here:

- The Unanimous AI company has launched a new software designed to generate predictions (and essentially, insights) through crowdsourcing. The method on which the software is based is known as "swarm intelligence" and differs from traditional crowdsourcing primarily by enabling the synchronization of the crowd's insights and facilitating interactions among them in real time. The company's CEO claims that one can thereby engage in group decision-making that stems from the "competition" among the participants in each instance in real time, and one can use machine learning to help the community members learn about the opinions of their fellows and revise the insights (and predictions) rapidly.[81]
- Recently IARPA launched a project called the Hybrid Forecasting Competition (HFC) whose goal is to find out whether interfaces between people and machines

77 https://www.crowdflower.com

78 "Artificial Intelligence for Digital Response". AIDR. http://aidr.qcri.org

79 https://www.wirewax.com

80 Debategraph. https://debategraph.org/Stream.aspx?nid=61932&vt=ngraph&dc=focus

81 Galeon, Dom, "A Swarm Intelligence Correctly Predicted Time's Person of the Year". Futurism, December 6, 2017. https://futurism.com/swarm-intelligence-correctly-predicted-times-person-of-the-year

can improve the ability to predict geopolitical events. IARPA is encouraging the public to register for the program and take part in the forecasts, while offering them an online interface that includes various technological applications.[82]

○ Israel, too, has a place in the field. The Israeli Epistema company offers its customers an online interface that encourages them to collaborate - as a group - in carrying out analytical tasks. The interface can also perform a second-order analysis of the discourse and identify thought patterns, analytical blind spots, and so on.

The emergence of this new intelligence discipline, CROSINT, mandates a change in how intelligence organizations function, particularly when it comes to operative approaches and compartmentalization, defining the nature and tasks of collection and research (and the links between them), and redefining the concept of secrecy while protecting the organizational information assets

Crowdsourcing is not an all-embracing solution for the challenges that intelligence organizations face. Yet wise and diverse use of the tools existing today at each stage of policy formulation in general and of intelligence practice in particular can certainly yield valuable results. The emergence of this new intelligence discipline, CROSINT, mandates a change in how intelligence organizations function, particularly when it comes to operative approaches and compartmentalization, defining the nature and tasks of collection and research (and the links between them), and redefining the concept of secrecy while protecting the organizational information assets. However, investing in tools that combine advanced technological platforms and the ability to decentralize the data-collection and knowledge-creation processes - a key aspect of the CROSINT discipline - will undoubtedly lead the intelligence organizations to the 21st century.

82 "Does (Human + Machine) x Geopolitical Forecasting = Hyperaccuracy?" HFC. https://www.hybridforecasting.com/?source=IARPA

Text as Data :Computerized Content Analysis as an Intelligence Tool

Dr. Carmit Valensi and Dr. Keren Sasson[83]

Introduction

In the big-data era, the quantity of information available and accessible to the researcher is growing. This technological development requires researchers to adapt analytically and methodologically to the need to seek new insights while utilizing and substantiating previous ones, thereby deepening the familiarity with the research object. In light of the growing application of the "text as data" approach to security and policy studies, this article seeks to add a research tool, based on the method of computerized content analysis, to the researchers' array of advanced tools, thereby deepening and enhancing the endeavor of strategic research.

83 Dr. Carmit Valensi - researcher at the Institute for National Security Studies; Keren Sasson - doctor of international relations, specialist in regional security, in comparative consideration, and in strategic culture, research fellow at the Davis Institute.

The point of departure for this methodological shift is that in order to identify, characterize, and assess the significance of the patterns and trends in the discourse of actors in the political domain, great importance must be assigned to how such actors produce a public discourse in accordance with their commitments, thereby validating and translating their interests into behavioral practices. In other words, discourse research based on the text-as-data approach can shed light on the prior processes that shape the actors' behavior.

The purpose of this article is to present the use of advanced methods of text-as-data analysis in intelligence research. To that end we chose to apply the methodology of computerized content analysis - one of the useful and major methodologies in the field of academic research. Applied by social scientists in general and communication researchers in particular, it is mainly found at the methodological point of overlap between qualitative and quantitative research. In this study we make use of the QDA MINER software, which has an honored place among content and discourse researchers in the academic sphere. Computerized content analysis adds an advanced analytical tool to the researcher's set of tools, one that emphasizes the relevance of textual materials (both open-source and collected) and the advantages of integrating the methodologies when surveying and constructing knowledge in broad and varied research contexts.

The software's capabilities are investigated through two test cases - the **speeches of the leader of Hizbullah** from 2007 to 2017 and the **propaganda array of the Islamic State**. By applying the methodology to these two test cases, this article attempts to shed light on the contribution of open-source information to intelligence research at the present time. Thus the article presents and investigates an additional research tool that can help in supporting existing assessments and in producing new insights in intelligence research.

Text as Data: Methodology

The main aim of this article is, as noted, to examine the use of the advanced text-as-data research methods as a key tool for seeking and investigating insights (unknown-unknowns) and for multidimensional validation of existing hypotheses (known-knowns). The research hypothesis is that Hizbullah leader Hassan Nasrallah's speeches can help in understanding his states of mind and offer a research lens that enables balancing between, on the one hand, the traditional focus on tactical aspects of the organization's activity, and, on the other, the consideration of the rationales that dictate them. In the second case, the propaganda array of the Islamic State can offer researchers a deep understanding of the development process this organization has undergone since its establishment as well as its mechanisms for dealing with

the recent changes, and can support and strengthen - or raise doubts about - our assessments about future directions of development.

In analyzing texts we will make use of this methodological array:

1. "Web scraping" of all of Nasrallah's speeches in Arabic and of all[84] the official magazines of the Islamic State (Rumiyah [Rome] and Dabiq) in English and in Arabic (so as to confirm the identity of the contents in the different languages). In addition, by writing Python code, all lines of the content of leading jihadist forums that discuss the Islamic State were scraped.
2. Use of the R language[85] to clean, process, and index the mass of information that has been collected in order to code the contents and define the main variables (time, events, contexts, etc)..[86]
3. Importing the ("clean") texts into a computerized content analysis software (QDA MINER).
4. Applying functions: topic modeling, link analysis, entailment analysis, density analysis, proximity analysis (for explanations of the functions, see the test cases below).
5. Statistical processing of the data with the STATA visualization software.

Next page: *a methodological model for conducting the research.*

Test Case I: Nasrallah's Speeches as a Key to Understanding Change and Continuity in the Organization's Security Concepts and Practices

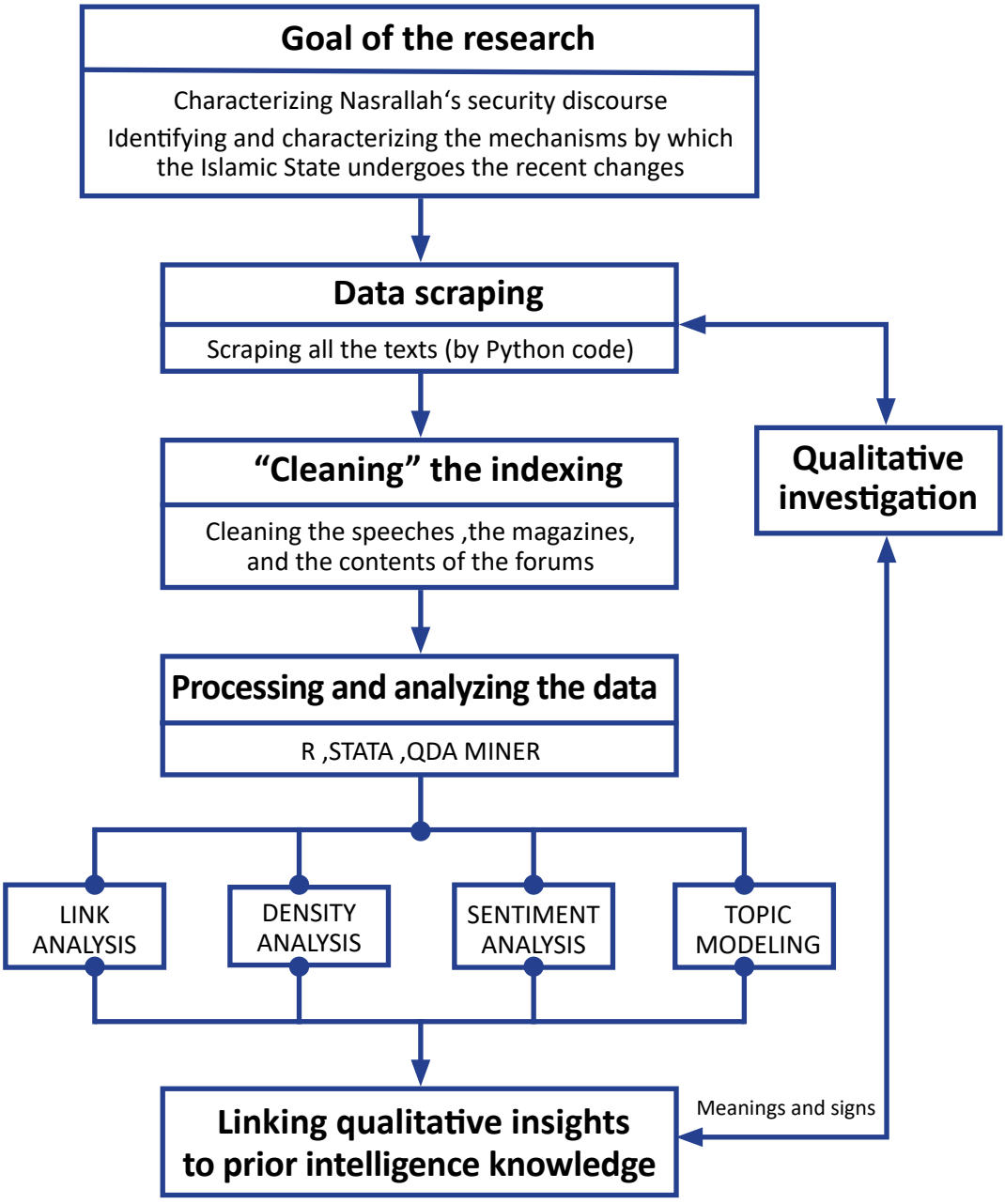
It is worth noting that at first, the content analysis was not oriented toward a focused research question but, rather, toward the objective of exploring the insights that emerge from the texts. In order to identify and map the patterns of change and continuity in Nasrallah's security concept, we chose to divide the data into two main time periods: the first, from the end of the Second Lebanon War (2007) to the organization's declaration of its involvement in the fighting in Syria (2013); the second, from that declaration until August 2017. Because the goal of the research was to investigate Nasrallah's security concept, we had to "clean" the texts that were

84 "Web scraping" refers to collecting and downloading data from the internet in mechanized fashion.

85 R language is a programming language with which one can perform different manipulations of the data (on the network and in general) and particularly for statistical analyses.

86 On average Nasrallah gives 71 speeches per year. About 47 of the speeches are given regularly for holidays and special days. In most years the number of speeches is similar and not far from the average. In the case of the Islamic State, 57 magazines (in two languages) were collected (Dabiq and Rumiyah) and about 1,324,000 lines of content from the discourse on the online jihadist forums.

Fig. 15: A model for conducting research that investigates the text as data.



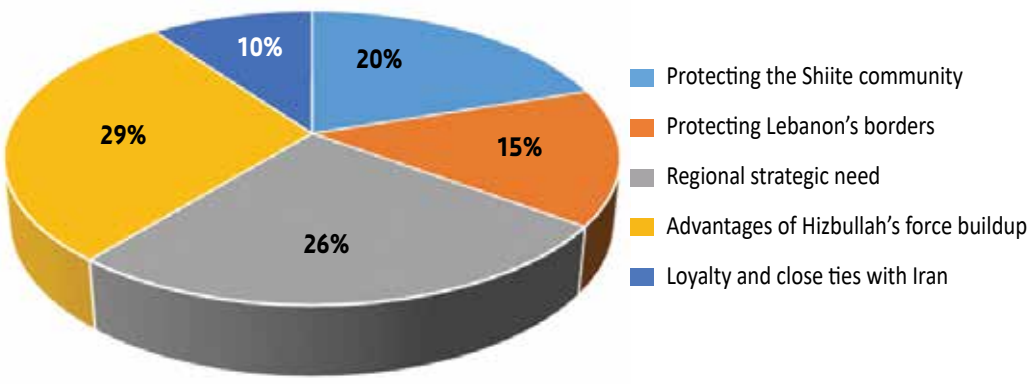
imported into the software and leave only the paragraphs that it identified as dealing with matters related to security issues. For this we applied a dictionary adapted to the concept of "security" (which underwent reliability and validation tests, both of them at an average of >.70), with which all the relevant paragraphs were extracted. After that stage we mapped the main issues on Nasrallah's agenda. For this we used the topic-modeling function, which extracts the most frequently used words in a text and groups them into orderly content categories.

An analysis of Nasrallah's speeches showed that over the years he had developed flexibility and adaptability both on the conceptual-strategic and operative levels, thereby substantiating the hypothesis that the organization's security concept is undergoing processes of change and adaptation to the changing strategic environment, and hence is expanding the range of threats it poses to Israel.

By applying the topic-modeling function, which enables mapping and characterizing the issues with the greatest volume and reference in the texts, we determined that:

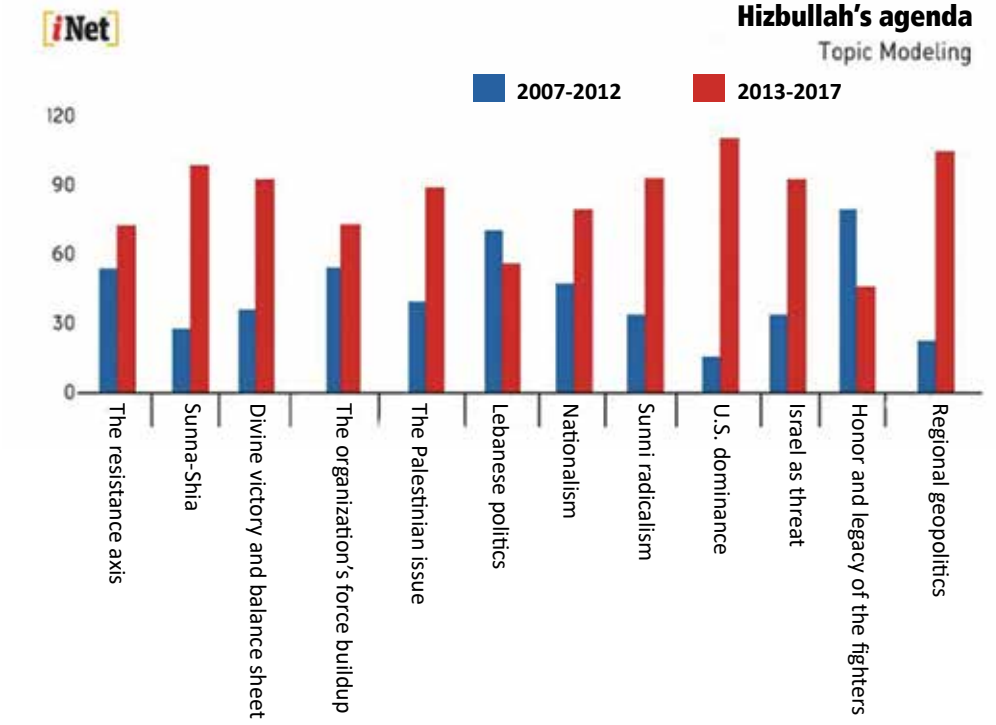
- Nasrallah's official declaration of Hizbullah's involvement in the fighting in Syria constitutes a turning point in his security rhetoric, indicating organizational changes on both the conceptual-strategic and operative levels. Subsequently we mapped, using the proximity-analysis function that measures relationships, proximity, or hierarchy between content entities/categories in the text, the elements of the framing with which Nasrallah justifies the organization's involvement in warfare outside Lebanon's borders:

Fig. 16: The framing of Hizbullah's involvement in the Syrian conflict arena



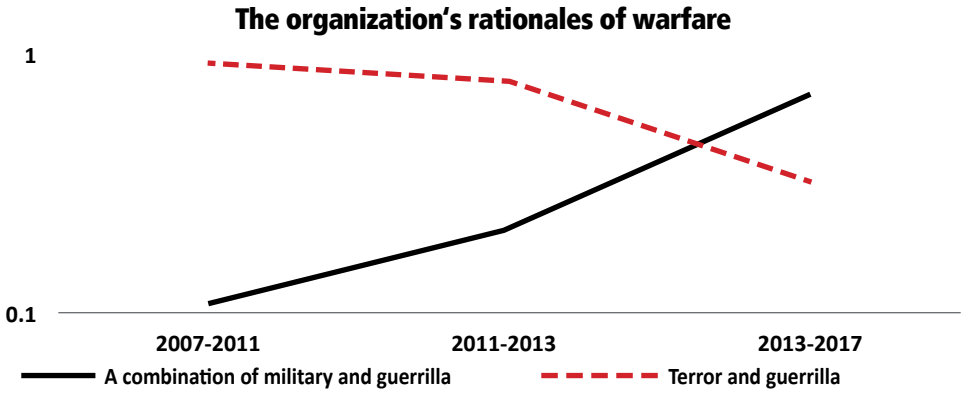
- The mapping of the main issues, those that receive the greatest volume out of all of Nasrallah's references in his speeches, reveals that the occupation with Israel has diminished but still remains a central issue on the organization's agenda. The trend of diminution does not reflect a reduced concern about Israel as an enemy but, rather, reflects the entry of other issues into the organization's agenda.
- A considerably increased concern with Hizbullah's force buildup and with enhancing Iran's role in that context.

Fig. 17: The change in Hizbullah's agenda.



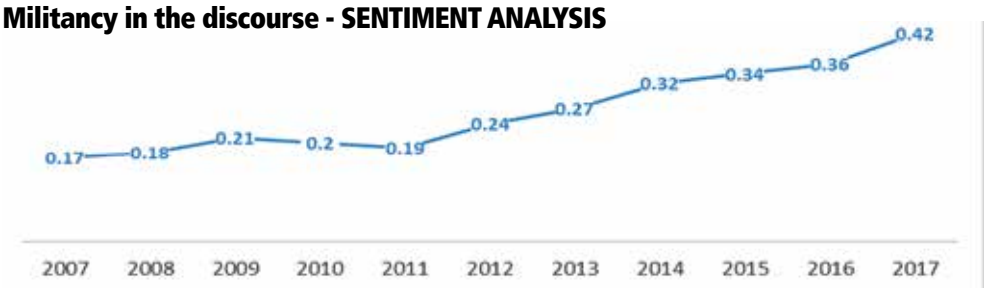
Through the use of built-in dictionaries to identify and characterize military rationales and practices, we identified a trend of (rhetorical) adoption and assimilation of hybrid military rationales with which Nasrallah depicts the development of the organization's military configuration as a combination of "traditional" guerrilla warfare and "regular" military warfare.

Fig. 18: The development of Hizbullah's rationales of warfare.



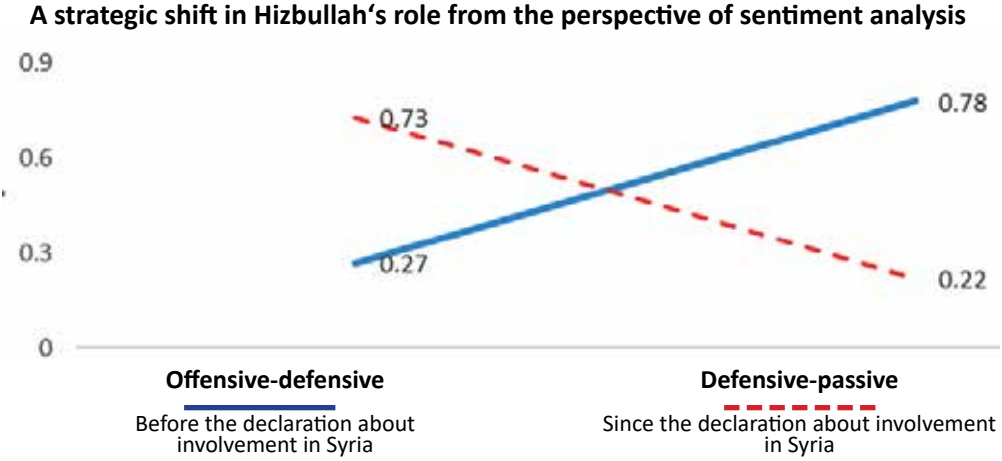
By "running" built-in dictionaries for sentiment analysis (in security-discourse contexts), we identified a rise in Nasrallah's belligerent or defensive tone when referring to security issues (primarily regarding Israel, but also in contexts of other Middle Eastern conflict arenas in which he is involved).

Fig. 19: Militancy in Hizbullah's discourse.



An examination of the semantic relationships in the discourse reveals Nasrallah's attempt to convey to his target audiences the strategic shift in Hizbullah's regional status and role from "passive actor" to "proactive actor".

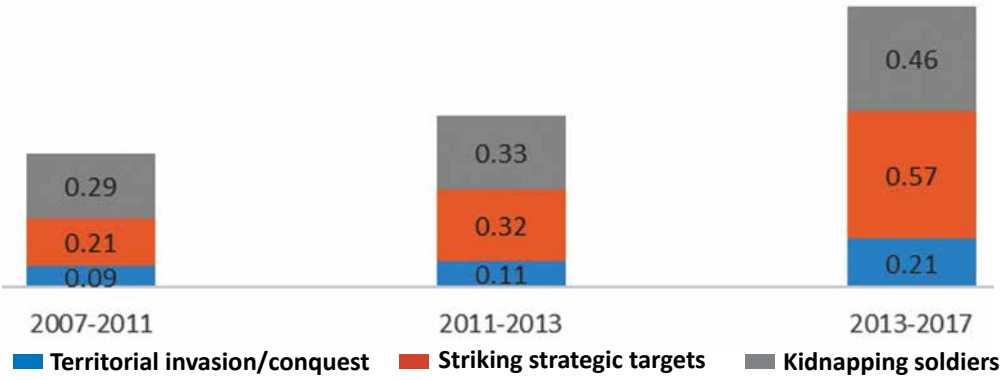
Fig. 20: A strategic shift in Hizbullah's role.



By applying the proximity-analysis and link-analysis functions, which examine and measure the strength of the relationships between the main themes and entities mentioned in the texts, we found evidence of:

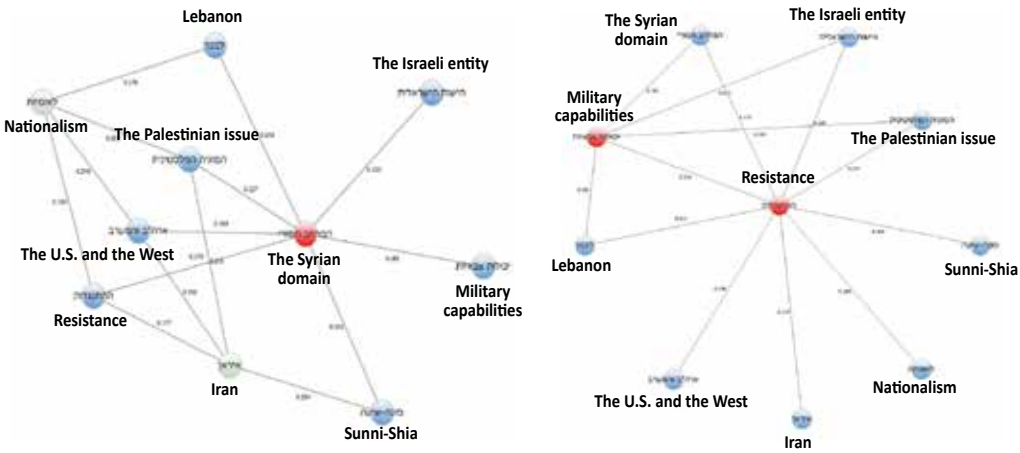
- A gradual distancing from the organization's image as an actor committed only to resisting Israel, while adopting an image of an actor concerned with and available for other conflict contexts.
- A rise in the cohesion of the relationships within the resistance axis, alongside a rise in references to Sunni-Shiite tensions as part of the organization's "just war" framing for the Syrian domain.
- Changes in the components of the threat to Israel, with the emergence of new threats that call for redefining the balance of deterrence between the sides.

Fig. 21: Changes in Hizbullah's approach to threatening Israel.



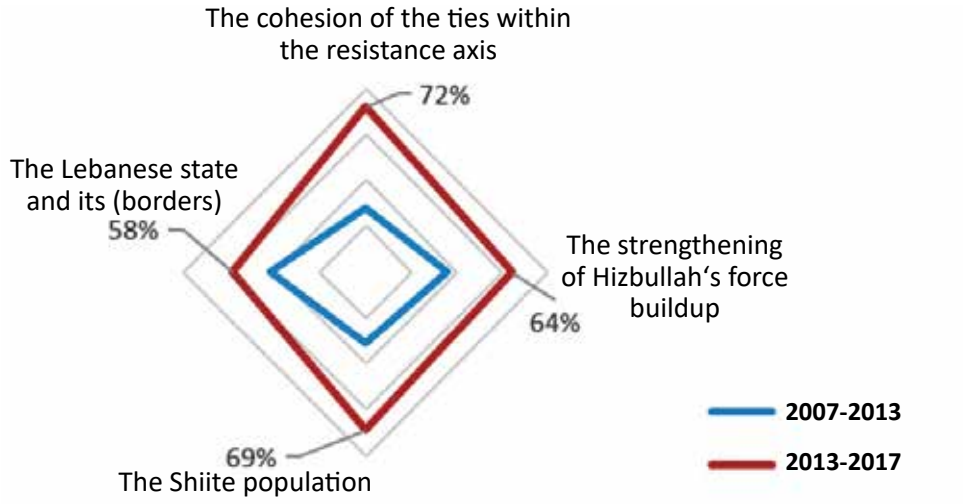
- Application of the link-analysis function, which enables identifying, mapping, and measuring the strength of the networks of discursive relationships in the texts, reveals a substantial rise in the strength of the semantic relationships that Nasrallah constructs in his security notions between Hizbullah's force buildup, the Lebanese state, Syria and Iran as allies, opposition to the United States and its allies, and deepening the Sunni-Shiite tensions in the Middle East.
- The findings indicate that when Nasrallah focuses on "resistance", the strongest semantic relationships are measured in the thematic proximity between Lebanon, the organization's force buildup, the Syrian domain, Iran, and in a relatively innovative fashion, the Sunni-Shiite tensions. However, when Nasrallah's discourse focuses on the Syrian domain, a strong triangle of semantic relationships emerges between Sunna-Shia, force buildup, and the Israeli entity. This reinforces the assumption that resistance entails advantages that accrue from the involvement in the Syrian domain, and that this involvement is "justified" or at least explainable as an important aspect of the escalation of the Sunni-Shiite tensions, mainly surrounding the Syrian conflict arena.

Fig. 22: Semantic relationships in Nasrallah's speeches.



The data extracted from the discourse networks in the texts allow an additional methodological manipulation aimed at identifying the discursive centers of gravity. This function is based on measuring the strength of the relationships between entities, references, tone, and contextual events, and offers an important analytical lens for characterizing the organization's centers of gravity - at least on the rhetorical level.

Fig. 23: The centers of gravity in the Hizbullah discourse.



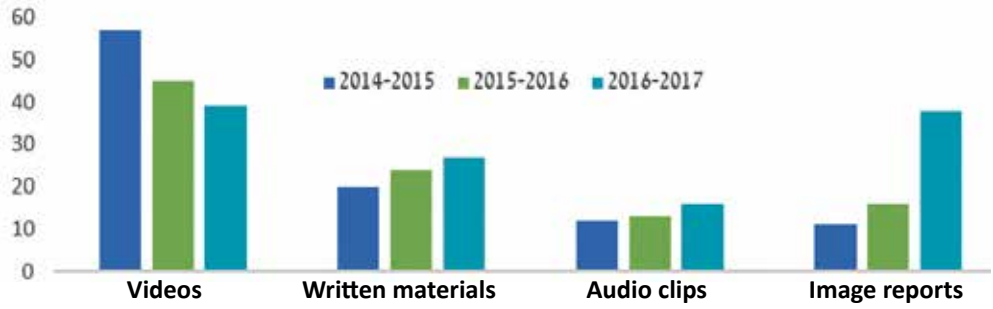
Hizbullah's centers of gravity, as indicated by the contents of the speeches, are: the Lebanese state (with emphasis on infrastructures, borders, and stability), the organization's force buildup, the cohesion of the resistance axis, and the support of the Shiite community. All these reflect systematic continuity in the elements of Nasrallah's security concept.

Test Case II: The Change in the Islamic State as Evidenced by the Official Propaganda Array

The use of quantitative content analysis is meant to provide an additional tool (alongside classic intelligence-qualitative research) for responding to these questions: What change processes has the Islamic State undergone at the present time? What are the Islamic State's possible directions of development on "the day after"? And finally, what can be learned from the intra-jihadist discourse about the change in the Islamic State? These questions were explored by analyzing the evolutionary process that the organization has undergone in recent years and the lessons to be drawn from it, and by conducting a two-dimensional analysis of the Islamic State's official propaganda array (through computerized content analysis of all the official magazines published in 2014 in the original language and in English) as well as an analysis of the global jihadist discourse about the Islamic State's current and future situation (through mapping and characterizing the contents of jihadist forums since 2014).

Web scraping of the products of the Islamic State's digital propaganda array (*Dabiq* and *Rumiyah*) yields several insights. First, there is a substantial decline in the propaganda array's activity with regard to its various products. As part of this trend, the use of written materials (the organization's official magazines) and of image reports increases, while the quantity of videos disseminated on the network decreases. This trend can be linked to the coalition's warfare against the Islamic State, which naturally compelled the organization to reduce its media activity.

Fig. 24: The Islamic State's media products.

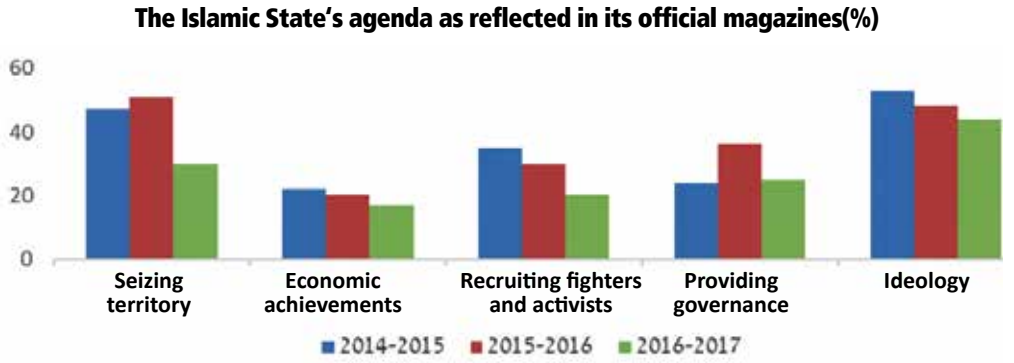


The official media products of the Islamic State(%)

A mapping of the Islamic State's agenda reveals that from 2014 to 2016 its official propaganda array focused primarily on reinforcing the "ideology" mechanism. Most of the magazines' contents dealt with the Islamic State's ideological vision, based on two main elements: publicizing and spreading the "true faith" and explaining the jihadist network and the needs of its people in Iraq and Syria.[87]

87 This was done using the R programming language to apply the topic-modeling function. The function enables ranking and mapping of the main topics that appear in the text and of those that receive the greatest weight/volume among all the texts. This function is based on identifying the most common words in the texts (over time) and placing them in structured content categories (themes). The findings in the graph are a ranking of the main topics that appear in Dabiq (the official magazine of the Islamic State since 2014) and Rumiyah (which has replaced Dabiq since 2016).

Fig. 25: The Islamic State's agenda.



At a later stage there is a decline in the official references to recruiting fighters and activists to the organization's ranks, and also in references to the Islamic State's economic achievements. However, despite the awareness of the changes it had to cope with, providing governance to its citizens remained central in the Islamic State's official discourse because this mechanism was always at the core of the organization's vision and constituted a main source of its power. Thus, despite the decline in the magazines' references to the governance issue, it retains its high frequency and overall weight in the discourse.

A prominent and important finding is the decline in the Islamic State's references to the issue of territorial control. This finding can be attributed to the fact that the Islamic State lost and continued to lose territorial control. An interesting point that emerges from reading and analyzing the texts that appear under this category is that not only did references to the importance of territory for the Islamic State decline, but the tone of the references to the issue changed as well.[88]

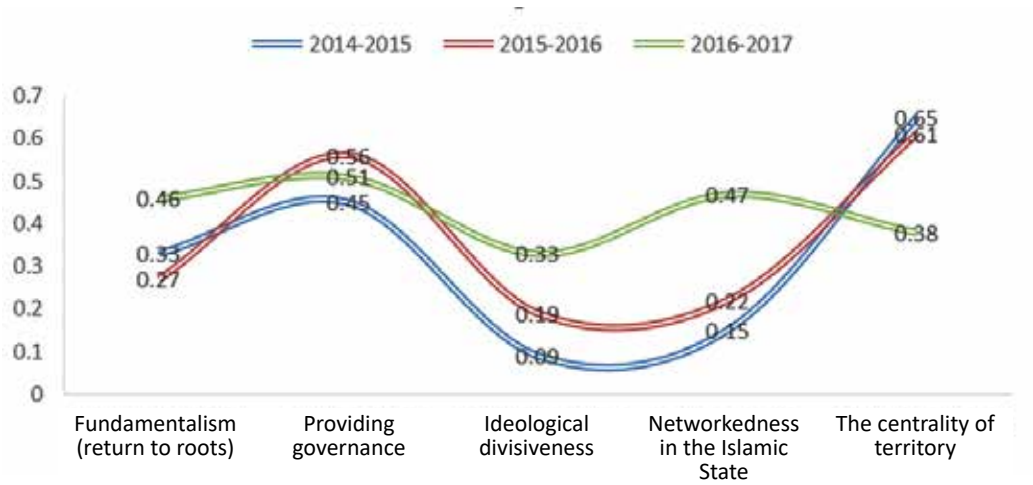
Analysis of the central issues in the Islamic State's official propaganda array[89] yields several insights. First, a trend of divisiveness emerges in the Islamic State's official discourse. The centrality of territory is in decline (as seen in the previous analysis), governance retains its centrality in the discourse (but less than in the past and relatively to the military discourse, which increased), and fundamentalism (return to roots) emerges as an important need for the Islamic State's leadership. Second, it appears that the Islamic State is undergoing a process of strategic rethinking, which

88 Analysis of the tone of the discourse uses the sentiment-analysis function in the computerized content analysis software known as QDA MINER. This analysis makes use of a dictionary (in Arabic) that examines specific reference to a positive or negative tone in the defined context (the centrality of territory for the Islamic State).

89 Done using the density-analysis function, which is composed of a list of adjusted "laws" that are entered into the system by the researchers, enabling the extraction of the main characteristics of the discourse as it emerges from the magazines.

emphasizes the distant enemy "in the West" and the need for Islamic State supporters to operate beyond the Al-Jazeera region and join forces with jihadists all over the world (see the category "networkedness" in Fig. 26). In other words, even though the Islamic State has not completely abandoned its local activity in Syria and Iraq, it appears to be in a process of change that involves moving from a local vision toward a more global strategy, one that bears similarities with that of Al Qaeda.

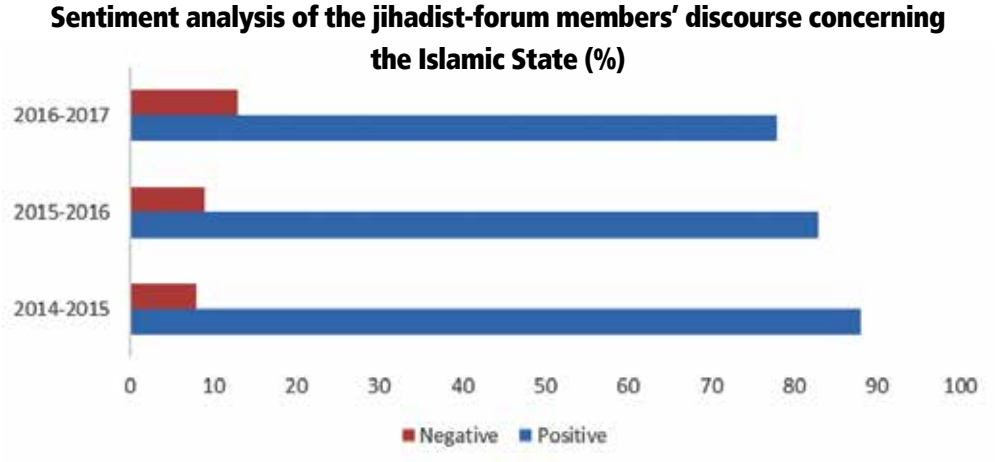
Fig. 26: The change in the Islamic State's discourse.



The Future of the Islamic State as Evidenced by the Jihadist Discourse

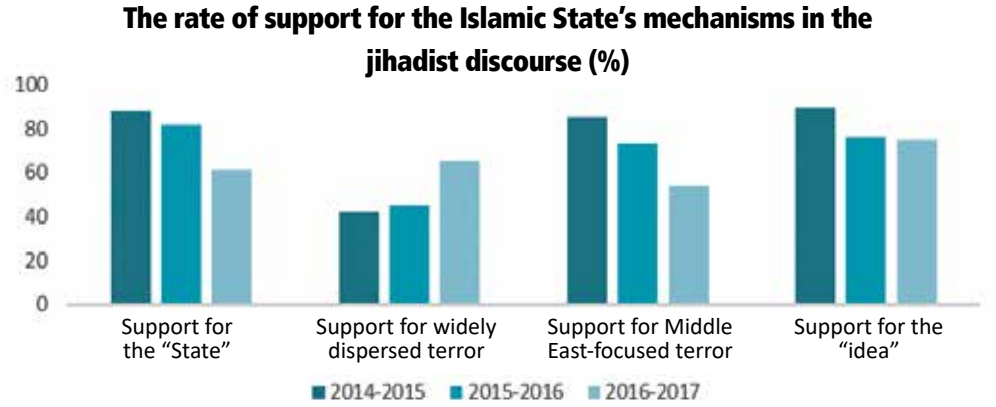
After analyzing the main characteristics of the Islamic State's official discourse, we examined the extent of congruence between this discourse and the broader discourse in the jihadist forums on the networks. First we looked into the prevailing sentiment toward the Islamic State among the members of the various forums. The aim of this kind of analysis was to identify patterns of change or continuity in the attitudes of Salafi-jihadist supporters and activists toward the organization, particularly in light of the changes it has recently undergone, and to map their support for the main mechanisms and their expectations of these mechanisms over the years. An aggregation of the data indicates high rates of support for the Islamic State among members of the forums over time (despite a small decline during the previous year).

Fig. 27: Analysis of the discourse of the members of the jihadist forums.



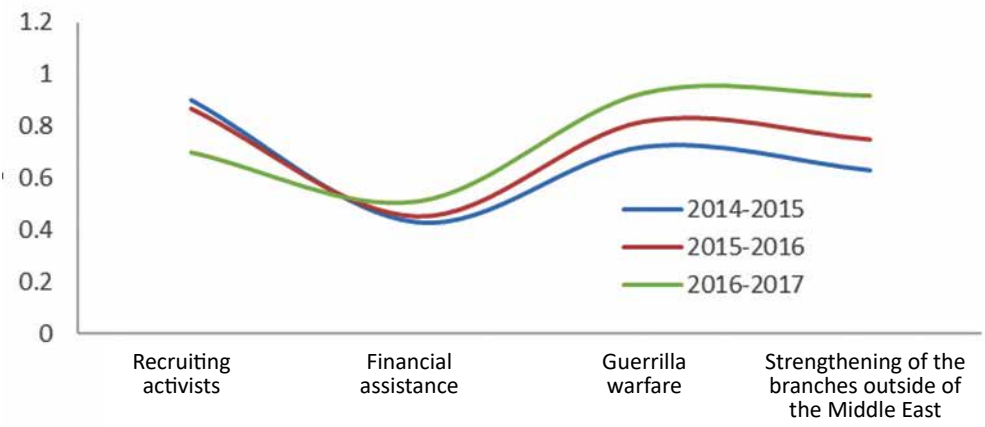
The second stage of analyzing the discourse of the jihadist forums involves exploring and characterizing the jihadist-forum members' support for the Islamic State's main mechanisms. The main findings of this analysis indicate that support for the "idea" of the Islamic State remains strong and stable in the jihadist discourse, while support for the "State" mechanism (in our assessment, as a result of the Islamic State's territorial losses over the preceding year) has declined.

Fig. 28: The rate of support for the Islamic State's mechanisms.



On the other hand, two especially interesting findings emerge from mapping the general discourse. First, there is a trend of growing support for widely dispersed terror by the Islamic State. That is, in the broader jihadist perspective, there is no need to keep focusing on Islamic State's belligerent activity in the Middle Eastern domain; instead the organization's activity needs to be expanded to the global arena (again there are evident similarities here to Al Qaeda and its global characteristics). The possible answers one could "extract" from the discourse, if members of the forums were asked to respond to this question: "What must the Islamic State do to ensure its survival?" yield the following findings:[90]

Fig. 29: Emphases and expectations of the Islamic State as evidenced by the forums.



- Recognition by the Islamic State's supporters of the need to strengthen its branches outside of the Middle East and thereby, indeed, rebrand the organization's activity as focusing on worldwide hostile activity
- A rise in the awareness, and expectation, of a return to the military logic of a guerrilla organization, linked to a decline in support for the "State" mechanism. The importance of recruitment remains as it was (but, again, in the global context), and the volume of reference to the need to mobilize financial assistance for the Islamic State's ongoing activity increases.

90 For this kind of analysis, a special code was used (based on manipulation of keywords in ontologies, scenarios, and probabilities of change in security contexts - both in terms of warfare and of strategy) in the R language to extract all the possible "answers" from the lines of content that were drawn from the forums, and processed and analyzed afterward in QDA MINER and STATA to extract and exhaust the insights. The findings presented in the graph represent an integration of the "answers" during the years defined for analysis.

- The findings of the analyses of the jihadist-forum discourse, then, buttress the assumption (and assessment) that the Islamic State is still the most relevant and dominant actor in the Salafi-jihadist arena and that it is maintaining its vital importance in the eyes of its supporters despite its military failures in recent years.

Conclusion and Implications

This article has illustrated the use of advanced methods of text-as-data analysis in intelligence research, applying a computerized content analysis methodology. Today this is one of the most useful and important methodologies for social scientists in general and communication researchers in particular. The content analysis performed in this document exemplifies a methodological integration between quantitative (computerized) approaches and methods and qualitative interpretation and analysis of the findings through use of the QDA computer software, one of the leading and most advanced content-analysis softwares in the world of academic research.

The different functions that were surveyed point to the complexity of the strategic-research field in the current era. They can be abstracted and made available by adapting the analytical tools to contemporary research. In light of the insights that were obtained from both the question-guided research (in the case of the Islamic State) and the data-guided research (in the case of Nasrallah), the articles illustrates and underlines the relevance of the text-as-data approach and the advantages of methodological integration (combining qualitative and quantitative methods) when it comes to utilizing insights and building knowledge both in this particular context and in other research contexts.

The authors of the article exemplify methodological integration: one comes from the world of quantitative research, primarily in the Middle East and political science discipline, while the second author has a background and experience in quantitative research in the international relations field

Regarding the work process, it should be noted that the authors of the article exemplify methodological integration. One of them comes from the world of qualitative research, both academic and intelligence, primarily in the Middle East and political science discipline, while the second author has a background and experience in quantitative research in the international relations field. Indeed the writing on the Islamic State was done in tandem. A qualitative analysis was conducted based on ongoing familiarity with the

research object and the historical process it has undergone (still in the period of its affiliation with Al Qaeda). On that basis, scenarios were written (involving competing and complementary possibilities) about the organization's future, among other things drawing on the historical lessons to be derived from its previous incarnations. At the same time, the quantitative research was performed, essentially involving the quantification of the data so that the scenarios proposed in the qualitative research could be ranked according to probability in a way that was more reliable and validated.

The assimilation of the integrated approaches can improve the research products of the intelligence community and in general. At the same time, the use of the quantitative methods requires prior in-depth knowledge about the content worlds of computerized content analysis, beyond the level of familiarity with the software that is used. Such understanding requires a change in the approach to training intelligence researchers and in the weight that is given to the study of quantitative issues, beyond the usual content worlds of history and the Middle East.

Finally, the integration of quantitative and qualitative approaches and methods, as carried out in this study, enables abstraction as well as deepening - that is, emphasizing and deepening the familiarity with the research object, alongside abstracting the research field and making its complexity accessible both textually and visually. As a follow-up study, the methodological tools presented in this document could be used to cross-check the insights that emerge from the "open" material with "other" materials and to utilize the intelligence insights in the particular context.

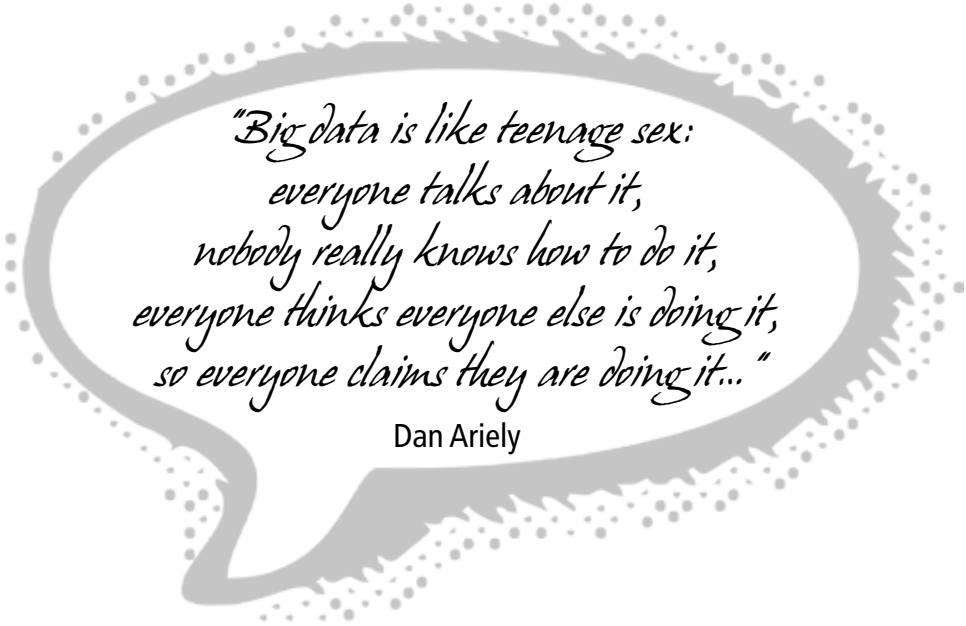
The Use of Mechanized Databases as a Complementary Research Tool: The Example of Research for Operative Purposes

Maj. A. - serves in Aman, and **Dr. Keren Sasson**

Introduction

In most "traditional" operative researches the researchers generally employ a methodology of examining new developments in light of the existing information about the research object, or of comparatively analyzing contemporary events and past events that have similar content. This methodology is assumed to enable the drawing of conclusions that will shed light on "formative factors", "variables", "trends", and "shifts" in the future battlefield, thus allowing the analysis of future challenges at a sufficient resolution for intelligence analyses in the domains of force buildup and deployment.

Despite the widespread use of this methodology, it is not free of limitations. First, in the absence of a necessary correlation between cases from the past and the current (or developing) reality, it remains difficult to identify future trends. In addition, the operative research continues to rely on a relatively small database that makes it difficult for the researchers to reach a high threshold of quality, up-to-dateness, intimacy,



and specificity. Traditional research that uses these methods also leaves great room (sometimes too great) for the researcher's interpretations of the materials collected and may sustain biases and subjectivity. Thus, in most operative (traditional) cases it remains difficult to validate qualitative assessments and conclusions by using quantitative evidence.

In recent years, with the substantial increase in the quantity and quality of data available and accessible on the network, new research methodologies and systems for utilizing data have started making their way into military and security researches as well, with growing use of mechanized databases. Recently researches of this kind have also begun to be conducted in the intelligence branch of the IDF, primarily as part of in-depth strategic investigations and investigations of society and media.

In contrast to the strategic level of analysis, the operative level has relatively lagged in this field, mainly because the tools used in the "open-source information market" have so far been seen as irrelevant to operative needs. Recently, however, awareness of the potential that these tools also harbor for operative-tactical research has been growing. For instance, over the past year research institutes affiliated with the U.S. army in general, and with the intelligence bodies in particular, have conducted a series of applied military researches that deal with the operational environment using methodologies that exhaust the information by means of mechanized databases.[91]

Methodology

This article examines the contribution of mechanized databases to security-military researches. It focuses on an operative test case that addresses the types of warfare occurring in the Middle East in recent years. For this purpose two main databases were chosen:

- GDELT (the Global Database of Events' Language and Tone), which was developed at Georgetown University and surveys, encodes, and makes available for public and research use 98% of all global media.
- ICEWS (the Integrated Crisis Early Warning System), a tool for crisis prediction

91 "Operational environment" is a currently used term that gives a more holistic characterization than the traditional description of the "battlefield". In Western military thought this term is defined as encompassing the conditions, circumstances, and influences that combine to determine how military forces are used. In this approach, operational "practice" is regarded as a system formed by the links and interactions between a large number of variables and subvariables: military variables (involving the nature of the enemy and the opponent, and the characteristics of the conflicts and the military systems), civilian variables, environmental variables, as well as a range of technological variables that have the power to influence events on the kinetic, cyber, or psychological battlefield. The definitions are according to the *Department of Defense Dictionary of Military and Associated Terms,* 2005 edition.

that was developed by the American defense establishment and collects and encodes open-source media data as a platform that enables predicting and warning of crises.[92] Our investigation centers on the researcher's ability to extract applicable conclusions from mechanized databases.

It is important to note that[93] we chose from the start to focus on familiar operative questions, without seeking to determine these tools' ability to improve the exploration of new or different questions.

Hence[94] this article is a preliminary study that examines the possible contribution of research tools from the big-data world to achieving operative insights on issues of changes in the enemy, in the threat, and in the Middle Eastern battlefield since the regional upheaval began.

The retrieval and analysis of the data were based mainly on the indices of:

- Geography: The acts of warfare that were retrieved and investigated were restricted to three "disrupted" Middle Eastern countries: Yemen, Syria, and Iraq.
- Time: The analysis of the data that were collected in the databases was restricted to the years 2012-2016 (2011 was not included because in the first stages of the regional upheaval, most of the hostile actions involved internal security and did not constitute warfare).
- Scope: The data that were analyzed include about 1.8 million "battle events" in the GDELT database and about 39,100 "battle events" in the ICEWS database (initial and critical testament to the differences between the work methods of the databases and to the degree of the databases' suitability for operative researches).

The investigation focuses on two main questions:

First, do mechanized databases enable the monitoring, scraping, and encoding of information relevant to the operative research? To address this question, we will comparatively characterize and analyze the mechanisms for encoding the media reports in the two databases, GDELT and ICEWS. An analysis of this kind is aimed at validating the findings, particularly on the ontological level.

Second, do the open-source media reports contain rich, structured, and precise information that enables the drawing of insights and operative conclusions through qualitative "manipulation" of data? To address this question we will compare, on the

92 It should be noted that currently the predictive function is not accessible to the general public (nor to the academic one) but is applied primarily within the domain of the U.S. security bodies.

93 Before the big-data era, security researchers regarded open-source media reports as insignificant or irrelevant to classified operational researches. The entry of these research tools into the security field, at different research levels, increases the need to assess their degree of suitability for military-operative research.

94 This is a relatively limited investigation that, as noted, allows one to address only the traditional questions in operational researches without exploring the ability to reveal the known-unknown.

one hand, the findings that were obtained in the context of analyzing the data of the databases, and, on the other, the intelligence insights that emerged in the context of basic monitoring and research that is conducted with "traditional" analytical tools. An analysis of this kind is aimed at validating the contribution of the methodology in question with regard to the content questions that occupy the operative researcher.

What Are Mechanized Databases of Open-Source Media and How Do They Operate?

The contemporary research field includes a range of mechanized databases that amass and process media reports, from those established by mass-media corporations to databases like Lexis Nexis. This article focuses on the two most important databases in the research sphere, which are prominent mainly because of the power of their mechanisms with regard to "reading comprehension" and the automatic encoding of events reported in the various media.

- GDELT (the Global Database of Events' Language and Tone): This database grew out of an ambitious project at Georgetown University led by Kalev Leetaru (of Yahoo) and Phillip Schrodt (a Georgetown political-science professor who works with automatic methods for encoding events). The project began in 2011, and in its first years it offered an analytical platform for researchers from the fields of media and international relations. Gradually the investigative use of the data stored in the database increased and spread to diverse disciplines. Three years after the project was launched, the data in the database became available to different investigative target audiences through user-friendly online interfaces (Google Big Query) that facilitate the wide and interdisciplinary use of materials amassed and encoded in the database.
- ICEWS (the Integrated Crisis Early Warning System): The database was developed by the Lockheed Martin corporation with funding from DARPA (the American agency to encourage innovation and research in the security field, which constituted a model for the establishment of the Research and Development Agency in Israel) and the U.S. navy. This is a database whose stages of development (collection, encoding, and analysis of media reports from a certain set of media outlets) lasted for about a decade and were closely attended to by leading experts and academics from different fields of research and knowledge (including political science, computational linguistics, and data science). The availability and accessibility of the database's data[95] were restricted for years

95 The inclusion of experts from research fields in the construction of the knowledge base is a force multiplier when it comes to building any such base and adapting it to the needs of the research that is intended to make use of it.

to the American defense community, and only in recent years have they been made partially available to groups of researchers from Harvard University for different validity tests and trial researches.

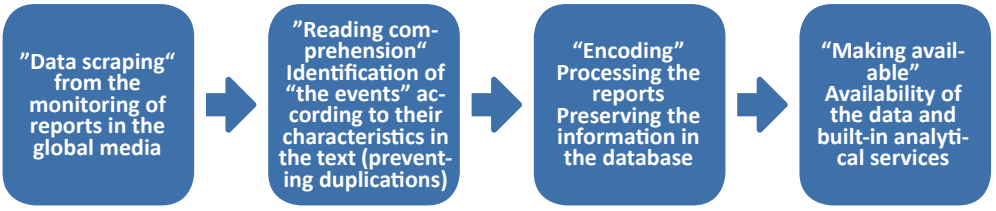
The databases differ from each other, among other things, in the objectives for which they were created and developed:

The GDELT project is intended to produce "a big data history of life, the universe and everything". The database is meant to "encompass the entire human society by constructing a worldwide catalog of human behaviors and beliefs, one that will link every person, place, organization, sphere, subject, news source, and event on the planet to a single massive network that consolidates what occurs in the world. All this while making note of the event's context, who is involved, and how the world relates to it, each and every day".[96]

The ICEWS database was established for security purposes of predicting events, mainly with an eye to force buildup. Apart from the concern with "photographing" the media picture of the situation, the developers of ICEWS wanted to develop predictive capabilities using data that were collected and encoded in the database. This was considered achievable by developing and applying a complex algorithmic. According to the developers, this algorithmic now enables the researcher to attain an accuracy of 80% in predicting and warning about crises.[97]

The operative method of mechanized databases for collecting open-source information includes identifying and scanning articles and reports from numerous and diverse media sources from all over the world, indexing, processing, and encoding the content (according to a built-in encoding book), and finally, analyzing and making the data available for research purposes. The work method of mechanized databases in general, and of those considered in this article in particular, can be seen as falling into four main stages:

Fig. 30: The work method for utilizing databases for purposes of intelligence research.



96 https://amanwiki.services.idf/wiki/GDELT

97 https://dataverse.harvard.edu>icews. It should be noted that currently the predictive function is not accessible to the general public (nor to the academic one) but is applied primarily within the domain of the U.S. security bodies.

Investigating the Suitability of the Databases' Work Methods to the Purposes of Operative Research

The **data-scraping stage** focuses on identifying media sources for the collection and storage of raw information and on monitoring these sources. This is a critical stage because the database's ability to amass and supply data is in fact dependent on it.

- This stage relies mainly on the ability of the algorithm built into the base to identify and continuously monitor the sources selected for the data retrieval. This includes: adding/subtracting sources, dealing with sources that are not media (such as social networks) or are not textual (such as videos), maintaining access to sources that are not free of charge or are closed for private use, as well as maintaining the ability to look back and retrieve data from historical sources.
- The databases described in this article differ in their approaches to identifying and collecting the sources of the media reports based on which the encodings are performed:
  - GDELT's approach is to monitor as large a number of media sources as possible regardless of the source's centrality in the local or global media. Thus the database collects information about events that have been reported in the global media since 1979[98] and gives identical weight to local, small media and to large, major media. As a result, media reports (written, broadcast, and online) are collected that have been published in 65 languages all over the world. The database also "scrapes" (at this stage only partially) articles and posts published in the open social networks.[99] One of the clear-cut advantages of the database, according to those who launched the project, is its ability to scan up to 98.4% of the global media content, with automatic updating at a frequency of 15 minutes.

98 At present a retrospective survey is being conducted of all the media information since 1980.

99 The database also includes an encoding of 0.5 billion video hours (originally in English).

- ICEWS's approach is to limit the range of the "scraped" information from the media outlets and to focus on sources known to be reliable, high-quality, and "validated", sometimes even if it entails taking the risk of "missing" information that did not gain wide media resonance. Hence the database monitors a "limited" number of about 6,000 major media in the world alongside correspondences in social networks since 1995. The database's data files are published occasionally, usually once a year, and do not include an ongoing or automatic update as in the case of GDELT (at the time this article was written, data until 2016 were available for viewing and downloading). In addition, the database does not cover reports in Persian, Arabic, or Turkish,[100] thus considerably limiting its ability to reflect Middle Eastern events in an inclusive, structured fashion.

The "reading comprehension" stage: This stage examines the data that have been collected and encodes them for particular events. When it comes to verifying the data and preventing categorical duplications, the algorithms that guide the information processing in the databases are validated by complex checks and tests (the results of which meet the standard research requirements). It should be noted that at this stage the most significant methodological challenges for the database are identifying the events, that is, "reading" and understanding the essence or nature of the event so as to label it accurately, and preventing double identifications, that is, distinguishing be-

100 A check that was conducted in conjunction with Harvard researchers who are available for developing the database and its data indicates that the Arab media will soon be acquired, and later the Arabic language will also enter development stages.

tween an event and a report about it (for example, if a certain battle is reported simultaneously by several media, the algorithmic of the database is supposed to know how to label and count all the reports dealing with it under one battle). When it comes to creating the link to the individual event within the totality of all the events, an important advantage of the databases in this regard is their ability to provide the researcher with a structured picture in contexts shared by the unique event under examination (power, sentiment, and centrality of the event).

The databases make use of *different* technology for processing and encoding the information collected from the media. This difference naturally produces significant gaps in the accuracy of the databases' findings. At the same time, in many cases a similarity in the phenomena that are presented can be identified from the data of the two databases:

- GDELET's automatic algorithm relies on basic, commercial translation and processing tools. External checks of validity and reliability show that the tools the database uses to identify events and avoid double reports are not sufficiently sensitive or accurate. For instance, in a query on the quantity of the warfare events in the Middle East in particular years, the GDELT database identified 1.8 million events while the ICEWS database identified only 40,000. This finding points to a significant weakness in GDELT's identification mechanisms. This weakness is a main reason for the database's limited acceptance in the academic world.

- The automatic algorithm of the ICEWS database makes use of the BNN ACCENT engine, which is designed for the analysis of natural languages of the Raytheon BBN corporation. The tool is highly regarded among researchers who use the database for its greater precision (including in operative matters), particularly in the context of its identification capabilities. According to the developers of the database, tests of the validity and reliability of data identified through this engine (for events of the "warfare" kind) find a precision capability of 74%.[101] The retrieval of warfare events in the Middle East in the particular years yielded a limited number of 39,110 such events. This numerical differential indicates the logical strong point of the database's identification mechanism as well as its greater suitability for operative vital information.

The "encoding and storage" stage: This stage focuses on encoding the data and preserving them in the database. Because of the large volumes of information, there is no workable way to preserve the data as original text files. Thus, in both databases, after

101 In general, the database also demonstrates an extraction and aggregation capability for events of 60%-80% when it comes to dividing events into different categories.

being processed the information is preserved in encoded form.[102] The data collected from the different sites is encoded by means of the **Cameo** (Conflict and Mediation Observation) encoding book, which contains distinct and agreed-upon encoding categories that have been used over the past few decades in political science studies and validated by hundreds of academic studies. This encoding book makes it possible to differentiate among hundreds of topics, thousands of attributes, and a few thousand subtopics within the numerous media reports amassed in the database.[103]

In operative research contexts, notwithstanding the various differentiations of the encoding book, researchers can rely primarily on data that have been encoded with a very limited number of keywords that do not express all the "richness" of the operative phenomena. Operative questions can rely on only two relevant Cameo encodings - attack or warfare. GDELT also enables the storage of "free" data, including those that are not known and encoded by Cameo but have been identified in the data-processing stage according to the indices of place, name, type, and actor. This includes the following features:

- The database identifies, calculates, and preserves sentiment data, AvgTone (on a scale of 10+/-).
- To each line of content the database assigns the "importance" index, which is calculated by the amount of attention directed to a particular event by the global media.
- In the special context of operative researches, this database is especially relevant because it also calculates and preserves Goldstein-index data for each encoded event, which entails assessing the intensiveness of a conflict/cooperation (here too on a scale of 10+/-).[104]
- ICEWS also offers several unique functions in addition to Cameo data. Even though the data open to the free use of the database are more meager, it should be emphasized that the database analyzes data relating to all the reporting that it ascribes to a certain event. Hence the benefit of using these indices depends on the accuracy of the database's identification mechanism, which, as noted, turns out to be higher than that of GDELT. The algorithmic of the database enables additional

102 The "basic encoding unit" in the database is an "event" that receives a code that is composed of several fields: date, media (in which there is a report about it), language, topic, subtopics, actors involved, country, sentiment, landmarks, and so on.

103 For instance, the encoding book distinguishes among 1,500 religious groups and 650 ethnic groups.

104 This is an index that was obtained and applied in political science studies already in the 1970s, and today researchers in the academic world continue to make extensive use of it. The scale's range is from -10 to +10, with negative numbers representing hostile events and positive numbers representing cooperative ones. These values are found in the intensity field of the data.

functions:

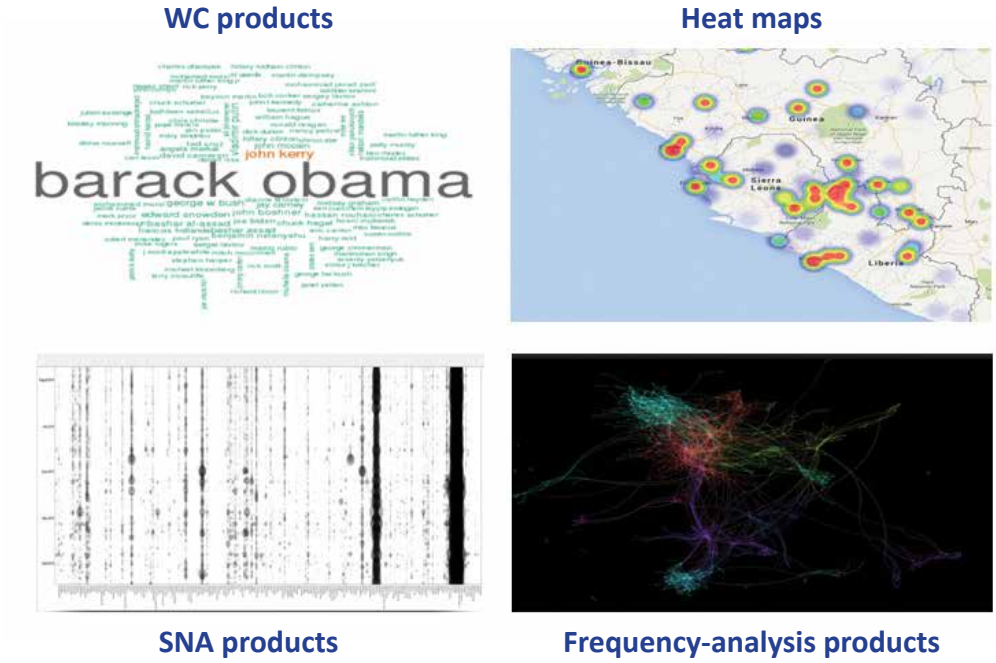
Building and encoding of "dictionaries" (bags-of-words) - databases that are built in computerized fashion under categories of actors, agents, and organizations or sectors. The transparency and validity of the dictionaries are of importance both as the raw material for research analyses and as a platform for the validation and accuracy of the database's "reading comprehension" engines. The intensiveness of the events is measured and preserved in the database by the Cameo index (on a scale of 10 +/-).[105]

The "making available" stage: This stage deals with making the information available that is amassed, encoded, and preserved in the databases, providing tools for analyzing the data and presenting them graphically by means of built-in analytical functions. Generally there are several levels of accessibility of the materials collected and encoded in the database:

- Retrieval: The relatively simple, basic possibility of extracting the data and analyzing them by using statistical-analysis softwares of different kinds (primarily SQL).
- Analysis with built-in devices: The databases enable automatic analysis of the data by built-in analytical services. As we understand it, there is a preference for processing and analyzing data by using certain tools that are external to the database when it comes to exhausting the insights to be found and performing more complex and varied manipulations of the data.
- Visualization: The databases offer a capability of (relatively sophisticated) graphic presentation of the data.
- Accessibility of the raw data in the databases: This refers to the accessibility of data that have been amassed by the database and retrieved by the encoding, but in their "raw" form (returning to the information itself before it was encoded), since in both cases the lines of encoding include the URL data of the articles (in some cases the database also preserves the text itself).
- The information in the GDELT database is available for free research use with periodic, ongoing updating. If the researcher does not have prior knowledge of the SQL language, he can use analytical services of the database itself that offer "built-in" retrieval (the quantity of the data retrieved is limited to 10,000 lines of content only), basic processing of the data, and visualization according to built-in patterns.

105 A familiar and accepted index used in quantitative researches in political science that was developed as a replacement for the Goldstein index, and which was used in the GDELT database.

Fig. 31: Examples of products based on the databases.

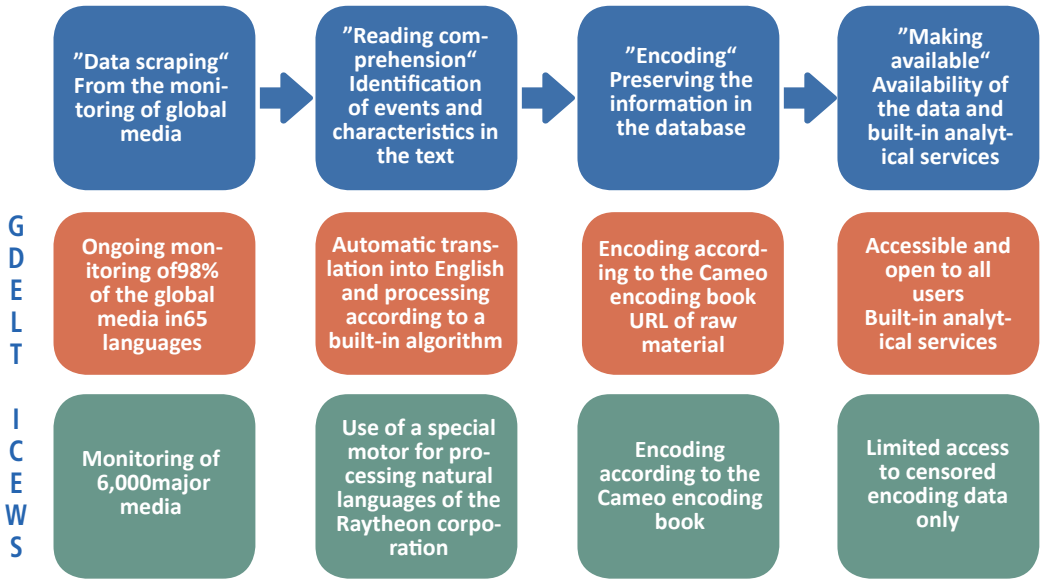


As mentioned, the ICEWS database was designed for the use of the American defense establishment, and most of its data and analytical functions are not made fully available and accessible for widespread use. To a limited extent, the raw data that the database obtains are made available through Harvard Dataverse. It should be emphasized that the limited quantity of the data available for "open" use hinders the creative utilization of the insights that the database may offer. Information and research tools that are made available to the public include:

- Built-in, censored data files (that do not include information on domestic U.S. issues) distributed according to years.
- The only processing device for the raw data that are made available to the public is an aggregation device that offers analyses that amalgamate the retrievals of the database according to a set of built-in indices: years, regions, countries, bilateral contexts, intensiveness, and so on. Some of the products of the aggregations of the retrievals are available at the site and some can be requested from Harvard researchers.

The main differences between the databases as well as their suitability for operative research are mapped in this figure:

Fig. 32: Differences between the databases.



Test case: Characteristics of Middle Eastern warfare, 2011-2017

Data retrieval from the two databases reveals that the volume of the events in the two of them differs and also is apparently larger than the real volume of the events. Apart from the explanation that this stems from the databases' inability to identify double reports and count them as a single event, it appears that the journalistic reporting often omits details that would distinguish between one event and another.

The "meagerness" of the coding precludes in-depth researches aimed at mapping the nature of the fighting, such as defense-offense relations or the topic of secondary forms of battle (advance, retreat, delaying, etc).. In addition, it is difficult (though not impossible) to use these encodings to investigate indices such as mobility (mobile-immobile-stationary), composition, size, and organization of the combat frameworks (platoon-brigade-division) and to assess the characteristics of the fighters' order of battle. All those limitations constitute a significant obstacle to the use of these tools for analyzing the enemy's combat doctrine.

In both databases, the characterization of the types of forces active in the different battle arenas is in fact dictated by those who submit the media reports, which are not consistent. A particular combat force may sometimes be reported as an "army" (to emphasize its political legitimacy) or sometimes as a militia (to undermine its status and prestige). For example, a media report about a militia called "the Victory Army" will cause the automatic encoding of the "actor" as a regular military force and hence distort the reality.

Methodological Issues That Arose from Analyzing the Test Cases

In researches that use databases of the kind surveyed here, there is a need to "normalize" the results obtained especially if they include data that were collected over time or encoded by the different databases, or even reported in different languages. A relatively simple way to[106] bridge the gaps is by adapting the accepted norms of measurement to a particular issue (normalization).

For example,[107] investigation of a quantity of warfare events over several years should be conducted in a wide context of the general increase in the volume of media reporting. Usually data can be normalized by using one of the familiar statistical methods:

- Presenting the data with the use of two different scales on the same graph.
- Presenting the data in percentages and in absolute values (this is suitable for presenting findings such as "quantity within a sample").
- Converting the data to z-values - a method that is suitable for values that are important both for a specific value and for a comparative distribution of the values along the scale.
- In addition to inaccuracies, partial reporting, and meager encoding, the existing coding method in the databases is inherently problematic when encoding a "military" event, particularly one characterized by several categories and subcategories simultaneously (for example, simultaneous investigation of movements, combat methods, weapons, etc)..
- The relative meagerness in the encoding of the data amassed and preserved in the database can be dealt with by importing the data and analyzing them more thoroughly at the second (complementary) stage of analysis, which will enable the researchers to perform different, more complex manipulations by using the method of multilayer analysis and applying more[108] sophisticated tools (such as STATA, R language, or Python) for in-depth, systematic analysis of the data that will facilitate delving into the media reports, filling the absences that were identified in the previous stage, and thereby completing the picture of the reality that is supposed to be investigated or described.

106 The orientational capacity of the reading comprehension engines of the same database can vary significantly in different languages.

107 Thus, for example, in research on procurement of military equipment, it is necessary to "normalize" the values of the procurement to the value of the currency over time so as to enable comparison of the data in relation to the budget or the GDP and to other comparative values as well.

108 This method is guided by the rationale of examining different stages and levels of analysis by using different research tools and methods. For example, retrieving data from databases is one method (or one level) of analysis. Quantitative manipulation through the use of STATA or any other statistical tool enables more complex and thorough exhaustion of the data and constitutes a different and additional research method and level of analysis

Conclusion and Recommendations

The big-data era presents opportunities as well as challenges for intelligence researchers in all the disciplines and for operative researchers in particular. Meanwhile the use of advanced tools and research methods, of operative, tactical, and micro-tactical kinds, is increasing. In the present era, huge quantities of knowledge and of open-source information make it possible to expand the writ of research by conducting investigations with wide-scale samples. The existing systems and the accessibility for researchers allow the investigation of phenomena in a broad regional (and even global) context, making relatively simple comparisons between different actors so as to identify and characterize patterns of change and learning over time. The use of these tools helps generate insights about various developments and dynamics that can be identified and mapped only from a broad context of events; investigating the databases makes it possible to publicize products with a low classification (accessible to consumers at all the levels) and thus surmounts the obstacles of compartmentalization and classification of sources. Also notable is the possibility of conducting a rapid exploration so as to disseminate points of interest that are relevant to further research (with different methods).

In the present era, huge quantities of knowledge and of open-source information make it possible to expand the writ of research by conducting investigations with wide-scale samples. The existing systems allow the investigation of phenomena in a broad regional context, making relatively simple comparisons between actors so as to identify and characterize patterns of change

The use of these qualitative tools provides a wide scope for exploration that can uncover unfamiliar phenomena. By using these methods one can find quantitative evidence for qualitative assertions and "intuitive" statements. In our opinion, the databases that were scrutinized here have a medium potential to contribute directly to research on vital operative information because of some ontological and methodological problems. However, we recommend using them in intelligence researches in light of:

- Their evident contribution as a complementary stage of operative research, particularly when it comes to making a rapid choice of research hypotheses, identifying trends, and collecting qualitative evidence to reinforce intelligence research findings, while exploiting the simplicity of use and availability of these tools.

- The databases enable the creation of a two-stage research system. In the first stage, the database should be used to point to the relevant sources of the text and to acquire them with its acquisition devices; in the second stage, the texts that were collected can be systematically analyzed with a text-analysis device. The processing and analysis of the findings with a two-stage research system of this kind will enable the researchers to maximize the existing knowledge in the databases and to refine the insights that emerge from the data in a manner that is more accurate and better adapted to the research question.
- It should be noted that, during the research, the tools offered great advantages in the political (foreign and domestic) field, in research on societies and populations, and in terror research. All this could be done relatively rapidly and simply while exploiting the advantages of open-source media in studies in these fields.
- In light of the evident advantage of the ICEWS database with regard to the identification and "reading comprehension" mechanisms, and despite the present difficulty in making full use of the database (which entails ensuring accessibility in conjunction with the American defense establishment), we recommend simultaneously using both databases.

In essence, a look at the methodology of exhausting information with mechanized databases underlines this methodology's contribution as a complementary research tool, one that gives the researcher what is needed to contend with the growing volume of the available information with which existing research propositions can be substantiated or refuted as well as quantified

In essence, a look at the methodology of exhausting information with mechanized databases underlines this methodology's contribution as a complementary research tool, one that gives the researcher what is needed to contend with the growing volume of the available information with which existing research propositions can be substantiated or refuted as well as quantified. This methodology brings the researcher to the forefront of the contemporary research endeavor because it enables a synthesis between the traditional familiarity with the research object and the insights that emerge from mechanized databases. These databases can therefore be used as an additional and enriching analytical stage. Thus, despite the inherent limitations of these research tools, in our assessment it is worth enhancing researchers' familiarity with research devices that can make use of the different databases.

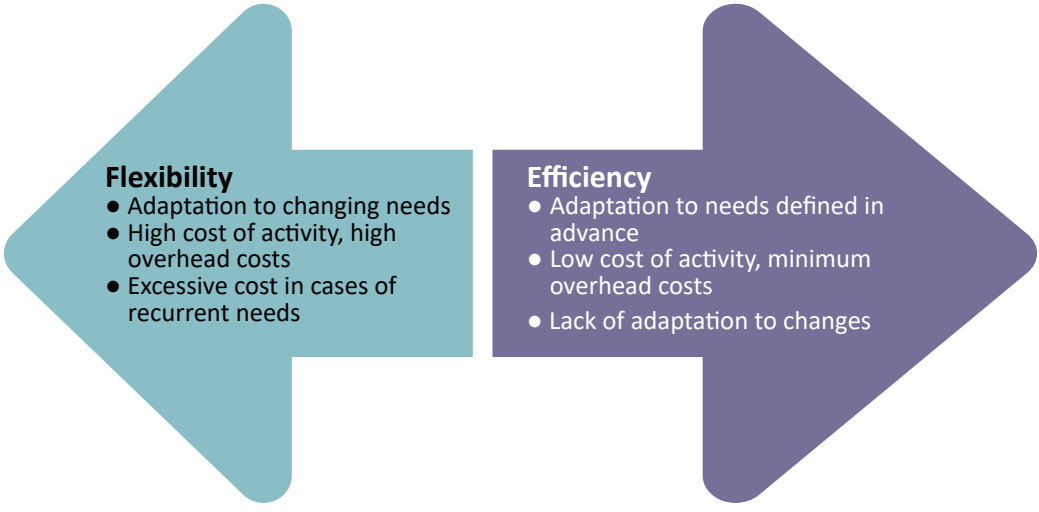
Operational Optimization in the Era of Variation, Big Data, and Infinite Change

0. - serves in Aman

The Old Intelligence Era: Built-in Tensions

Intelligence practice is in a state of built-in tension. On the one hand, the nature of the intelligence research objects is dynamic, unique, varied, and constantly updated. The basic objects of intelligence practice - states, organizations, people, processes, events, and so on - and the research processes for clarifying their nature and significance are not at all regular; all stem from the needs of military, security, and national activity, which are by nature context dependent, unique, dependent on time, space, and context, and not foreseen or given in advance. On the other hand, the intelligence organizations themselves - large, complex institutions that depend on numerous and expensive resources - must operate under economic constraints and demonstrate efficiency and optimal exploitation of their resources. These constraints spur the organizations into processes of operationalization and institutionalization that are manifested in fixed rationales, technological institutionalization, broad infrastructures, and so on.

Fig. 33: The tension between flexibility and efficiency in intelligence practice.



The above-noted tension hovers between flexibility and efficiency. On the one hand, the more the intelligence process tends to flexibility the less efficient it will be, because flexibility will come at the expense of the consolidation, structuring, formalization, and mechanization of aspects of activity that enable the saving of resources. On the other hand, the more the intelligence process tends to efficiency the less will it be flexible and adaptable to changing and particular needs. If, for the sake of functional efficiency, the organization builds, defines, and consolidates processes (and thereby improves its efficiency by saving on overhead costs at every stage of activity), it will not be able to adapt itself to each new need without impairing these constructions, which will require considerable resources.

The degree of organizations' efficiency with regard to their objectives can be gauged by considering the components of the organization and checking which of them maximize flexibility and which maximize efficiency. An analysis of this kind is comparable to analyzing the space in which the organization functions - its variance and change, and whether there is a lack of adaptation in flexibility or in efficiency with regard to the changing nature or the regularity of the space. In an intelligence entity this inquiry should focus on the two basic domains of reference for the process of crafting the intelligence assessment - the intelligence world of events itself, and the types of information by means of which this world is investigated.

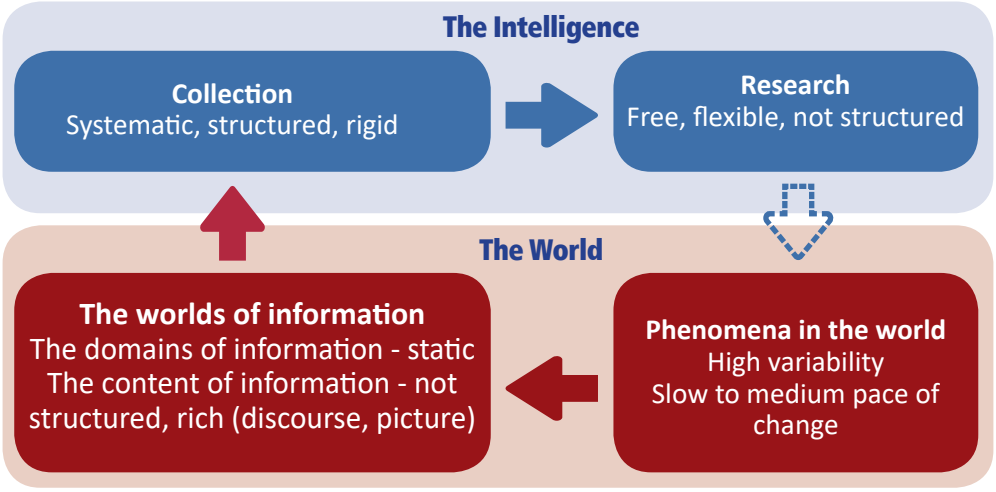
Historically the intelligence process made a separation between two main components: the collection processes and the research processes. The research processes required great functional flexibility because the research product - intelligence assessments about the research objects - had to give an optimal account of the unique and special situation of these objects. Thus the research organizations developed flexible and free processes of production based on textual products with free wording, as well as processes of knowledge production and assessment based on relatively free access to processed information, research discourse, and a synthesis of processed materials up to the level of research products (surveys, compilations, and so on).

In contrast, the collection processes indeed required maximal efficiency because of the nature of their activity - namely, bringing large quantities of information items from the relevant, immense informational environment into the research domains. To achieve their objective amid the quantity of information, the collection bodies developed work processes and systems that maximized the efficiency of the filtration of the copious information, of the identification of the relevant information within it, and of its rapid transfer to the research bodies. In particular, the research bodies were constructed in a rigid fashion so as to deal with three domains of information: the optical (VISINT), the signals (SIGINT), and the human domain (HUMINT).

This "operational efficiency" was achieved at the expense of the flexibility of the collection system, and did not enable (though usually the need did not arise) the making of changes and adjustments for each new need and research object.

Fig. 34: Between the world and intelligence in the old era.

In the human and informational reality in which these processes developed, it appears



that the above-described design of the organization was valid for two main reasons.

The first reason is the nature of the threats in that era, which were organized in large, hierarchical, and relatively stable human frameworks such as states or armies. Such a situation created a need to understand the intentions, motives, and rationales of key individuals in these frameworks, which determined the other side's steps in the different hierarchies. The variety and quantity of the research objects were relatively limited, and their rate of change was slow. And, apart from them, it was not overly important to describe and understand numerous other entities. In such a state of affairs, the attempt to obtain and learn as much as possible about what was happening in the mind of the enemy was vital, and optimal learning focused on his discourse and verbal expressions, which created an inherent and systematic need to obtain abundant information about his conversations and statements. The issue of "capabilities" also dealt with limited domains and entities, reflecting the power and physical situation of these organizational frameworks.

The second reason is the limited variety and relatively slow rate of change of the technological and communication world with which intelligence had to deal at that time. This relatively slow change made possible the establishment and

institutionalization of large networks for obtaining, treating, and distributing numerous news items from the basic domains (SIGINT, VISINT, HUMINT) in the same fashion and in a relatively regular process. Overall, as far as results were concerned, it appears that there was a relative congruence between the pace, variety, and flexibility of intelligence practice, in accordance with the operative and strategic needs, and the approach to collection at that time.

The Big-Data Era, Accelerated Change, and Their Implications for the Intelligence Endeavor

In recent decades we have been witness to two meta-processes that humanity is undergoing. One is the massive and accelerated spread of technologies all over the globe, as computing and communication capabilities get relatively cheaper and are made massively available. Once unavailable to most of the public, today they are in almost everyone's possession. The second is the ongoing erosion and dismantlement of the large human organizing frameworks, and the "flattening of the pyramid" such that the bases for identity, for the capacity to choose, and for the fulfillment of the individual's desires are now more elastic and open to him than ever before. Clearly the second process is an outcome of the first, which essentially has altered the balance of power between the individual and the resource-rich entities and organizations that used to determine reality for him, and today are less powerful and less able to control the individual's reality.

With the advent of the big-data era, the quantity and the variety of the digital manifestations of phenomena from the physical world have multiplied radically, and have drastically increased the potential to gain numerous and complex insights. Yet, at the same time, the costs and the resources required for obtaining, treating, and storing the information have also increased drastically

Two of these processes undermine the basic assumptions of the intelligence endeavor. First, the strategic environment, with the threats and the opportunities that it harbors, is no longer limited to state or large hierarchical organizations that are managed by key individuals. The asymmetric threat results from a situation in which, because of the spread and penetration of technology, each person or small framework can exert influence and constitute a threat. This change not only increases the goals that intelligence must pursue but also the variety of the objects as well as the rate of emergence of threats and changes,

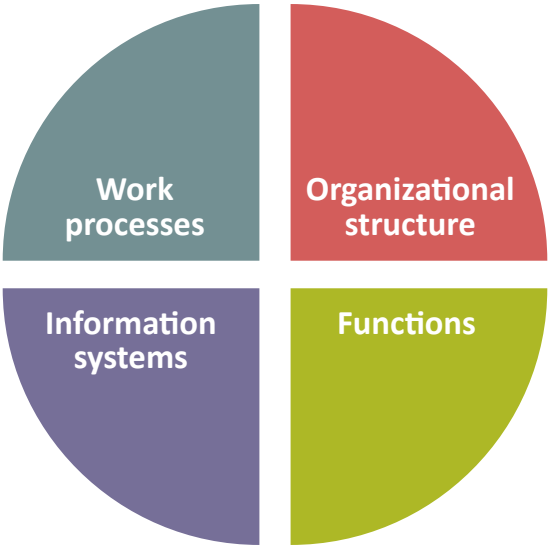
which is accelerating all the time. Second, technological-cultural development is also constantly accelerating, and many human processes that in the past were carried out with few and basic technologies are now shifting to numerous and diverse platforms that offer a specific response for each issue and are accessible to all.[109] This state of affairs changes beyond recognition the technological, communication, and informational reality within which the intelligence collection systems designed in the past must operate.

The challenge stemming from change in the technological and informational environment is many times greater. The advent of the big-data era has seen an expansion both of the quantitative aspect that requires treatment, and of the diversity and variability of the materials that require treatment. The quantity and diversity of the digital manifestations of phenomena from the physical world have increased radically, and this, in turn, has drastically increased the potential to gain more numerous and complex insights than ever. At the same time, however, the costs and the resources required to obtain, treat, and store the information, along with the investment that is needed to identify and exhaust the desired specific pieces of information, have drastically increased as well. All this, as noted, occurs in tandem with the accelerated rate of occurrences in the world, the emergence of changes and of further changes, which, in turn, has accelerated the emergence of gaps in the intelligence processes.

Clearly, in the era of change and huge diversity in which we live, the validity of institutionalized, rigidified processes is briefer than ever. With reality dictating an ongoing need for readjustment and flexibility, a new break-even point is needed in the tensions between flexibility and efficiency. However, the state of the resources, and the cost of dealing with the huge diversity of types of information, dictate the need for institutionalization and efficiency. This break-even point actually exists in the multidimensional space that determines the organization's functioning - the work processes, the manpower and training, the technological infrastructure, the organizational structure. Each of these must find its own break-even point, and all of these together must find an "operational optimum" in light of the amplitude of diverse outputs that are required of the intelligence endeavor, which differ in nature and in the pace at which they can be produced.

109 Consider, for example, the applications installed on your mobile phone. In the past, all that could be done with these applications could be done (if at all) only through a telephone conversation between people or face-to-face. Nowadays applications, which include textual and built-in information, substitute for processes of checking information, purchases, financial transactions, transportation and traffic, dealing with medical information and receiving medical services, navigation, receiving government or municipal services, social processes, giving updated information, sharing experiences and media, finding romantic partners, recording physiological data, physical training, keeping a log of events and tasks, invitations to meeting and events, technical support, invitations and managing welfare and leisure activities, and so on. More and more human processes are shifting to the digital dimension.

Fig. 35: The work processes in the intelligence endeavor in the big-data era.



To enable the finding of a suitable break-even point between, on the one hand, flexibility and efficiency in the work processes and, on the other, the rigidifying and institutionalizing of the work processes, the intelligence process should be broken down into its logical components, and the significance of rigidification or elasticity with regard to different needs must be understood. This requires, first, describing the general nature of the intelligence process as a logical prototype of the research and collection processes, in which intelligence must deal with numerous and diverse arenas. Such a prototype is described in the concept of knowledge development with its three components: conceptualization, known unknowns, and unknown unknowns.[110]

As for the conceptualization and the systemic learning process that implements it, these do not deal with treating and obtaining information but, rather, with interpretation and with crafting a strategic conceptualization of the research arena. Thus the main effect of the era of big data and accelerated change is the frequency with which the systemic learning cycle, and the formulation of the picture of the world, must be conducted. Because this process is in any case a human one based on critical

110 The concept of knowledge development - a methodological description of the full intelligence process - was formulated by a learning group two years ago and described in the first issue of this journal. This concept views the intelligence endeavor as essentially one of ongoing learning, with the learning process consisting of an interaction and synergy of three functions: the conceptualization element, the known-unknowns element, and the unknown-unknowns element. For more, see Head of the Design Department in the Research Division, Aman, "An Intelligence Knowledge Community as an Operative Mechanism That Provides Strategic and Systemic Flexibility to Aman". *Intelligence in Theory and in Practice,* no. 1, May 2017.

and interpretive discourse, there is no strong aspect of "institutionalization" here stemming from the slow pace of the old world,[111] and the process can be accelerated as much as necessary vis-à-vis the rate of change in the arena or in the research field.[112] Here it should be noted that process-related concepts from the old era, such as the "annual intelligence assessment", appear to be less relevant in the new reality. In addition, the way in which the conceptualization and learning processes rely on information products may also need to change in light of the changes in pace, and the need to exhaust the wealth of the information in the learning process in a way that perhaps does not exist in the traditional approaches of transferring individual bits of information.

As for the component of the known unknowns, here the analytical-investigation approach should be seen as a logical prototype for the process of conceptualization and finding answers to factual questions through information. As noted, the analytical-investigation approach divides the intelligence process of obtaining a factual answer into several stages, and motivates the intelligence research bodies to produce orderly theories about the world and the information, in light of which and through which one can examine information and identify findings (signatures) in it. From these findings, according to the investigative theory, one can make inferences about phenomena in the world (characteristics) - from which, in turn, according to the same theory, it will be possible to make inferences about the research object itself.

Fig. 36: The analytical-investigation approach.

111 There is indeed no institutionalization of the interpretive process in the operative mechanism and in the intelligence organization. However, it is possible to construct hardened and institutionalized processes on the operative side as a consequence of the arena-level conceptualization, such as IDF force-buildup plans, operative plans, orders of battle, training, and so on. From the perspective of a particular conceptualization, it is certainly possible that operative bodies in the military will build up their force vis-à-vis the described arena-level threat, and it will be difficult for them to change quickly as the threats and their interpretations change (an example is the classic infantry training for conquering simulated Syrian positions, which is based on plans stemming from an operative concept that apparently is no longer relevant). The need for operative flexibility in the current era is, however, beyond the scope of this article.

112 A case of an extreme need for conceptualization and rapid, frequent interpretation is what occurs during warfare, in which the state of affairs, the operative rationales of the Israeli side, and the nature of the enemy side change rapidly and are affected by the intelligence-operative process itself.

The strength of the analytical-investigation approach is that it shows the organization the theoretical basis for producing insights about the world and enables it to examine them critically. An integral part of the requisite examination of the intelligence theories about the world, and of the information at the basis of the investigation approach - apart from determining the theories' validity in the first place - is the examination of how long these theories will be valid and their degree of changeability or durability with regard to both information and the world. Such examination allows the organization to institutionalize or elasticize its form of organization for conducting investigations. Such investigations will be based on theories that appear to the organization to have long-term validity, and to show stability and durability with regard to both the world and the information. That, in turn, makes it possible to determine work processes, organizational frameworks, supportive technologies, and so on for these theories. It should be emphasized that investigations that deal with dynamic and changing fields are likely to require a more flexible organization, beginning with the work plan itself and also in aspects that facilitate its implementation such as training, technologies, sources, and so on.

For example, an investigation designed to identify radars among essential bits of information about aerial superiority may perhaps turn out to be relatively stable, in light of the relative durability of the theory about the world and the information - radars (characteristic) produce electromagnetic energy, electromagnetic energy (signature) is absorbed by an ELINT sensor, and thus the investigation process can be relatively durable, with regular checking of electromagnetic signatures in the ELINT range and inference of the existence of radars from these signatures. Alternatively, an investigation designed to identify lone-wolf terrorists, with their great variability and the different ways in which they are manifested in the information, will have to change and replace the characteristics that it seeks and the signatures in the information that it checks in each case separately, also taking into account the passage of time. In this case the investigation plan will change periodically as the understanding of the reality and of the information changes.

The Information Infrastructures in the Intelligence Enterprise at the Present Time

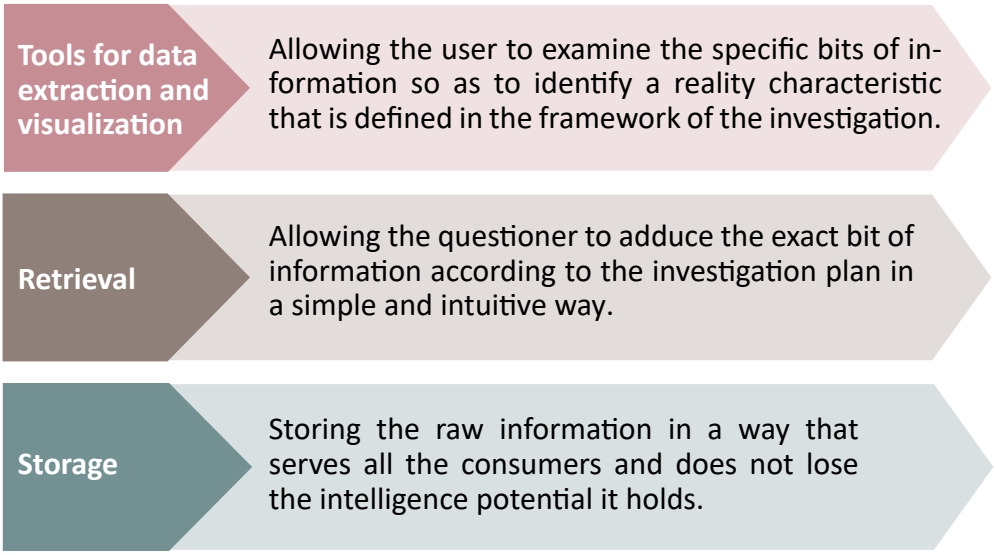
Intelligence theories and work plans require a technological infrastructure for their implementation. From the technological standpoint, implementing investigation plans entails identifying the relevant information in the organization's databases, questioning it in a way that will accurately reflect the investigative idea (by obtaining the precise information that constitutes the signature in the information, as defined in the investigation plan), performing different manipulations of it, and finally examin-

ing it to determine whether the signature appears in it as required by the investigation plan. I will refer to this entire process "data extraction".

In the big-data era, given the quantity, variety, and pace of the information alongside the numerous and diverse tasks, considerations of resources and cost-benefit ratios entail that the technological infrastructure cannot be ridigified to extract data for every purpose. Hence there is a vital need for technological infrastructures for data extraction that will give the different consumers of the information a great deal of freedom of action. They will then be able to adapt the above-described process to their needs in a precise fashion, in accordance with the work plans. In addition, because sometimes the validity of a theory about the information (that is, a definition of a signature in the information as representing a phenomenon in reality) can be prolonged, the information consumers must be able to "fix" the process that they produce as much as they wish, in a way that accords with the dynamism of the information in the arena, and to use this asset again and again.

In order to understand what technological components are needed to enable such flexibility throughout the intelligence-information endeavor, it is worth depicting the stages of consumption of the information during data-extraction processes. Such a depiction appears in this diagram:

Fig. 37: Stages of the process of extracting the intelligence data in the big-data era.



Thus the extraction process relies on a layered functional structure, with each layer based and dependent on the layer below it. Such a description explains why, for freedom of action and flexibility at the upper floors, flexibility at the lower layers is a precondition. In the old era, in light of the relative stability of the information world and the way reality was represented in it, such layers were built rigidly so as to enable specific checks of separate, predefined signatures. These structures resulted in "silo-like" systems that could interrogate a specific database in a specific way, and present the data in a manner that was specific and predefined. However, when the information world changed and the way in which different phenomena appear in it changed, the systems were not suited to the extraction processes that were desired, and great technological investment was needed to carry out adaptations - in the form of the database, in the retrieval capability, and in the forms of display of the information for the user.

A flexible approach entails that the technological layers do not constitute an obstacle to extracting present or future intelligence potentials, and that the responsibility for defining the exact form of extraction will be transferred to the user:

- At the layer of the databases, this involves the desire to keep the raw materials in their most complete raw form and not change or give up parts of the information, such as by adjusting data structures so that one can "fix" a certain intelligence data structure (and thereby also give up parts of the information that do not belong to the standard used). Such a structure is defined according to a specific form of consumption and is not aware of any potential for extraction that has not yet been considered and, hence, is not required.
- In the retrieval layer, it involves giving the intelligence consumer a maximal capacity for expression so that he can precisely express the intelligence idea he is trying to implement as a basis for adducing the exact information from the databases. This capacity for expression must meet new tests: that it provide exhaustive coverage of all the requests for data retrieval for each question and intelligence need; and that it not be overly technical but, instead, adapted to the intelligence consumer as its main user. For example, the SQL retrieval language allows very high flexibility in gaining access to the information, but using the language at high levels requires rigorous technological training that people who extract data do not necessarily possess. At the same time, reports prepared in advance (that run an SQL code) are indeed a product that intelligence people can consume easily. However, having been prepared in advance and "rigidified", they do not grant the intelligence user freedom to express the intelligence idea he has brought to the extraction task, but only the idea that was incorporated in the reports in the first place.

- At the layer of visualization and illustration, a flexible approach entails the need to give the user platforms for self-production of modes of presenting the information, for using a tool of filtration or analysis, and for choosing the appropriate "language of representation" in order to identify the signature. For example, the user must be given a capacity for independently generating graphs of different kinds, for geographic presentation through the use of a tool of analysis and illustration, for presenting a text and analyzing textual data within it, and so on. In this case, too, certain systems have structured and rigidified a certain mode of presentation and illustration that is intended to identify a specific characteristic in a specific way, and thus have denied users the ability to investigate the data in a manner that suits the needs of their specific task of extraction.

Achieving an operational optimum requires understanding the effects of creating such flexible and generic infrastructures. Because of the quantities of information, the information era creates real difficulty in retrieval and in performing technical calculations during the processes of retrieval and extraction. Technological solutions usually involve creating rigid patterns for the information, which accord precisely with consumption and thus improve the performances. However, as noted, the price of these solutions is that they reduce flexibility for the questioner. Hence there is a need to formulate a strategy of work processes that perform optimizations wherever needed, leave room for flexibility wherever possible, and generally are in ongoing dialogue, with high awareness of the implications of the mix of storage methods for the needs of the intelligence endeavor.

Semantic Retrieval

In the context of the above description, it is worth elaborating on the difficulty that exists at the retrieval stage and on the idea of semantic retrieval, which can offer a solution to the difficulty. The task of retrieving accurate and relevant information for a specific intelligence purpose is especially difficult in the Information Age because it entails a built-in tension between three components:

- The huge variety of the information;
- The variety of the intelligence questions, the investigation plans, and the signatures defined in their frameworks; and
- The need for the thorough technological training that is required to retrieve data capably and accurately.

Because the role of the one who extracts the information is to identify characteristics of reality based on phenomena in the information, he must combine expertise in identifying phenomena in reality and in information (an intelligence and disciplinary

expertise) with the ability to retrieve numerous and varied data from a vast quantity of databases that are very diverse in structure and content (technological and "informational" expertise). Experience shows that the effort to train functionaries who can operate within this tension does not succeed. Instead the functionaries are drawn to one of the extremes while forgoing the expertise required at the other extreme, or they make use of technological expertise in the SQL language, in the databases themselves, and in the innumerable data structures, dealing with the information layer in its technical sense while neglecting their full intelligence and disciplinary expertise and detracting from their relevance to the intelligence task. Or they may deal with the intelligence aspects but keep a distance from the complexity of the information world with its nuances, thereby wasting the opportunities if offers for their purposes.

This gap cannot be bridged by transferring the responsibility to the functionaries themselves. From a cognitive standpoint, it is impossible to encompass the full intelligence potential that lies in thousands of types of information to be found in the intelligence endeavor, and at the same time manage to devise an exact solution for each specific intelligence need. The idea that could contain the key to overcoming this gap and enable the organization to achieve optimization in the way in which it constructs itself for the tasks of information retrieval - given the diversity of sources and types of information, and of questions and kinds of extraction - is the idea of semantic retrieval. Such retrieval entails consuming information not according to its physical form in the databases (i.e., tables, records, data platforms) but, instead, in light of its significance for intelligence. In this approach, when it comes to identifying a certain signature in the information with regard to contacts, which are scattered through numerous tables of raw data, we will not want the functionary to be responsible for knowing where such information is located

Semantic retrieval entails consuming information not according to its physical form in the databases but, instead, in light of its significance for intelligence. In this approach, when it comes to identifying a certain signature in the information with regard to contacts, we will not want the intelligence worker to be responsible for knowing where such information is dispersed but, rather, to be able to inquire, and the intelligence-technological infrastructure will have to mediate his request for all the relevant information

but, instead, to be able to inquire about "contacts" in general, and the technological infrastructure for the preliminary inquiry will have to mediate his request for all the relevant information.

Such ideas are not new either in intelligence or in the civilian sector. The idea of preparing specific databases for business or intelligence purposes has long been in existence, and indeed well exemplifies the rigidifying approach - that is, preparing a table of data that suits the mode of consumption. The principle behind the idea of semantic consumption as something relevant to the information age is that we are unable to foresee how the person extracting the information will want to retrieve data. What is required, then, is a generic semantic approach that will suit every purpose, based on a prototype of the form of access to the information and not on concrete and specific needs of access (that is, reports that can already be written).

What, then, is the prototype of information consumption? When we access information, any information, we do not regard it as a potpourri of undeciphered data but, instead, anchor its significance to a network of the entities and the relationships within it. This is not by chance: the information itself results from a process of documentation (human or mechanized) of occurrences in the world that can themselves be anchored to a network of entities and relationships.

For example, contact information about telephone conversations between consumers of a cellular service provider constitutes documentation of a phenomenon in the world in which person X has telephoned person Y. The informational representation of the phenomenon can be a table of telephone calls in which each record contains the identification of the two sides (their telephone numbers) and the date of the conversation, and the record as a whole represents one call. Specifically, this is a list of telephone calls between telephone numbers; generically, however, it is a list of contacts between entities.

The idea of generic semantic retrieval is based on this approach of identifying and characterizing the content of the information by using a model of entities and relationships. If we invest in identifying the basic elements of the databases - the entities and the relationships - we can show the user an ability to consume information according to its intelligence significance, not according to its physical structure. Such an approach can be facilitated through a process of semantic mapping - a process in which one draws links between every column and table in the database to produce an assemblage of intelligence entities and relationships in a semantic data model. Such mapping makes it possible to produce technological networks that can mediate between a semantic query in an intelligence language (entities, contents, and relationships that have intelligence significance) and the technical queries that are needed to retrieve the required information from the databases themselves with their huge

diversity.

For example, a user can carry out a query that is formulated as follows: all the people having a telephone that made more than 30 calls to a list of 30 implicated people, and that also was located near an Israeli community during the preceding week, and who have a first-order relative who has been detained or arrested by Israel in the past. This query could automatically be translated into a mechanized process of retrieval and fusion of data from all of the organization's databases that contain information about people, telephones, ownership of telephones, communities, locations, conversations, and arrests. Even if the raw information is dispersed among dozens of different databases, in different formats and with different technologies, semantic mapping makes it possible to filter this complexity and obtain results from the standpoint of the questioner's intelligence language.

Such an approach to data consumption provides flexibility and very broad freedom of action for a plethora of changing intelligence needs. In addition, it is not rigid and limited like reports that are written in advance and return solely the answers that were defined for them; nor does it require deep technological understanding for the writing of complex retrieval queries. Likewise, because it performs filtration and simplification for the consumer, it does not require highly detailed knowledge of each database at the information level. Of course, creating a semantic-retrieval layer is not cheap from either a technical or intelligence standpoint (given the need to perform a mapping of the information). In the big-data era, however, such an approach constitutes a very appropriate break-even point that allows great flexibility in the face of the changing needs and information.

The Source Infrastructure in Intelligence Work

Underlying the data-extraction processes that constitute the atomic stages of the investigation and consolidation processes is the information that forms the foundation of the intelligence organization. The organization does not have special value in itself, but only insofar as it serves the extraction processes that are relevant to the investigation and, thus, to the intelligence solution that is offered with regard to the needs and the information that have been specified. This requires a well-ordered strategy for deciding on the prioritization and characterization of the types of information that must be obtained, and hence, also, on the technological force-buildup processes and the concrete work plans for obtaining the information.

It is worth recalling that the information required for the work processes (investigation and consolidation) in the intelligence organization is chosen specifically as reflecting certain characteristics of reality that we will want to identify. Hence the intelligence requirement, and the investigation or consolidation plan, are what

should dictate which information is sought and at what pace it will be obtained. The more that the space for activity is diverse and flexible, the more diverse and dynamic will be the requirements concerning the sources. At the same time, processes of developing sources, and of obtaining and conveying information, usually entail a great investment of resources and a lengthy process of development and stabilization. Hence, at this level, too, there is a need to balance between, on the one hand, the development and deployment of rigid collection tools and sources, for which the yield and relevance of the materials they obtain justify the time and cost of developing and deploying them, and, on the other, more generic and flexible collection platforms that generate new balance points. This requires sufficient investment in the development of platforms whose final and exact orientation, given the needs of the work plans, can be achieved at a relatively low additional cost.

Sources that are built in a flexible and versatile fashion indeed make it possible to transfer the control of the collection processes to the intelligence personnel, who in any case are the ones who best know what pieces of information they need to promote the investigation or consolidation process. Such an approach, if based on a sufficiently robust infrastructure, makes it possible to free up technological bottlenecks of control over the sources, and to transfer the authorities and the bulk of the work to the intelligence-creation processes themselves, while improving the congruence between the intelligence need for information and the concrete actions to obtain it. At the same time, such an approach requires the training of intelligence-operational manpower that can carry out these tasks optimally in the framework of the work teams. Despite the extra investment needed to establish accessible platforms and not specific tools, and to train manpower to operate these platforms, such an investment enables a significant shift of the operational balance point in the world of the sources to a domain that will allow much greater functional flexibility, and thus also enhanced congruence and fuller exhaustion of the information potential in the world of the sources.

Organizational Structures in the Intelligence Endeavor

It is clear that the flexible methodologies for intelligence practice, primarily the analytical-intelligence approach, require organizational structures and the development of professions that foster the flexible, goal-oriented behavior that is inherent to these approaches. This contrasts with the rigid and contractual organizational structures of the past, particularly those designed to filter and utilize raw materials (such as "silo-like" audio-production efforts), which are not well suited to the flexible concepts of work that are developing at present.

Hence it appears that the organizational frameworks best adapted to the current era are those that are multidisciplinary, oriented to an intelligence task, and organized ac-

cording to an intelligence work plan that is based on the investigation method. These frameworks are precisely suited to the task and to the work plan that is formulated, and it is supposed to be possible to establish them and dismantle them as required when their tasks have been completed and new tasks and work plans are formulated.

At the same time, an organizational idea of this kind creates other difficulties. From the standpoint of the intelligence endeavor, the frequent dismantlement and formation of ad hoc work teams, without continuity or a defined intelligence focus, can hinder the deep, ongoing development of knowledge about the nature and components of the intelligence arena. From the professional-disciplinary standpoint, independent work in the framework of combined investigation teams could divert the disciplinary expert (having audio, textual, VISINT, or other expertise) from his professional focus, leading him away from a critical mass of other experts in his discipline who will no longer be working alongside him as in the contractual efforts. Such a situation could lower his professional level, weaken professional mechanisms of criticism, and retard his progress in the profession.

To provide a solution for these issues, organizational structures are now emerging that are intended to realize all the advantages. Such an approach is now being tried in Aman as an intermediate structure, and appears that it will provide a solution both for the need to organize in task-oriented combined teams and for the need for a definite, ongoing intelligence foundation, while also enabling sufficiently broad professional

backing for the different disciplines. This approach is based on a relatively large organizational structure with dozens of functionaries, which is directly subordinate to the arena structure. This is essentially a sub-arena structure that deals with a field and a campaign that are well defined in the arena. In addition, this suborganization is characterized by a unified command and by resource autonomy.

Conclusion

In the current era, the challenges involved in building an intelligence organization on an operational scale are greater than ever. The risk of moving to one of the extremes - too much efficiency, which will detract from adaptation and flexibility, or too much flexibility, which will require more resources than are available - is higher than ever. In this article I have surveyed the different aspects of the intelligence mechanism's activity, and have given them forms that will make it possibly to identify and perhaps even adopt approaches to practice and force buildup that are suited to the degree of efficiency or flexibility that is required from the perspective of the force-buildup personnel.

Alongside devising processes and a strategy for force buildup, we need to create control and inspection mechanisms that will constantly focus on the task we are engaged in: Are we too rigid in places where flexibility is called for? Or might we have developed mechanisms that are too flexible and turn out to be extraneous and wasteful, and actually could be made more rigid? All this is required if we are to maintain the same optimum of operational activity.

This optimum point is always in motion. Because, in the current era, the cost of an error is rising steeply, the force-buildup processes of the intelligence endeavor must be navigated meticulously and adroitly in order to maintain the balance of resources and investments in relation to the benefits to be derived, within the narrow boundaries of the price that can be paid.

Approaches to Intelligence Research in the" Post-Truth "Era

Brig. Gen. (res). Itai Brun
former head of the Research Department of the Military Intelligence Directorate

Introduction

Intelligence, according to a widespread definition, is knowledge about the enemy and the environment that is requisite for decision-making. Intelligence research is supposed to develop this knowledge and thereby help decision-makers understand the present and think about the future. How is this knowledge developed? What is the methodology that fosters its development? How should theory influence the practice of intelligence research? What is the role of the "truth" in the effort to develop intelligence knowledge? And what is the role in this knowledge development of the huge databases (i.e., big data) that are accumulating among the intelligence organizations?

To begin with, it needs to be asked whether intelligence has a theory or a methodology at all. A 2002 article by Michael Warner described the problematic theoretical basis of intelligence. Although its title concerned the lack of an agreed definition of intelligence, the article dealt mainly with the lack of a theory. "In a business as old as recorded history", Warner wrote, "one would expect to find a sophisticated understanding of just what that business is, what it does, and how it works. If the business is 'intelligence', however, we search in vain... So far no one has succeeded in crafting a theory of intelligence". Warner's article is still relevant, and current books about intelligence also call for the formulation of such a theory, which still does not exist.[113] This state of affairs also affects the methodology of intelligence research. This article seeks to contribute to the discourse about intelligence research by depicting three widespread, typical approaches to research practice, which can be called: the educative, systemic, and scientific approaches. The article characterizes the approaches, identifies each one's theoretical basis, and highlights the differences between them. The main conclusion reached is that the approaches are distinguished primarily by how they perceive reality - its existence, its attributes, and the proper way to clarify and understand it. These approaches represent three fundamentally different worldviews concerning the nature of intelligence. They also differ from each other in how they view the benefit to be derived from big data.

113 Michael Warner, "Wanted: A Definition of Intelligence". Studies in Intelligence 46, 3 (2002): 15-22.

Intelligence research is at the forefront of coping with the changing reality. There is not, and apparently also cannot be, a single good way (an approach or a method) to contend with the profound uncertainty that exists regarding the enemy and the environment. Although this does not mean "Anything goes", the state of affairs is quite close to that. Hence, the three approaches described here certainly have a place in the research organizations. Moreover, combining them can foster a better and more complete understanding of the reality. Nevertheless, in my view there is a clear advantage, especially in this era of "post-truth" and big data, in placing the approach described here as the "scientific approach" at the center of the research effort that seeks to clarify the complex reality.

The Educative Approach

In the movie *The Hunt for Red October* (1990), Jack Ryan, the CIA analyst, discovers that the Soviet submarine captain Marko Ramios is planning to defect to the United States with his nuclear submarine and its crew. In an unforgettable scene, this possibility arises in Ryan's (played by the young Alec Baldwin) mind during a discussion hosted by the president's national security adviser and attended by many. The discovery does not emerge from an orderly process but as a sudden thought, when a series of details fall into place for Ryan in a single pattern that, he believes, points to the right solution for the intelligence question. When one of the generals sitting around the table asks how he can know what is happening in the Soviet captain's mind, Ryan replies with great vehemence, "I know Ramios".

During my service I saw not a few research personnel who tried and even succeeded to understand the other side's decision-makers and predict their moves because they "knew them"

This scene is, of course, completely fictional, but it is also a classic example of a research approach based on an effort "to get into the enemy's mind" by becoming very knowledgeable about him. During my service I saw not a few research personnel who took this approach, and seemingly even succeeded, just like Jack Ryan, to understand the other side's decision-makers and predict their moves. More than once, what happened in reality was very similar to what happened on the movie screen.

Yehoshafat Harkabi, a former head of Aman, represents this approach in the most explicit fashion when he describes "human intelligence". He refers to it as:

the intelligence that has penetrated and reached the person, his thoughts, his attitudes, his moods and reactions, and the intelligence that can assess the person, the adversary, the enemy, that can assess him not only in a numerical way but qualitatively.

Human intelligence is the intelligence of sensing, a kind of sensing that is achieved after many years of work and of studying the adversary. In this kind of intelligence, we do not just posit the ordinary human being and perceive the adversary in his image; instead we penetrate his thoughts and see him as he is.

Thus Harkabi describes an ability to "penetrate the thoughts of the decision-maker and to know him even better than he knows himself". This ability is based on profound, long-term familiarity with the enemy. It leads to the kind of "intelligence of sensing" that was so well epitomized cinematically by the process Jack Ryan underwent during the discussion hosted by the national security adviser.[114]

This is a classic positivist approach, and its roots are in the humanities and social sciences. It maintains that a single reality exists, which is an actual situation, one with a regularity that can be discovered. According to this approach, intelligence researchers can discern this regularity and thereby understand the reality and even predict the decision-makers' future forms of activity. According to this approach, the main method for understanding the reality is the intelligence researchers' profound familiarity with the enemy and the environment, which is grounded in meticulous study of the past and the present, intimate familiarity with the enemy's culture and language, and painstaking reading of relevant texts. Intelligence researchers who favor this approach believe that this familiarity enables one to "get into the enemy's mind" and decipher his inmost thoughts.

Nowadays there is an updated version of the educative approach that tries to identify a regularity in the data that patterns the activity of the enemy. This approach views the technological developments as offering new possibilities to enhance the familiarity with the enemy, which can now make use of enormous databases about the enemy and of the concomitant statistical analysis

This is sometimes described as an inductive approach to developing intelligence knowledge. Here it is called the "educative approach" because, if reality is unitary and objective, it is very logical for intelligence research personnel who favor this approach to assume the role of "educators" of the decision-makers, clarifying for them the reality that they do not always want to recognize. Yehezkel Dror describes this dynamic: "If the leaders are inclined to misunderstand the reality…one has to provide them with accessories that will help them recognize the reality and under-

114 Yehoshafat Harkabi, Intelligence as a State Institution (Maarchot and Intelligence Heritage Center, 2015). (Hebrew)

stand it. This is one of the roles of the intelligence assessment officers, especially the senior one among them... They must serve as educators of the leaders so that they can better understand the reality and its dynamics, including its future development".[115]

The educative approach regards the profound familiarity as lying in understandings about the research objects that stem from knowledge of their culture, language, and history. In the big-data context[116] there is now an updated version of the approach, very different in its attributes but similar in essence, which tries to identify a regularity in the data that patterns the activity of the enemy. This version is grounded in the same basic notion that such regularity exists and that the technological developments offer new possibilities to enhance the profound familiarity with the enemy, which can now make use of enormous databases about the enemy and of the concomitant statistical analysis. According to this version, the larger the database the greater the ability to identify the laws in question. Hence this version of the approach invests great effort in developing the ability to analyze - utilizing the vast quantity of data in the databases - the patterns, types of activity, and contacts that can help in discerning the regularity. It also posits that a deviation from the regularity will be detectable as an anomaly.

Despite its occasional successes, the educative approach's ability to serve as a leading and systematic approach to intelligence research is very doubtful. It has many shortcomings, methodological and practical. First and foremost, the basic assumption that future modes of activity are determined by those adopted in the past is problematic and misses the dynamism of human life. People make decisions in response to the changing reality and not infrequently surprise even those closest to them when they alter their perspectives and diverge completely from their previous behavior patterns. But that is not the only problem with this approach. In the complex world of today, it is very doubtful that positioning "decision-makers" at the center of research thinking can successfully address the range of factors that affect how the reality develops. A further problem with this approach, well illustrated in the movie, is that it is not subject to supervision and criticism. That is, it is not possible to examine how the researcher reached his conclusions. Intelligence research is both an art and a skill based on systematic work, but this approach takes it too far in the direction of an art that relies on the ingenuity of particular researchers.

In my view, essentially this approach succeeds (and, again, it succeeds to a considerable extent) until the decision-maker on the other side changes his outlook, or the reality becomes more complex and makes his outlook less relevant. It is then

115 Yehezkel Dror, "Intelligence as an Educator of the Leader". *in Intelligence and the Leader* (2004), 19-25. (Hebrew)

116 Michael Milstein, "It Won't Change…It Changed, It Will Change". Intelligence: In Practice, no. 2, 2017, 59-67. (Hebrew)

more prone to error than other approaches. The people who are closest to us surprise us, despite our deep familiarity with them and their behavior patterns. Thus a more systematic methodology for clarifying and understanding the reality is needed.

The Systemic Approach

In the second half of the 1990s, a new set of concepts, unknown until then, began to penetrate the Israeli intelligence discourse. It entailed the frequent use of such notions as "systemic intelligence". "knowledge development". "systemic frameworks". "discourse". "conceptualization". "intelligence campaign". "adversary system". "context". "systemic idea". "rationale". "intelligence superiority". "relevance". and "tensions". Against this backdrop, a new approach emerged. It questions some of the basic understandings that have underpinned intelligence research, particularly regarding the nature of intelligence at the higher levels.

This can be called the "systemic approach". In describing it, Aharon Ze'evi Farkash and Dov Tamari say that the intelligence for a campaign "is not a set of facts but, instead, a subjective observation and understanding of the emerging phenomenon, which is produced in tandem by intelligence and the commander (or commanders) of the anticipated campaign". In their view, "The purpose of the intelligence for a campaign is to provide the commander of the campaign and his superiors with the conditions with which to apprehend the adversary and our side in a single system, which enables self-scrutiny in light of a new phenomenon, in a new context of the emerging crisis".[117]

The systemic approach asserts that in the systemic and strategic world (in contrast to the tactical world), the process of developing intelligence knowledge is one of creating or constructing a new reality and not the reflection or discovery of an existing one. Proponents of this approach do not regard the "truth" as a main criterion and replace it with another idea, that of relevance to the unique context

The roots of the systemic approach lie in constructivism and in the theory of complex systems. In its more extreme version, it asserts that in the systemic and strategic world (in contrast to the tactical world), the process of developing intelligence knowledge is one of creating or constructing a new reality, not the reflection or discovery of an existing one. The intelligence research personnel who favor this approach believe

117 Aharon Ze'evi Farkash and Dov Tamari, **And How Will We Know? Intelligence/Operations/Statesmanship** (Tel Aviv: Sifrei Aliyat Hagag, Hotza'at Yediot Aharonot, 2011). (Hebrew)

that there is no objective reality that exists separately from the consciousness of the people who deliberate about it. The reality, in their view, is fictitious and primarily subjective. They do not view research activity only as an attempt to understand entities that exist beyond the border, and they do not measure their success in terms of the completeness of the intelligence picture. Even less do they regard the "truth" as a main criterion. They replace it with another idea, that of relevance to the unique context.

This approach is based on an accurate understanding of the deep connections between one's own side and the other side, on seeing them as a single complex system. What is required, then, is to create the conditions for study and understanding of the unique context in which decisions are made, primarily by crafting the conceptualization of both our own forces and the enemy. The research process emphasizes the need for joint understanding, in contrast to other approaches that emphasize competition. The intelligence research personnel who favor this approach see themselves as an integral part of the circle of decision-makers, and they view intelligence knowledge as a main component of operative and strategic knowledge development. This approach jibes well with approaches that regard intelligence as a main factor in influencing reality.

This approach, too, has had some successes, and undoubtedly has made a very significant contribution to the current intelligence concept. In its less extreme version it has fostered a fundamental and appropriate change in the understanding of the role of intelligence research personnel as partners in policymaking, in planning operations, and in force buildup. It has also contributed to the analysis of the enemy and the environment as a complex and dynamic system, and to the importance now attributed to analyzing the issue of the concrete context of events.

At the same time, this approach too, in my view, is unsuited to serve as a leading and systematic methodology for intelligence research. It has three shortcomings that distance the research personnel from the truth and from the effort to discover it. The problem begins when the concern with conceptualization and "subjective understanding" comes at the expense of the effort to clarify the complex reality; it continues when the research personnel's great concern with the effort to influence the reality leads to allocating less attention to considering the reality itself; and it intensifies when these processes lead to a change in the professional identity of the research personnel, who cease to regard themselves as leading the effort to clarify the reality.

The Scientific Approach

In the lobby of the original CIA building in Langley, Virginia, the head of the agency from 1953 to 1961, Allen Dulles, emplaced the maxim: "And ye shall know the truth and the truth shall make you free". The maxim is from the Gospel of John (8:32), and it was also chosen by Johns Hopkins University and other academic institutions as a motto that expresses the importance attributed to the idea of "discovering the truth" in a democratic society.

The effort to discover the truth is indeed common to intelligence research and to science. Isaac Ben-Israel is the classic representative of the "scientific approach", which believes that intelligence should adopt the approach to research that science pursues As he remarks: "In human culture there is an additional institution for clarifying the reality, which has already been established for several hundred years, namely, science... For several hundred years science [meaning natural science] has accumulated great experience in deriving scientific predictions from empirical findings, that is, from observations about nature... Therefore it appears that it is worth studying the philosophy of science in order to draw conclusions from it about the philosophy of intelligence".[118] The scientific approach to intelligence research is based on critical realism and its roots are in the exact sciences. It maintains that there is indeed a reality, independent of our thoughts about it, that intelligence research can clarify and even understand. But proponents of this approach are well aware of the possible biases that can occur both when obtaining the raw material and when processing it. Hence they believe that the knowledge they possess consists of hypotheses that are always subject to scrutiny and criticism. In their view, indeed one cannot achieve perfect objectivity in understanding reality, but one can approximate it. Those who favor this approach regard its main tools as similar to those on which the concept of "scientific investigation" is based, primarily the elements of raising hypotheses, perpetually casting doubt, and testing the hypotheses through ongoing debate.

A serious debate cannot be held in a place where there is a single opinion, a single explanation, or a single possibility. Therefore, at the center of the scientific approach to intelligence research stands the idea of "competing possibilities". This idea promotes a research process that creates a basis for debate by posing a wide variety of explanations (regarding the present) and possibilities (regarding the future). Most important, the idea helps bring the assessment process to light, while adopting a clear standard in a way that enables extensive evaluation and criticism of the basic assumptions and the research process. At the same time, presenting the additional pos-

118 Isaac Ben-Israel, *Dialogues about Science and Intelligence* (Tel Aviv: Maarchot, 1989) (Hebrew); Isaac Ben-Israel, *The Philosophy of Intellligence* (Tel Aviv: Misrad Habitachon, 1999) (Hebrew); Isaac Ben-Israel, "Intelligence as an Institution for Clarifying Reality". *in Intelligence and the Leader* (2004), 68. (Hebrew)

sibilities in no way eliminates the role of intelligence research in pointing to what it considers the most reasonable explanation or possibility when doing so is feasible.

Thus, according to the idea of "competing possibilities", the research process begins with posing explanations and possibilities and continues with discussion of these in light of the information that the research personnel possess. The idea is applied differently to the different groups of research questions. The more one moves on a spectrum from "secrets" to "mysteries", the less important the information and the more important the tools and methods that can assist in understanding the present and pondering the future. Here it is appropriate to use such tools as "war games", "red teams", "playback", "scenarios analysis", and even "crowdsourced intelligence". These tools increase the chances of overcoming the problems of failure of imagination and sticking to a conception, which are the main causes of failure in the effort to clarify reality.

The scientific approach sees big data as a new and exciting tool that in some cases can help in deciding between the competing possibilities or in ranking them by probability. The point is not necessarily to pose the possibilities themselves; the idea of competing possibilities entails a creative "green light" to extend research thinking to a search for explanations and possibilities, which can be conducted apart from the information or even in its complete absence. After the possibilities are posed, the scientific approach raises questions and tries to discern in the information the patterns, modes of activity, and relationships that are of high "diagnostic value". What counts here is the ability of the information to help in deciding between the competing possibilities or in ranking them by probability. In the view of the scientific approach, then, the key to utilizing the enormous databases lies in asking the right questions and conducting a critical discussion of the search findings.

The scientific approach sees big data as a new and exciting tool that can help in deciding between the competing possibilities. After the possibilities are posed, the scientific approach raises questions and tries to discern in the information the patterns, modes of activity, and relationships that are of high "diagnostic value"

Conclusion

The role of research personnel does not consist only, of course, in clarifying the reality, understanding it, and presenting it to the decision-makers. Because of their unique

position, because of the knowledge they have, and because of their ability to develop new, relevant knowledge about the enemy and the environment, research personnel are also deeply involved, to the point of full partnership, in processes of policymaking, of operational planning, and of force build-up at the different levels. However, their main advantage, in my view, is their ability to characterize the enemy and the environment at the decision-makers' table and to lead the ambitious effort to clarify the dynamic and complex reality.

In my eyes, the scientific approach is preferable because of its methodical nature, but primarily because of its basic stance toward the issue of reality and the possibilities of clarifying and understanding it. This pertains all the more to the present period, in which research personnel must contend not only with the uncertainty harbored by the dynamic reality but also with deliberate attempts, at home and abroad, to undermine the ability to clarify and understand the truth. Furthermore, they have to cope with an ongoing effort to question the *need* to understand the truth. The post-truth era thus puts intelligence research at the forefront of the endeavor to clarify the truth, and makes it a major representative of the truth to the public and the decision-makers.

According to the scientific approach, the reality that intelligence research deals with is not a trick of the imagination. It is a real entity that can be understood and described. This is the case even if the researchers' view is strongly influenced by their experiences, background, basic beliefs, the information they have, and by a long series of perceptual distortions, which probably cannot be corrected in any case. And indeed, in a world that is chaotic, frenetic, and full of contradictory interpretations, intelligence research personnel continue to search for truth. This truth - "intelligence truth" - is not absolute truth, since it will never be possible to know everything. There will always be a gap between reality as the intelligence researchers understand it and the reality itself. There will always be doubt as to whether they really understand the reality. At the same time, however, intelligence truth is not a fiction like the concept of truth in other fields of knowledge. The clear goal of intelligence research must be to come as close as possible to an accurate description of reality with regard to the enemy and the environment. Making the scientific approach the centerpiece of the research endeavor can help intelligence research personnel to avoid errors in this endeavor. It focuses the research personnel's identity on the task of discovering the truth and clarifying the reality, and gives them a methodology that can contend with the difficulties inherent in this ambitious task. It is also more open to the inclusion of other approaches as important means of posing competing possibilities.

Technological developments have led to a situation in which intelligence organizations, like other entities, hold information that has accumulated in huge databases. When it comes to intelligence questions of the "who", "whence", "when", and even

"how" types, these databases have huge advantages for intelligence, which have already been proven. But on the broader questions of clarifying reality, which this article addresses, the approaches are distinguished also by how they view the benefit to be derived from the databases. On this issue as well, the scientific approach, in my view, should be preferred. It allows one to investigate the information in light of competing possibilities, in a way that will, as much as possible, facilitate the discounting of errant possibilities and the espousal of valid ones.

The Approach as a Guide to Technological Intelligence Force Buildup

D. P. - until recently a senior director
in the Israeli Security Agency

The first issue of *Intelligence - in Practice* dealt with the different aspects of **jointness**, and for good reason. Already for a considerable time, the challenge of jointness has not involved merely another level of organizational excellence or a qualitative expression of a solidary culture. It is a necessary outcome of the changing and dynamic reality, which demands that all the elements of knowledge and ability, both within and between the organizations, be mobilized. One of the main reasons for the urgency and necessity of maximizing jointness is the global technological reality. Borderless networks enable a capability of development and learning that the adversary enjoys as well. This capability, which is almost equally available to all players, steadily erodes the traditional relative advantages that the intelligence organizations maintained in the past.

Despite the positive developments in recognizing jointness as a key element of the organizational culture, individual organizations have difficulty generating synergic approaches and successfully and fully implementing technological force-buildup processes

Jointness, then, is an endeavor for which the organizational culture, the combat doctrines, the interface between organizational structures, and the processes of force deployment and force buildup are mobilized to maximize the aggregate knowledge and capabilities in a way that will enable ongoing, successful coping with change while preserving Israel's qualitative advantage over its adversaries. It indeed appears that among the community's organizations, the conceptual change has passed the requisite threshold and is now being practically instituted after years of "talk". The change is occurring mainly thanks to a clear, transparent conceptual discourse that percolated from the top down, with strong support from the chiefs of the intelligence organizations. The field echelons view this discourse as authentic and as promoting phenomena of integration and synergy in the various niches of the organizations.

Despite these positive developments, as the barriers between organizations are lowered, individual organizations have difficulty generating synergic approaches

and successfully and fully implementing technological force-buildup processes. The problem does not lie in absorbing technologies, or in the development capabilities themselves; the technological components of the community's organizations are vigorous and creative. However, when one considers the full chain of value, which is measured at the last link of the chain, when one analyzes the general organizational ability to apply advanced technological capabilities, when one assesses the effectiveness and efficiency of incorporating a new capability - and not only among a handful of experts or an ad hoc team of experts formed to tackle a particular task, but, rather, in large-scale scale organizational processes - then the gap opens. It then becomes clear that the advanced and creative operational capability, which may have been successfully applied at the outset in a local pilot, is not being assimilated and effectively implemented among the field units, and indeed is not fulfilling its purpose.

The gap, then, is not found in the technological units or the field units, but in the full range of activity that begins with force buildup and ends with force deployment. It is no secret that in the decentralized world in which most of the organizations operate, when it comes to technological force-buildup processes, we will almost always find gaps of efficiency and effectiveness regarding technological capabilities when we reach the end users, despite the outstanding personnel who work in these units.

The reasons for this state of affairs are many:

- Difficulty in exactly defining the requirements, which already creates, at the outset, a gap in meeting the expectations. If in the past any requirement would be likely to "hold water" for a long period, today requirements are unstable and change rapidly.
- A growing diversity between the different arenas and hence a diversity of requirements between the different units of a given organization. Thus one hears within the organization multiple voices of consumers of the same development resource.
- A diversification of specializations that hinders the building of structures that will last over time. Every day new professions develop that further stretch the professional common denominator. Because of the technological and professional focus, organizational structures are devised that consist of small islands of expertise, whose ability to wield influence laterally is more limited.
- Competitiveness and a desire to safeguard one's "spiritual ownership" and credit sometimes detract from the ability to involve all the stake-holders in real time.
- Ongoing difficulty for the end users, who are replete with new technological capabilities, to absorb and assimilate at the required pace.
- The hindrances that were noted were situated all along the chain of value. Above all of them, however, is a much more basic and consequential problem: the weakness of the approach at a time when the organization's existing characteristics

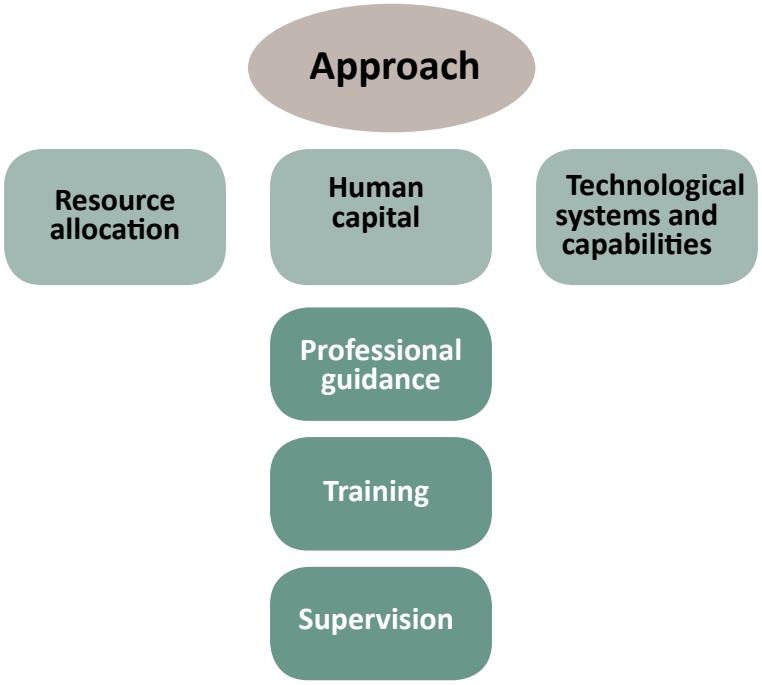
erode the commitment to it.

The professional approach requires a practical definition of the direction of professional activity, in a way that will make it possible to assess the challenges posed by the present and future reality. An example of a professional approach from the world of intelligence content is that of "unified production", in which the producer must be trained to handle all the kinds of communication used by the adversary. Only then can one produce a complete intelligence picture, and in a relevant amount of time, of the adversary's various kinds of communication. An opposing idea is that particular specializations are required to contend with the adversary's effective, high-quality production of all these kinds of communication.

The structural stage of defining the concept is critical because the force-buildup process is coherent and ongoing, hence requiring a flow of synchronized actions in order to succeed at the task. One can describe the process as having three main components, beginning with the definition of the approach:

The "head": One approach is adopted and formulated. The formulation of the ap-

Fig. 38: Between formulating the approach and the implementation and supervision.



proach includes the answer to the question, **why** must we act in X manner? (No matter how trivial it may sound, we must remind ourselves more than once that the answer to the question "Why?" must precede the answer to the questions "What?" and "How?")

The "body": The level at which the necessary resources can be defined. This concerns the appropriate human capital, the requisite technology, and how the resources are to be allocated. Clearly, this component is directly subordinate to the approach.

The "legs": The level of implementation, the combat doctrine, the procedure, the professional guidance, the training, and the supervision.

The approach is the "glue" that unites all the elements of the force buildup. It is the guide that enables the details to be synchronized. The definition of the approach is a critical factor for interorganizational communication that helps in overcoming decentralization and remoteness between units. The clearer the approach, the more freedom of action can be created down along the chain of value; everyone looks upward, understands "Why", and acts accordingly. The approach determines and affects how the technological systems are characterized. When it comes to arranging the requisite manpower and their mode of operation (in corresponding organizational structures), the approach certainly influences the professional guidance, the "How?" and hence also the training systems and the instruction involved.

The unified-production approach makes it necessary to characterize the systems that will be suited to the individual worker, who must deal with a wide range of types of communication. According to this approach, home systems will be required that will include and enable maximal connectivity between the different applications. At the same time, clearly this approach will reject any solution based on stand-alone systems, which will make the approach difficult to implement. At the center stands the producer who is responsible for an objective, not the professional expert in the field of WEBINT or SIGINT. In line with this approach, the candidates for recruitment will be characterized; and the manner of training them, along with the appropriate organizational structure in which they will operate, will be defined.

One can similarly view the developing field of VISINT, in which the competing approaches are the post-event intelligence approach, which involves an investigation, and the prior-intelligence approach (which collects and warns). Here too we will find that each conceptual decision will lead to a completely different path for the force-buildup process in the value chain as a whole. It is clear, then, that defining an approach generates a set of decisions, which must be coherent at all points of the force buildup.

Here it is worth considering a relevant example of a clash that is intensifying and developing before our eyes. Cloud technologies, disappearance of contents, big data,

AI, accumulative suspicions, and other concepts and challenges compel us to make a strategic decision between two approaches. The first approach posits, then, that technological force buildup requires a focus on data-based intelligence, since the emergent reality restricts access to communication contents while the global development of the data field offers a way to compensate for this gap. The contending approach maintains that we do not have the privilege to forgo an uncompromising effort to reach every possible piece of content related to the adversary's cyber SIGINT activity. While for purposes of this discussion it is not important who is right, it is clear that every force-buildup value chain is affected by such a conceptual debate.

Each decision will entail different directions of development, a different focus for investment, a different adaptation of manpower, and the creation of appropriate organizational structures, along with training and guidance frameworks. We must not adopt an "Anything goes" method and come out sorely lacking. The need to make an organizational decision on an approach to be adopted becomes critical when force-buildup personnel are decentralized and independently managed in different organizational frameworks. Unless a decision is made, the door is opened to competing approaches, which, as they develop, pose a danger to jointness and synergy.

The problem does not lie in competing ideas, in different and challenging thinking, or in creativity that permeates the organization from its margins. These are important and valuable. The problem lies in the fact that decentralized development units operate according to their own independence and authority and seek to implement their approach. Insofar as this is not coordinated with the force buildup in general, the predictable result is faulty performance regardless of the approach's appropriateness or congruence with reality.

Here I should note that the organizational culture must control and direct the creativity and free spirit of the development units, presenting the same compass for everyone. We must understand the inherent

Main aspects of approach- and jointness-driven organizational change

✓ Crafting the approach is what sets the force-buildup process in motion.

✓ Force buildup that is integrated and effective cannot operate under competing approaches.

✓ Joint crafting and formulation of the approach, as close as possible to the "production line", will improve the result significantly.

conflict between our desire, on the one hand, to develop innovation, initiative, and agility, and, on the other, to manage large units with a limited and regular range of "customers". Intelligence organizations cannot be startup organizations in the full sense of the term. Apparently a partial solution to this problem is to empower "startupist" R&D personnel and separate them from the general force buildup. Indeed, in an effort to contend with the phenomenon, all the units in the community have built and are building mechanisms intended to shorten lines of management and command, and to improve processes of direction and supervision. And yet still, paradoxically, the power of free creation and innovation, which is an organizational growth engine, can at a certain point become an obstructive factor rather than a propitious one. Hence we must focus on the question of who holds the keys to the approach: the force-building actor who has the responsibility for winning? The development actor who truly understands where the technological opportunities lie? Or the head of the organization?

In most of the organizations, the authority to set the approach is at the senior command and administrative level, and the lower floors have not reached a decision on the organizational approach. The organization has not yet answered for itself the question "Why?"; meanwhile the middle echelons are already busy over their heads with "What?" and "How?" No wonder, then, that the prevailing feeling is that everyone is trying to grab pieces of a jigsaw puzzle that are flying in the air.

Today the organizational dynamics, alongside the decentralization in the force-buildup areas, enable different approaches to influencing the bottom-up lines of development. The development personnel can translate the requirements of development in a way that makes it harder for the staff personnel and others to construct the rest of the value chain efficiently. Likewise, a force-deployment operator in the field units can develop an internal work process, which he sees as proper, that will not accept a solution that is appropriate and supportive of manpower or technological capabilities because these are subordinate to a different approach. And thus, driven by everyone's good intentions, we reach a situation where the managerial focus, including the professional goals, will be local and specific instead of directed at overall organizational success. The connection between the different components will be made at too high a level, and thus jointness will be squandered and organizational performances will be impaired.

So What Needs to Be Done?

- **A joint conversation starting at the stage of germination:** Assuming that approaches germinate somewhere at the lower levels, one should make sure that, already at the germination stages, the conversation is a joint one that includes the

staff and field personnel who are relevant to implementing the approach.

- **Specifying the "first among equals" who will lead the discussion:** The staff worker at the lowest level who has broad responsibility and influence must be designated the "first among equals". It is he who will lead the discussion of the approach, together with his counterparts in the development units, before obtaining the senior echelon's approval. The existing situation, in which the crafting of the approach develops upward from a single unit to the approval of the head of the unit before it is adopted by the organization, is a sure recipe for closure, contrariness, and obstacles. Already at the junior level, jointness is an essential requirement for deep, joint learning about the dilemmas and challenges without fear of professional disagreements. Those will only improve the result. The more that broad agreement about the approach emerges, the greater will be the ability to allow freedom of action for the partners, and the performance can be led with greater responsiveness to the relative advantage of all the stake-holders in the process. Only thus can one develop the organizational culture that believes everyone has a relative advantage, and only by working together can synergy be created. Without a joint framework for discussing the issues, such a culture will not develop.
- **Involving the senior management level:** A further obstacle to the development of the required organizational culture is the tendency of the senior management echelons to interfere before the stage of crafting the approach has been completed, thereby undermining the open conversation. For the joint task to be completed, the group of junior-level managers, which has engaged in the initial conversation and clarified the approach, must present the approach jointly to the senior echelon in a way that dispels the team members' affiliation with their own unit and augments the joint mission. Clearly, in light of the many obstacles described above, this is not a simple task.
- **An organizational declaration of an approach:** At the end of the process, a single document (some will call it "founding") must be written that declares the choice of an organizational approach. This document will be signed by the person with organizational authority, and from that moment it will become an available and accessible guide to every relevant stake-holder and will thereby serve to clarify the approach, which is, as noted, the heart of the matter.

Counterintelligence in the Western Countries and Big Data

Dr. Avner Barnea[119]

Introduction

Big-data[120] information systems have developed in the business world since the 2000s because of the need to deal more effectively with the huge quantities of information that business firms collect, especially in the social-media era. Subsequently, counterintelligence[121] (CI) organizations, like intelligence organizations in other fields, realized what big data could contribute to their activity. The main reasons for the delayed reaction were conservatism and unfamiliarity with the capabilities developing in the business world. The new challenges with regard to preventing terror and cyber attacks impelled the CI organizations to incorporate big-data systems that could improve their effectiveness. At the same time, almost no information on big-data use by CI is published, hindering public surveillance of the activity.

Development of the Use of Advanced Information-System Tools in CI

The last 50 years in the CI field can be divided into three periods:

- During the first period, which lasted until 1989, the year in which the Soviet Union collapsed and the East European (or Warsaw Pact) countries were liberated

119 Research fellow at the National Security Studies Center, University of Haifa; head of cyber, security, competitive intelligence, and crisis management studies, MBA program, Netanya Academic College; former senior Shin Bet official.

120 The definition of big data comes from the business sphere: **"Big data** is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation".
https://www.gartner.com/it-glossary/big-data
Mark Lowenthal, one of the former top CIA officials, defined big data this way: "The ability to amass and manipulate large amounts of data on computers offers, to some, tantalizing possibilities for analysis and forecasting".
https://www.afcea.org/content/big-data-way-%E2%80%A8ahead-intelligence

121 Counterintelligence is clandestine intelligence activity aimed at countering terror, countering radical political subversion, and preventing espionage by foreign countries. Cyber security is becoming part of CI. See an extensive definition ordered by the U.S. government:
Executive Order 12333, as amended, United States Intelligence Activities:
https://www.cia.gov/about-cia/eo12333.html
For more on the topic: Shulsky, A., & Scmitt, G. (2002), Silent warfare: Understanding the world of intelligence, Potomac Books, 99-128.

from the communist regime, the dominant issue in CI was counterespionage. In the Cold War era, the Soviet Union and its East European satellite states devoted huge efforts to clandestine collection, primarily seeking information about the West on the assumption that quality intelligence would help improve their military capabilities and help close the technological gaps between the Eastern bloc and the Western states.

- The second period, 1989-2001, was an interim period in which, while the Soviet threat came to an end, it was not yet clear on which issues the Western intelligence agencies would focus.
- The third period extends from the end of the 1990s, and particularly from September 11, 2001, to the present. In this era the main threats involve Islamic terror and the cyber domains. Intelligence organizations must develop new and substantial capabilities to counter these threats, which differ fundamentally from the classic espionage of the first period.

The First Phase in Counterintelligence: Counterespionage

This era was characterized by a struggle between the Western countries, led by the United States, and the Soviet Union and its communist satellite countries. Both sides gave high priority to espionage and to efforts to counter it. CI still lacked advanced information-system tools, and the quantities of information were relatively small compared to today. Most of the security information collected was obtained through clandestine means. Both sides assigned relatively low priority to open-source collection. To a large extent the CI organizations' approach to information was intuitive, while giving[122] priority to information obtained clandestinely without comparatively examining similar information that could be collected openly.

This period was characterized by a search for information that was manual and almost devoid of automation. At that time CI personnel had doubts about automatic processes for handling information, which they regarded as imprecise and as not contributing to human understanding. The challenge was to identify information in internal databases of the intelligence organizations, and it was quite difficult to identify information that was incomplete without efficient automatic systems. Hence the first domain in which information systems entered CI was that of single-value identification of people and organizations, primarily of CI targets, with the aim of improving the assessment of their dangerousness.

122 Twersky, A., & Kahneman, D. (2005), "Judgment under conditions of uncertainty: Heuristics and biases", in Rationality, Fairness, Happiness, Daniel Kahneman et al., ed. Maya Bar-Hillel, University of Haifa (Hebrew). See also Kahneman, D. (2013), Think fast, think slow, Tel Aviv, Matar, 136. (Hebrew)

At that time automatic information systems were still new. It was still much before intelligence personnel thought in terms of big data, and CI organizations had an ambivalent attitude toward these systems: on the one hand, a need was felt for advanced devices; on the other, there was a security concern about exposure of the CI activity, especially to external information-system experts. There was also a mistrust of these instruments' effectiveness. At a time when the quantities of information still had not reached the level of an explosion, many CI personnel believed in manual systems and in the human memory. The reigning mindset among CI personnel was one of excessive confidence in their expertise with a tendency to view the contribution of automatic systems as negligible.

Then as now, business firms were the leaders in the information-systems world as they developed tools for business in a competitive reality. At that time some intelligence organizations believed they could develop sophisticated devices by themselves while also fearing to expose them, and these attempts entailed financial investments that were very large as well as unjustified. From the end of the 1980s, intelligence organizations began to understand that they were not equal to the advanced capabilities that computerization companies were developing, and that the right course of action was to acquire these tools and adapt them to their needs.

In the period of the struggle between the blocs espionage was given high priority by both sides, and so was the activity to counter it. CI still did not have advanced information-system tools, and the quantities of information were relatively small compared to today, with most of the security information obtained through clandestine means

The next stage involved learning how to retrieve textual information. Because, at that time, the information systems had limited capabilities in this regard, keywords were used. However, it soon became clear that keywords were a problematic approach because their number was growing rapidly and it was difficult to achieve uniformity when linking the information obtained to the relevant keyword. In retrospect, the automatic information systems of the 1970s and 1980s fostered progress in CI capabilities but did not meet the expectation of giving these organizations a significant advantage over their opponents. Many years passed before automatic indexing became part of intelligence work and made "keywords" almost irrelevant.

The Second Phase of Counterintelligence: The Interim Period

The end of the 1980s to 2001 constituted the interim period. In Western countries, especially the United States, intelligence organizations' budgets were substantially reduced and the intelligence effort shifted from a focus on the Soviet Union and its satellite states to a global and decentralized approach in which ad hoc threats were dealt with as they arose. The American intelligence community had to be more flexible in responding to changing needs; other Western intelligence organizations also had difficulty adjusting to the new situation.[123] The intelligence endeavor was also influenced by conclusions related to the outlook of Francis Fukuyama in his book The End of History and the Last Man (1992). In it he raised the possibility that the fall of the communist bloc[124] was not just another event in the long human history of conflicts between worldviews, ideologies, and types of regime but instead marked the end of this history, a transition to an era in which peace and liberal democracy would prevail in the world without being challenged by any other ideology, and conflicts between states would be limited to specific pockets. Thus Western decision-makers assessed that in the future there would be less need for security and intelligence organizations, including CI organizations. Shortly thereafter the historian Samuel Huntington came out with a completely different theory, propounded in his 1993 book *The Clash of Civilizations*.[125] In the post-Cold War era, Huntington saw civilization as a key factor in the world of international relations. However, because the civilizations differed from each other in their basic values and worldviews, he believed that a clash between them was imminent and that it would be driven mainly by the religions.

Despite the uncertainty about where the intelligence organizations were heading, meanwhile another process - the development of the internet - began. It led to a rapid growth in the quantities of information that were open to all. It began to appear that the open and accessible information could be more helpful in the security-intelligence field as well, and could complement clandestine collection. Still, many CI organizations were in a state of denial about the potential contribution of open-source collection and continued to give the highest, and often exclusive, priority to collection by clandestine means. In the United States the congressional intelligence committees had already been calling to enhance open-source collection capabilities

123 Marrin, S. (2012), *Improving intelligence analysis: Bridging the gap between scholarship and practice, London, Routledge.*

124 Fukuyama, F. (1993), *The end of history and the last man,* Oram, Tel Aviv. (Hebrew)

125 Huntington, S. (2003), *The clash of civilizations*, Shalem, Jerusalem. (Hebrew)

since the 1990s,[126] noting that the information obtained was often of no less quality and that this could also allow cuts in the intelligence organizations' budgets.[127] In this period the computerization of the intelligence organizations, including the CI organizations, continued, later leading to the use of big-data tools. However, for the most part the challenges did not appear to be major and the cost of computerization was still very high. Thus the introduction of automatic information systems continued but not with large investments, even though it was clear in the business world that computerization was the next revolution notwithstanding the bursting of the "dot-com" bubble at the end of the 1990s. In Israel computerization developed more rapidly in CI than in most Western countries because of Israel's unique problems. Unlike in other Western countries, in Israel thwarting terrorism had been a central issue for CI since 1967.

With the end of the Cold War and the rise of the internet, it began to appear that the open and accessible information could be more helpful in the security-intelligence field as well. However, CI organizations continued to deny the potential contribution of open-source collection to the CI endeavor

Toward the end of the second phase, dozens of search engines began to develop in the business world; the aim was to collect information from the internet. Also developing were capabilities and tools to help analyze information and understand its significant by drawing connections between the different bits of information, or what is known as link analysis.

These capabilities and tools were originally developed for researches in the social sciences[128] and the computer sciences, and later formed the basis for the development of important analytical computerization capabilities, mainly in the field of terror prevention.

126 Best, R., & Cumming, A. (2007), "Open source intelligence (OSINT): Issues for Congress". *Congressional Research Service,* https://fas.org/sgp/crs/intel/RL34270.pdf

127 Steele, D. (2008), "The open source program: Missing in action". *International Journal of Intelligence and Counterintelligence,* Vol. 21, No. 3, 609-619.

128 Barnea, A. (2005), "Link analysis as a tool for competitive intelligence". *Competitive Intelligence Magazine*, Vol. 8, No. 4, July-August.
See, in the context of operative intelligence aspects of the use of link analysis:
Barnea, A. (2017), "The 'Lone Wolf' Phenomenon: New challenges in the era of overload of information". *International Journal of Intelligence and Counterintelligence*.

The 21st Century as a Turning Point - the Third Phase: Thwarting Terror and Cyber Threats

The third phase of CI begins with the terror attack on the United States on September 11, 2001. In the Cold War era, intelligence developed in the strategic domain with an emphasis on long-term warnings and predictions[129] alongside a considerable improvement in collection[130] and counterespionage capabilities. Later, and especially after 9/11, CI had to adapt rapidly to new situations and focus more on tactical issues of warning about and fighting terror[131] by nonstate organizations, a different sort of threat from what intelligence had been used to. It[132] took the intelligence organizations time to adjust to the new situation. It now became clear that thwarting terror was the main challenge for CI, and later a completely new field - cyber security - emerged. In the present era, two other components of the CI discipline - countering political subversion and counterespionage - are of lower priority for CI in the Western countries.

One of the important lessons of 9/11 was that there was no lack of information. The investigatory commission that examined the events that had preceded the attack[133] reached the conclusion that American intelligence had possessed quality information on Al Qaeda's intentions to attack targets within the United States including a possible date for the attack. The main failing of the American intelligence community was an inability to formulate a clear threat picture, resulting from the fact that none of the relevant U.S. intelligence organizations had had a full picture of the threat posed by Al Qaeda; instead the relevant intelligence was dispersed among the different organizations, primarily because of a long-standing lack of cooperation and an unnecessary compartmentalization. The commission recommended a change in approach to compartmentalization and to information securi-

129 Davis, J. (2007), "Intelligence analysts and policy makers: Benefits and dangers of tensions in relationships", in Johnson, L., ed., Strategic intelligence: *The intelligence cycle*, Praeger Security International, New York, 143-165.

130 Fingar, T. (2011), "Analysis in the U.S. *Intelligence Community: Missions, Masters, and Methods", in Intelligence Analysis, Behavioral and Social Scientific Foundations*, Baruch Fischhoff & Cherie Chauvin, eds., National Research Council of the National Academics, National Academies Press, Washington, DC.

131 The U.S. State Department defines terror as "the unlawful use of force and violence against persons or property to intimidate or coerce a government, the civilian population, or any segment thereof, in furtherance of political or social objectives".
Byman, D. (2017). "Should we treat domestic terrorists the way we treat ISIS?" Foreign Affairs, October 3. https://www.foreignaffairs.com/articles/united-states/2017-10-03/should-we-treat-domestic-terrorists-way-we-treat-isis?cid=nlc-fa_twofa-20171005

132 Cavelty, M., & Mauer, V. (2009), "Postmodern intelligence: Strategic warning in an age of reflexive intelligence". *Security Dialogue*, Vol. 40, No. 2, 123-144.

133 "The 9/11 Commission Report". (2004). http://www.9-11commission.gov/report/911Report.pdf

ty and set standards for information sharing in the U.S. intelligence community.[134] This included the framework of fusion centers located throughout the United States, which coordinate various security activities, primarily involving terror prevention, among the intelligence organizations, police forces, and other organizations.[135] The priority given to these standards over conservative values of excessive compartmentalization and exclusive holding of information by all the intelligence organizations laid the conceptual groundwork for the incorporation of highly advanced information tools for CI purposes - namely, big-data systems.

American intelligence, along with intelligence organizations in other Western countries, launched an intensive effort to collect information on terror threats within countries, including within the United States, and also on the global level so as to thwart these threats in advance. This effort was facilitated, among other things, by a new policy dubbed the War on Terror, and it was accompanied by new and rapid legislation that gave the U.S. administration capabilities that completely altered the balance between, on the one hand, human rights and individual freedom, and, on the other, security interests of states that seek to protect their sovereignty and their citizens.[136] The transition to digital communication systems fostered highly sophisticated capabilities in the internet-collection or open-source

After 9/11 the Western intelligence agencies launched an intensive effort to collect information on threats within countries and on the global level so as to thwart these threats in advance. This effort was facilitated by a new policy dubbed the War on Terror, and it was accompanied by legislation that gave the U.S. administration capabilities that completely altered the balcned between individual freedom and security interests

134 See also the book by the former White House adviser on fighting terrorism, Richard Clarke:
Clarke, R. (2004), *Against all enemies: Inside America's war on terror,* Free Press, a subsidiary of Simon & Schuster.

135 de Castro Garcia, A., Matei, F. & Bruneau,T. (2017), "Combatting terrorism through fusion centers: Useful lessons from other experiences?" *International Journal of Intelligence and Counter Intelligence*, Vol. 30, No. 4, 723-742.

136 On September 18, 2001, about a week after the 9/11 attack, the U.S. president approved the use of military force against those responsible for this attack, including nations, organizations, and individuals anywhere in the world. Subsequently the USA PATRIOT Act was passed, which gave wider authorization, without precedent in the Western democracies, for the interception and disruption of terror activities including within the United States. Later the act was strongly criticized for granting almost limitless powers to the law-enforcement authorities.

(OSINT) field alongside abilities to monitor information. The immediate results were enormous quantities of information that were collected, and great difficulties in using this information to identify suspicious figures and potential terror operatives - further[137] underlining the need for ultra-sophisticated information systems.

In those years the need for a dramatic improvement of the information systems also arose in the business field. By the mid-2000s various information systems were being used in the business organizations, each serving a certain component of the organization - for example, information for marketing purposes, for sales, for human resources, for management, for operation, and so on. There was an organizational difficulty involved in managing each of these systems, which did not communicate with each other and required special skills to operate them. The use of information that was already contained in these systems was not exhaustive because of the difficulty in operating them, but the main problem was that each of these systems stood by itself, and they did not enable the organizations to perform cross-systems operations with the aim of improving their business capabilities. Thus the need arose to develop systems capable of handling large quantities of information in different fields within an organization. Such systems were developed by leading computerization companies such as Orkal, IBM, SAS, SAP, and others, and these platforms were later given the name "big data".[138] The point of departure was that these new tools, which, mainly because of the digitality of the information, can handle information of any kind - not just written information (texts) but also voice, video, and target locations (location intelligence - LI) - could help in better exhausting information that the organizations already possessed. The underlying premise of the development and marketing of these systems was that they could help organizations make better decisions, based on the information to be found within the interorganizational systems,[139] in areas of market research, competitor analysis, management of the organization's units, and so on. Leading companies worldwide began to install big-data systems, and within a few years these systems became standard in large and medium-size companies.

137 Lim, K. (2015), "Big data and strategic intelligence". Openbrief. https://www.openbriefing.org/publications/report-and-articles/big-data-and-strategic-intelligence/

138 In 1999 the academic literature used the big-data concept for the first time regarding information systems, and from there the name expanded to include relevant information systems that developed later: Bryson, S., Kenwright, D., Cox, M., Elsworth, D. & Haimes, R. (1999), "Visually exploring gigabyte data sets in real time". Communication of the ACM, Vol. 42, No. 8, 82-90.

139 Jones, M. & Silberzahn, P. (2013), "Three reasons why big data doesn't make you smarter: Lessons from the world of intelligence". Forbes, Feb. 7, https://www.forbes.com/sites/silberzahnjones/2013/04/11/play-it-like-steve-jobs-three-questions-for-business-leaders-to-ask-when-surprise-hits/#27c33ad4765d

The assumption[140] that they would make organizations "smarter" underpins the use of these systems for business intelligence - BI;[141] thus the term "data-driven organizations" entered the business field, referring to organizations that make effective use of big-data systems in their ongoing activity.

The use of big-data systems in the business field is constantly researched, revealing how these tools are used and how much benefit is derived from them; there is also considerable criticism of these systems' insufficient analytical capabilities. The use of these systems in the CI field,[142] however, is different, and there is almost no systematic research on the subject. At the same time, it is clear that intelligence organizations, including CI, came to realize the possible contribution of big-data devices long after these advanced systems were already operating successfully in the business world.[143]

This delay stemmed from several factors: conservatism and hesitancy about entering new fields; insufficient familiarity with the business world, where tools develop rapidly mainly because of the competitive environment; security concerns about exposing sensitive information in systems that might not be sufficiently protected and safe; and the need to adapt commercial big-data systems to the CI organizations' unique needs. The big-data devices in the business world are known to be more varied than those used by intelligence, including CI, because of the wide variety of business activities.[144]

According to Edward Snowden, who left the NSA in 2013, the worldwide collection apparatus of American and British intelligence,[145] which is especially aimed at terror prevention, has gone fairly far in the big-data field. Snowden gave a peek into the enormous system for collecting and maintaining information that was estab-

140 According to a 2011 report by the McKinsey consulting firm, every company in the United States that employs over a thousand people has information stored in its systems at a magnitude of 200 terabytes: https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

141 A distinction should be made between BI systems whose focus is on handling information that is within the organizations and the concept of CI - competitive intelligence - which is the field that deals with monitoring threats and identifying opportunities for business firms in the competitive environment by collecting open-source information.

142 Court, D. (2015), "Getting big impact from big data". McKinsey Quarterly, January. https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/getting-big-impact-from-big-data

143 For example, only in 2013 did the CIA tell of an agreement that was signed with a business entity on the issue of cloud computing, which is required in the field of big-data systems: http://www.businessinsider.com/cia-presentation-on-big-data-2013-3

144 Lyon, D. (2014), "Surveillance, Snowden, and Big Data: Capacities, consequences, critique". Big Data & Society, July–December, 1–13.

145 Greenwald, G. (2013), "NSA collecting phone records of millions of Verizon customers daily". The Guardian, June 6. Available at: https://www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order

lished after 9/11, including the use of big-data devices.[146] He emphasized two aspects of CI work in the context of fighting terror: the huge quantities of information that have been amassed, and the huge capabilities for storing the information in big-data systems and efficiently retrieving it. He also noted the great difficulty involved in filtering the dangerous targets out of the reams of information, and he presented inside information that indicates the difficulties of working effectively in such a situation. These systems pose similar difficulty in the business domain.

Edward Snowden highlighted some challenges for CI work in the terror-prevention context. He pointed to the enormous quantities of data and the difficulty in storing the information. He also noted the great difficulty in filtering the dangerous targets out of the reams of information, and he presented inside information that indicates the difficulties of working effectively in such a situation

It is common to ascribe to big-data systems in the CI field abilities to absorb hundreds of thousands of information units in a second, and, by using analytical tools, to perform an analysis that points to suspicious figures and warns of imminent terror attacks. But an analysis of the situation regarding big-data systems in the business field, which is less complex than the intelligence world, shows that the number of failures in qualitative (not quantitative) analysis is quite high; presumably the terror prevention, has gone fairly far in the terror prevention, has gone fairly far in the situation in the CI domain is no different.[147] This point was considerably buttressed by Snowden's revelations, which reinforced similar observations by former intelligence officials[148] that information collection has reached enor-

146 Lyon, 2014. In 2015 the then British interior minister (now prime minister) Theresa May acknowledged the collection of very large quantities of personal information by the British intelligence community and spoke of bulk power, which refers to the use of big-data systems:
Mathieson, S. (2015), "How MI5 and MI6 gather your personal data for surveillance". Computer Weekly, June 17, http://www.computerweekly.com/news/450298621/How-MI5-and-MI6-gather-your-personal-data-for-surveillance

147 Moore, S. (2015), "How to Prevent Big Data Analytics Failures". Gartner, December 18. http://www.gartner.com/smarterwithgartner/how-to-prevent-big-data-analytics-failures

148 See,e.g., the 2009 book by former NSA official James Bamford:
Bamford, J. (2009), *The shadow factory: The NSA from 9/11 to the eavesdropping on America*, Anchor Books First, New York. On this issue see also a far-reaching survey of democracy, ethics, and intelligence:
Konstanopoulos, I. (2016), "Democracy and ethics vs. intelligence and security: From Wikileaks to Snowden", in George Bitros and Nicholas Kyriazis, eds., Democracy and an Open-Economy World Order, Springer International, 3-24.

mous and, in the opinion of many, unreasonable magnitudes,[149] posing difficulties for effectiveness as well as ethical issues, given the scope of collection and the lack of transparency about its scope.[150]

Although Snowden did not apply the big-data concept to the computerization devices that the NSA uses for information storage and retrieval, the material he provided, including official NSA presentations that were uploaded to the internet, reveals the extensive use of these systems. The terms that appear in the NSA presentations that Snowden conveyed - *massive, dragnet, surveillance* - are also indicative.[151] Snowden's revelations also gave rise to the concept of "big data and surveillance", which means that one cannot conduct efficient surveillance of people, groups, and organizations, including capabilities to identify individuals within large populations, without the use of big-data systems. The concept of surveillance has developed and it is also called "actionable intelligence", which means that, by using the information hoard in the big-data bases, one can identify individuals who constitute a risk and act against them specifically.[152] Snowden spoke much about the "metadata" concept, which entails that identifying suspicious individuals or small groups requires wide-scale collection of a large variety of sources usually originating in communication.[153] For example, a software that the NSA developed called Co-Traveler can connect between cellular phones that are associated with suspicious individuals and map the contacts between them, monitoring their communication if necessary. To a certain extent the things Snowden exposed were already known,[154] but his unprecedented presentation of original NSA materials added urgency to the matter, including assessments not only about the magnitude of the activities conducted but also about their degree of effectiveness. Snowden also showed how media companies in the United States, such as Apple, Facebook, Google, YouTube, Microsoft, Skype, Yahoo, and others, including internet service providers (ISP), cooperated with the intelligence organizations even though they did not have to, and highlighted the far-reaching extent of cooperation between the administration and the communication providers,

149 Jeffreys-Jones, R. (2017), *We Know All About You: The Story of Surveillance in Britain and America*, Oxford University Press.

150 See Konstanopoulos, "Democracy and ethics".

151 Sotteck, T. and Kopstein, J. (2013), "Everything you need to know about PRISM". The Verge, July 17, https://www.theverge.com/2013/7/17/4517480/nsa-spying-prism-surveillance-cheat-sheet

152 Gandy, O. (2012), "Statistical surveillance: Remote sensing in the digital age", in Ball, K. S., Haggerty, K. and Lyon, D., eds, Routledge Handbook of Surveillance Studies, London and New York, Routledge, 125-132.

153 Friedman, G. (2014), "Keeping the NSA in perspective". Stratfor, April 22, http://www. stratfor.com/weekly/keeping-nsa-perspective

154 Andrejevic, M. and Gates, K. (2014), "Big Data surveillance: Introduction". Surveillance & Society, Vol. 12, No. 2, 185-196; Ball, K. S. and Snider, L., eds. (2013), The Surveillance-Industrial Complex: A Political Economy of Surveillance, London, Routledge.

something that had not been made public previously.

Snowden's revelations clarify the extent to which the privacy of great numbers of people in the United States, Britain, and elsewhere in the world has been compromised. This situation stems from the magnitude of the activities carried out, most of which had no connection to terror.[155] Yuval Noah Harari asserts: "This universal tendency to exaggerate the size of the threat is very problematic because it causes the waste of invaluable resources".[156] Heavy public criticism in the United States has prompted intensified public supervision of the NSA by the congressional intelligence committees, as well as legislation requiring special approvals that restrict the scope of intelligence collection aimed at citizens who are above all suspicion.[157] President Obama referred to the matter in 2014 when he called for the protection of privacy - a "comprehensive review of big data and privacy" - in the wake of Snowden's revelations.[158]

As the use of big data in the field of national intelligence, primarily for CI purposes, increases and the devices used grow more sophisticated, the business world is debating the contribution of big data to improving organizations' abilities to be more competitive and gain an advantage over competitors. Most of the criticism does not concern the technological capabilities of big-data devices but, instead, the extent of proper use of the data that the organizations possess. It is quite commonly claimed that the companies that develop big data have not developed good analytical tools that can help analyze the information obtained from these devices.[159] This is also the position of Jones and Silberzahn, prominent researchers in the field of business-management and intelligence systems.[160] Thanks to improvements that have been made in these devices in recent years, particularly in the analytical sphere, alongside the addition of new professions to the information field such as data analyst and data sci-

155 Brown, M. (2015), "NSA Mass Surveillance: Biggest Big Data Story". Forbes, August 25, https://www.forbes.com/sites/metabrown/2015/08/27/nsa-mass-surveillance-biggest-big-data-story/#578b4e092c13

156 Y. Harari (2009), "What is terror? From the Middle Ages to the twenty-first century". Zmanim, 108, Fall. (Hebrew)

157 Hattam, J. (2016), "Spying after Snowden: What's changed and what hasn't". The Hill, December 15, http://thehill.com/policy/technology/310457-spying-after-snowden-whats-changed-and-what-hasnt

158 White House (2014), "Big Data and the future of privacy", available at: https://obamawhitehouse.archives.gov/blog/2014/01/23/big-data-and-future-privacy

159 Gilad, B. (2015), "Your Big Data Analytics Can't Save Your Company". Academy of Competitive Intelligence, available at: http://www.academyci.com/2015/01/06/big-data-analytics-cant-save-company/

160 Jones and Silberzahn (2013).

entist,[161] progress in using the information stored in the business big-data systems has begun.[162] Today one cannot see business activity in the fields of commerce, health, finance, social media, and even intelligence[163] that does not make use of big-data systems. One of the senior CIA officials recently noted that the organization has established a Directorate for Digital Innovation that is "the first new directorate within the spy agency in more than 50 years".[164]

Big data has also made an important contribution to thwarting cyber threats.[165] The development of open-source collection capabilities for the social media (Socmint), which is the new intelligence arena, comes from the business sector[166] and has flowed into the CI field, mainly in the context of counterterror and identifying cyber threats in advance, and it requires the use of big-data systems.[167] It appears that these systems' contribution to the analytical process is in need of improvement; meanwhile they are helping more in substantiating assessments that have already been made.[168]

---

161 "A data scientist is a professional responsible for collecting, analyzing and interpreting large amounts of data to identify ways to help a business improve operations and gain a competitive edge over rivals".
http://searchbusinessanalytics.techtarget.com/definition/Data-scientist
At its website the CIA seeks to hire members of this profession:
https://www.cia.gov/careers/opportunities/science-technology/data-scientist.html

162 There are many studies on the topic of big-data systems, usually by the world's leading consultancy firms. For example, McKinsey: https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/how-companies-are-using-big-data-and-analytics
Accenture:
https://www.accenture.com/t20160106T194441__w__/fi-en/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Digital_1/Accenture-Global-Operations-Megatrends-Study-Big-Data-Analytics-v2.pdf
http://www.ey.com/Publication/vwLUAssets/EY_-Big_data:_changing_the_way_businesses_operate/%24FILE/EY-Insights-on-GRC-Big-data.pdf
And there are also academic studies, such as:
MacAfee, A. & Brynjolfsson, E. (2012), "Big Data: The management revolution". Harvard Business Review, October.

163 In 2012 the chief technology officer (CTO) of the CIA, Ira Hunt, addressed this issue on the CIA's websise and pointed to the organization's challenge of recruiting analysts in the big-data field:
https://www.cia.gov/news-information/featured-story-archive/2012-featured-story-archive/big-data-at-the-cia.html

164 Konkel, Frank, "How the CIA Is Making Sense of Big Data". Nextgov, March 16, 2016
http://www.nextgov.com/big-data/2016/03/how-cia-making-sense-big-data/126722/

165 O'Brien, S. (2017), "Challenges to cyber security and how Big Data analytics can help". Datameer, May 4, https://www.datameer.com/company/datameer-blog/challenges-to-cyber-security-and-how-big-data-analytics-can-help/

166 Scaachi, M. (2017), "Competitive intelligence and unstructured data". Competitive Intelligence, Vol. 20, No. 1, Spring.

167 Murdock, J. (2017), "Spies in the age of social media: Ex-CIA experts reveal challenges of modern espionage". International Business Times, July 19,
http://www.ibtimes.co.uk/spies-age-social-media-ex-cia-experts-reveal-challenges-modern-espionage-1631042

168 Lim (2015).

Conclusion

American intelligence fluctuates between repeated, not always successful attempts to provide a solution to a security threat and unbridled, ad hoc reactions to threats. An example of the latter was the mass detention of U.S. citizens of Japanese background after the Pearl Harbor attack (1941) for fear that some among them had engaged in sabotage or spying against the United States. The shock of 9/11 triggered unprecedented collection activity based on rapid and imbalanced legislation, which became all the more controversial after Snowden's revelations.[169]

The result was that the issue of individual rights in the digital era, and of the place of the individual in the security discourse, was given greater weight. CI organizations in the Western countries, and in Israel as well, need to focus on more accurate identification of the threats while infringing on individual rights and privacy as little as possible. Improvements in the analytical capabilities of big-data systems, alongside the professionalization of the analysts who use these devices, will help produce this change. At the same time, what is publicized in other Western democracies indicates that in Israel, too, there is room for public discussion of how to balance between security needs and individual rights, and how to ensure that the power possessed by CI will be better controlled.

169 In January 2014 a New York Times editorial called for Snowden to be viewed as a whistleblower because of his important revelations and said he should be allowed to return to the United States and granted a pardon: https://www.nytimes.com/2014/01/02/opinion/edward-snowden-whistle-blower.html