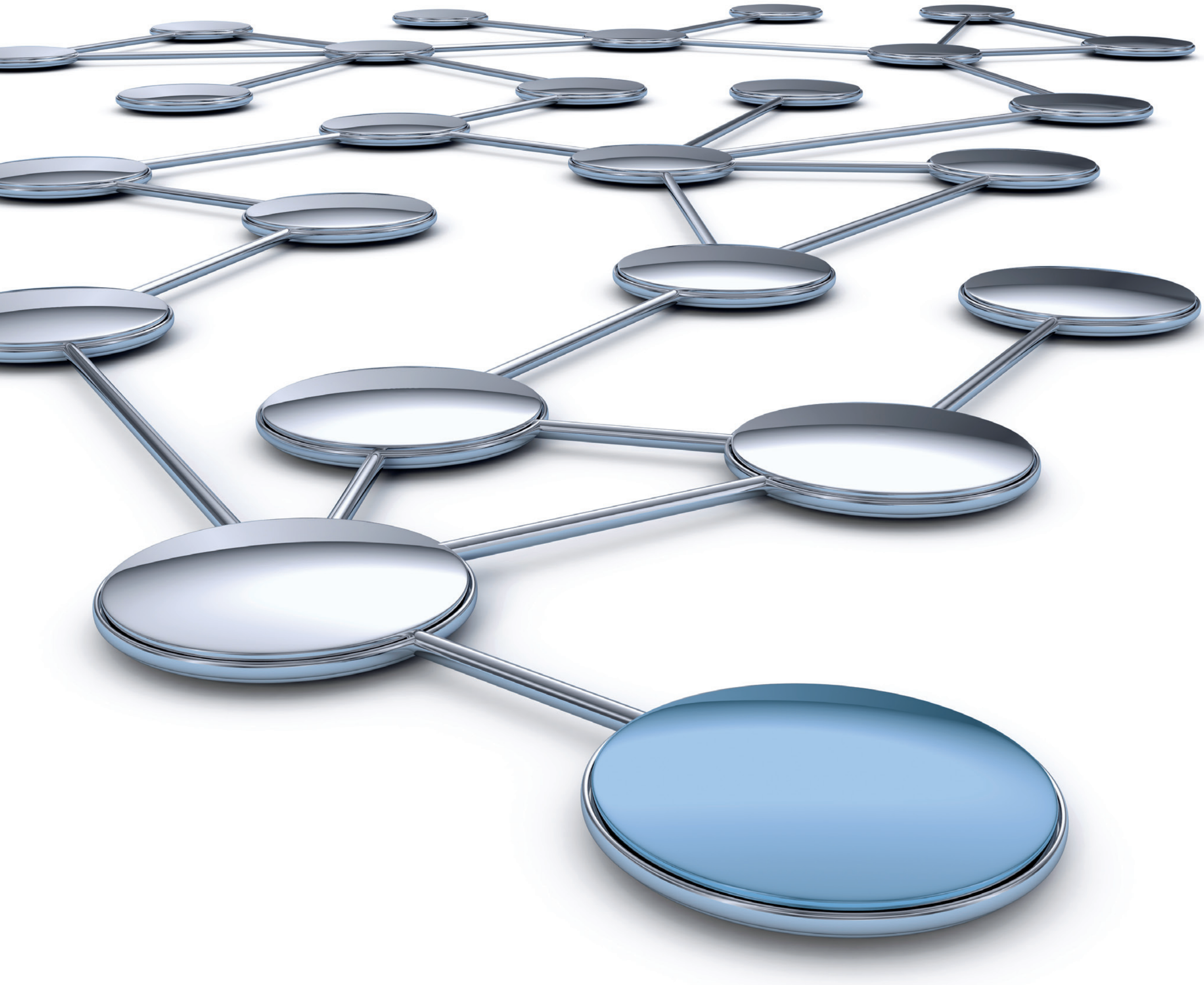

OS/NT

Report 1/2010

ISN ETH Zurich

INTERNATIONAL RELATIONS AND SECURITY NETWORK
ETH Zurich | Leonhardshalde 21, LEH | 8092 Zurich | Switzerland
Tel: +41 (0)44 632 04 24 | Fax: +41 (0)44 632 14 13
www.isn.ethz.ch | osint@sipo.gess.ethz.ch

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Authors: Felix Juhl, Chris Pallaris, Florian Schaurer

International Relations and Security Network (ISN)
ETH Zurich
Leonhardshalde 21, LEH
8092 Zurich
Switzerland
Tel.: +41 44 632 04 24

osint@sipo.gess.ethz.ch
www.isn.ethz.ch

Project supervision: Andreas Wenger, Director CSS; Victor Mauer, Deputy Director CSS

Disclaimer: The views expressed in this report do not represent the official position of the Swiss Federal Department of Defense or any other governmental body. They represent the views and interpretations of the authors, unless otherwise stated.

ISN
ETH Zurich

OSINT

April 2010

Key Challenges

Data Life-Cycle
- Ensuring a long term data retention plan.

Sensing
- Monitoring development of voice and data translation tools.

Intelligence Decision Making
- Applying interactive data analytic technology.

Validation Regimes
- Extending validation and verification regimes to data other than text.

Requirements Framework
- Developing standardized OSINT procedures for the intelligence cycle.

Strategic Forecasting
- Fostering joint simulations and scenario-planning.

Business / Competitive Intelligence
- Auditing if an agency's objectives can be achieved more efficiently by partly outsourcing tasks.

Ambient Intelligence
- Following the evolution of intelligent technologies.

Information Abundance
- Anticipating the informational risk implications of future IT developments.

OSINT Report 2010

Technology Trends

Data Life-Cycle: Strategies for Long-Term Data Retention

Government agencies worldwide are investigating the implications of long-term storage. Driven by the requirement to retain data indefinitely for operative, legal, administrative or historical purposes, many IT managers are grappling with how to preserve (electronic) data, documents and information for 50 years or more. Some of the issues they face include estimating the lifespan of storage materials and challenges such as technology obsolescence, selecting the right storage media, hardware compatibility, the longevity of operating systems, obsolescence of the formatting and structure, and document security. To ensure lifelong data handling adaption of a policy for continuous digitization and digital preservation is needed. In practical terms, this means scanning documents and moving them on to the newest versions of the software, or emulating older applications on current software, so they can replicate the older software's behaviour and still access older digital files that were stored on older media.

A long term data retention plan needs to be compiled in order to set up a (redundant) secure storage system environment, continue digitization and digital preservation, use adequate cryptographic methods and, eventually, form a digest of the file to detect any changes that may be due to deterioration in storage.

Sensing: Automated Near Real-Time Translation

Imagine walking around in foreign countries and talking with inhabitants in their mother tongue. It is a future that is on the fast way to becoming a reality. Researchers are actually closing in on the technology and foresee its application in the coming years in a very familiar device: smart phones. Google, which is at the forefront of this development, presented its progress lately, bringing near real-time translation, both voice and text, to a smart phone. At the Mobile World Congress in Barcelona, Spain, Google CEO Eric Schmidt showed off a smart phone feature that allows a user to take a picture of German text and have it quickly translated into English using optical character recognition software and Google Translate technology. For the past seven years, the military has

NOTES

been deploying a one-way communicator called the Phraselator P2 from Voxtec that conveys thousands of phrases with pre-recorded translations. Voxtec hopes to have a limited two-way communicator available in the next 12 to 18 months that will work with specific use cases and languages.

> Technologies like Languageweaver, based on GALE, can handle near real-time translations of several Arabic languages. Applications for smart devices will become generally available within the next 2 to 3 years. Therefore, the development of automated voice and data translation tools and add-ons for PSTN/GSM/HF interception solutions need to be monitored while decoders and mediators for local lawful interception systems need to be developed and deployed.

Intelligence Decision Making: Next Generation Interactive Analytics

Today more than ever, the volume and velocity of incoming data coupled with shrinking windows of time make intelligence and investigative analytics very challenging. Intelligence agencies' operations demand technologies that allow analysts to connect to data, explore the data interactively and share results through collaborative analysis. Over the past few years, most of the innovation in analytics has been seen in the area of automated information analysis. These techniques have been successfully applied in the private sector and in some security environments (e.g. in the domain of intelligence-led policing) to mitigate risks and channel resources more effectively. They remove the analyst from the equation by attempting to reveal all relevant insights automatically. However, in most investigations, the most important component is human judgment. Fortunately, modern technology has caught up with these challenges, namely with three breakthroughs in analytic technology:

1. Interactive data visualization
2. Collaborative investigative analysis
3. Unified data views

These three innovations attack the problem at its core and expose what the counterpart wants to hide most. Together, they represent the foundation of a new approach in intelligence and investigative analysis called Interactive Analytics (IA). IA is an investigation-centric approach to analyzing and understanding data in support of more accurate identification. This approach leads to improvements in detection, reporting and case resolution. IA provides the ability to explore, detect and confirm hidden relationships across disparate data sets. The exploration can be conducted in an unconstrained manner while investigators exchange insights as they are uncovered. Interactive data visualization allows the investigator to represent massive amounts of data in different visual representations and, in so doing, isolates important facts and patterns. But it is more than just pictures. It gives the investigator the freedom to ask questions through direct interaction with the visualizations. As the investigator discovers new insights, he can easily navigate, drill down, and develop vivid profiles. He can link together individuals, suspicious events, past events, current and historical alerts, accounts, acquaintances, background checks and transactional behavior. This is a powerful approach to investigative and predictive analysis that leverages the ultimate pattern recognition machine: the mind of the investigator. The ability to detect or anticipate non-obvious relationships and confirm inferences very quickly are important characteristics of interactive data visualization and analysis.

> Setting-up an OSINT workspace environment framework and defining interactive processes for exploring and analyzing unstructured and structured information, data and knowledge to discern trends or patterns, is indispensable for deriving solid insights and drawing trustworthy conclusions. The OSINT process includes communicating findings and effects constant change within the analyst's working environment. Therefore, a powerful data integration technology that allows analysts to quickly connect data from multiple sources and explore the unified data using rich charts, maps, time lines, relationship graphs and more, is of utmost importance.

Conceptual and Practical Challenges of Intelligence

Validation Regimes: Improving the Evaluative Capacity for Non-Text Information

Traditional regimes for the verification and validation of open source information are proving inadequate to the task of OSINT analysis. Traditionally, such regimes were established to evaluate the accuracy, reliability and credibility of text-based information. However, text based data constitutes a shrinking percentage of the overall data consumed by OSINT professionals. It is generally acknowledged that the common analytical process assessing the reliability of an intelligence source and the assessed level of confidence or credibility of the information according to the "Admiralty System" (as elaborated in NATO's STANAG 2511) is enhanced radically by the use of open source information (OSINF). Yet, far too often it only remains a benchmark by which classified sources and methods are judged to be relevant and cost-effective ex post. While OSINF proves to be increasingly helpful for complementing, contextualizing and supplementing information already collected through other channels, it is also indispensable as a "first resort" for validating the overlap of massive unstructured data sets, enabling pattern-matching and measuring amplifications in information transmissions, adding both structure and trustworthiness to the demand-driven analytical product. Appreciating that only resilient and flexible validation regimes will flank both immediate analytical results as well as strategic long-term gains, organizational agility is mandatory.

> Validation regimes will need to be extended to include audio-visual data, mashups, user-generated content, etc. The absence of commonly accepted criteria for the validation of non-text information types underscores the need to develop such guidelines for analysts working with open source information (e. g. within EU's OPTIMA project).

Requirements Framework: Defining the Workplace and Reforming the Workforce

While most government agencies have always and extensively used OSINF in some way or another, a common application-driven "doctrine" of OSINT, being a framework for the proliferation and management of security-sensitive knowledge in general, is still lacking, defining the workplace, adequately training the literacies needed, sharing best practices and diversifying acquisition and analysis between government and non-government subject-matter experts fusing and making coherent information from various sources under time pressure. It needs to be understood that a solid, yet critical, creative and conceptual analytical tradecraft is key to measuring the value of any given information and to derive actual knowledge from raw data. The "collection bias" must be overcome and it must be realized that OSINT is not limited to mining the internet or monitoring the press, but also requires considerable social and cultural skills from the players involved. It is the – originally Swedish – concept of a comprehensive and collaborative multi-national, multi-agency, multi-disciplinary, multi-domain information sharing (M4IS) and sense-making (M4IS2) approach that will ultimately lead the way to an intelligence culture fostering awareness and appreciation for the role a well-defined OSINF/T actually can play "connecting the dots" in an uncertain environment.

> A standardized catalogue of requirements for intelligence personnel dealing with OSINT will help clearly defining the role of OSINT within an agency´s intelligence cycle and encourage a regular exchange of experiences with other services, private businesses and academia. Continuous training through modular workshops, expert forums and ad hoc conferences is indispensable for sustained success.

Strategic Forecasting: Closing the Gap Between Detection and Decision

The development of unsupervised, self-learning and topic-sensitive classification algorithms clustering similarities as well as discrepancies (e. g. NATO's COHARS project) from the information collected is a major step towards precisely modeling scenarios, monitoring trends and identifying root sources as well as denial and deception. The applicability of any intelligence product eventually lies in its influence on (mainly political) decision-making, therefore it is essential to not only intuitively look at past experiences in order to predict the future, but even more to formalize, visualize, merge and relate information on a given topic constantly and real-time for early warning. It has to be considered that simulations and situation awareness do not only reflect potential developments and plausible outcomes, but they also provide excellent insight into the "human factor", i. e. the unique behavioural patterns of the people taking part, making accurate scenario planning yet another valuable and challenging tool for the training of OSINT professionals. Today, early warning and scenario-building no longer mean just screening known parameters with the help of static indicators, but scanning information for weak signals of potential, unlikely threats and unknown risks. While most major intelligence failures in the past have been failures to act, not failures to see, interpreting those intricacies requires informed "out of the box" thinking and needs to facilitate a closer collaboration between analysts and decision-makers.

Political decision-makers and other end-users of intelligence products need to have a basic understanding of how information is collected, processed and analyzed. Regular joint simulations and scenario-planning are the method of choice to determine corridors of action and to increase the intelligence professionals' awareness of political rationale.

Business / Competitive Intelligence: Privatizing Tradecraft

Business intelligence understands that the competitive advantage information can provide for strategic market analysis lies much more in its potential for fact-based support to decision-making rather than in the source being classified or openly available. As business intelligence is first and foremost concerned with having the right perception of the marketplace (comprising adversaries, competitors, products and clients), it is invaluable for risk management as well as for counter intelligence, preventing information leaks within the own company and anticipating threats from the outside. The competitive learning process resulting from the use of a vast range of relevant analytical models and data warehouses also affects government agencies: the lessons learned there need not be learned again, but adapted or, where applicable, out-sourced wisely. In this very field, government agencies and private companies are competing with each other for dictating the pace of innovation and, thus, for recruiting the best talents. When it comes to actionable intelligence, conclusions, recommendations and sound judgment need to take center stage while any research supporting the analysis (basically presenting the facts found and sources compared) is of lesser importance.

Intelligence professionals need to closely keep track of how commercial players and academia are making use of relevant technologies and how they handle the delicate underlying legal restrictions and opportunities. It is advantageous both for the producers and consumers of intelligence products to audit if their objectives can be achieved more efficiently and cheaper by at least partly out-sourcing tasks based on open source information.

Information Science

Ambient Intelligence: Merging Technology and Application

Ambient Intelligence (AmI) is a paradigm derived from the concept that through technological innovation the environments in which we live and work are becoming more and more intelligent. This is based on the integration of advanced electronics into these environments. People, animals, machines, objects, etc. interact with their environment and with each other. This interaction is responsive, pesonalized, adaptive, context aware, anticipatory and embedded. The consequence is a completely new set of smart products (handheld devices, displays, wearables, sensors, etc.) and services which constantly emit information trails that challenge the user's privacy and might be exploited systematically. AmI has several requirements for the underlying software and hardware architecture, of which the most important are:

- Ubiquity: The system should be integrated inconspicuously in the environment of the user.
- Interoperability: Devices and environments communicate and cooperate.
- Autonomy: A system has to be autonomous and reactive.
- Dynamic: An ensemble has to be extensible by new devices.
- User-centered: The user (or the "smart player") is in the center, not the devices.

It is obvious that this very idea of "pervasive" computing, shifting the focus from technology as such to the user, not only considerably empowers the users, but that it also increases potential security threats, both for the user and the information technology behind.

The technological, societal, economic and security implications of AmI need to be understood in order to make the best use of them. With real, physical and digitally augmented environments merging together, making sense of this synthesized reality encompasses the diffusion and interconnectedness around us.

Information Abundance: Open Data and Informational Risks

The volume of information is proliferating beyond our current capacity to exploit or manage it effectively. By some estimates, the volume of digital information increases ten-fold every five years. Given the securitisation of the national agenda, it is inevitable that much of this data will have some security relevance. Current trends will drive new business models, management philosophies, operational technologies and human literacies. Many of these will have relevance to the conduct of open source intelligence. It is necessary, therefore, that such trends are monitored carefully so that the practice of intelligence remains ahead of the curve. The impending "industrial revolution" of data presages the arrival of an "information economy" as organisations seek to exploit the commercial value of freely available information. Governments are driving this trend under the banner of "Open Data". It is understood that there is huge economic potential in the information generated by governments. Moreover, the release of such data satisfies growing public demand for openness and accountability. Yet, the effective management of the tensions inherent between greater openness on the one hand, and the release of potentially sensitive information on the other must be learned. Every consumer of open source intelligence is also an unwitting provider of open source intelligence to others. The constant proliferation of data underscores the continued challenge of information overload. Increasing reliance on data generating systems and sensors produces vastly more information than most organisations have the capacity to sort, process or store effectively. Thus, greater emphasis must be placed on the importance of metadata and the proper indexing of information to improve its findability.

A taxonomy of information risks that encompasses both human and artificial systems needs to be developed and communicated as broadly as possible. While it is essential to anticipate the informational implications of future IT and intelligence gathering programs, careful consideration should also be given to ensure the interoperability of metadata and, thus, the findability of all data collected.

Additional Sources / Links of Interest

New Intelligence Standards for Interactive Analytics (Centrifuge)
http://www.centrifugesystems.com/industry/government/index.php

Ensuring the compliance through long-term auditability, Secure Time Stamp Server (Thales)
http://iss.thalesgroup.com/Products/Time%20Stamping/Time%20Stamp%20Server.aspx

Avoiding a Digital Dark Age (American Scientist)
http://www.americanscientist.org/issues/id.8795,y.0,no.,content.true,page.1,css.print/issue.aspx

Information Hazards: A Typology of Potential Harms From Knowledge" (Future of Humanity Institute)
http://www.nickbostrom.com/information-hazards.pdf

Data, Data Everywhere (The Economist)
http://www.economist.com/specialreports/displayStory.cfm?story_id=15557443

The State of the Internet (Flowing Data)
http://flowingdata.com/2010/03/01/the-state-of-the-internet/

Next Generation Connectivity: A review of broadband Internet transitions and policy from around the world (The Berkman Center for Internet and Society)
http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/Berkman_Center_Broadband_Final_Report_15Feb2010.pdf

It's Not Information Overload. It's Filter Failure (Clay Shirky at Web 2.0 Expo)
http://www.youtube.com/watch?v=LabqeJEOQyI

Intelligence-Led Policing: The New Intelligence Architecture (US National Criminal Justice Reference System)
http://www.ncjrs.gov/pdffiles1/bja/210681.pdf

Competing on Analytics (Babson College)
http://www.babsonknowledge.org/analytics.pdf

Information Evaluation as a Decision Support for Counter-Terrorism (Thales/NATO)
http://ftp.rta.nato.int/public//PubFullText/RTO/MP/RTO-MP-IST-086///MP-IST-086-14.doc

How to identify credible sources on the Web (US Joint Military Intelligence College)
http://daxrnorman.googlepages.com/5-CompleteThesis-May08.pdf

Emerging Threats in the 21st Century: Strategic Foresight and Warning (CSS)
http://www.isn.ethz.ch/isn/Digital-Library/Publications/Detail/?ots591=CAB359A3-9328-19CC-A1D2-8023E646B22C&lng=en&id=47160

Scenarios for Ambient Intelligence in 2010 (IPTS)
ftp://ftp.cordis.europa.eu/pub/ist/docs/istagscenarios2010.pdf

NOTES

ISN
ETH Zurich

INTERNATIONAL RELATIONS AND SECURITY NETWORK
ETH Zurich | Leonhardshalde 21, LEH | 8092 Zurich | Switzerland
Tel: +41 (0)44 632 04 24 | Fax: +41 (0)44 632 14 13
www.isn.ethz.ch | osint@sipo.gess.ethz.ch

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich