Preface

Search delivering a list of documents to read is good. Next-generation information access with maps and dashboards is better. Cyber OSINT delivers what law enforcement and intelligence professionals want—high-value, actionable decision support tailored to need.

I started work on *The Enterprise Search Report* in 2003.[3] The profiles of important vendors, background information, and critical analyses ballooned from 400 pages in the first edition to more than 600 pages in the third and final edition I wrote. *ESR* was a milestone in the search-and-retrieval literature. The volume combined profiles of vendors with implementation and cost information.

But as I was completing the third edition of *ESR* in 2006, our research had documented a radical and fresh approach to information access that seemed to be gaining momentum. A handful of vendors were developing next-generation systems. Keyword search was taking a supporting role in a much larger drama.

Traditional search in these next-generation systems was a utility, a function effectively delivered with open source search systems like Lucene or Flax. The next-generation information access systems made extensive use of what I call for lack of a better term "smart software" and tools to automate many functions. For example, reports as well as simple to complex visualization were generated by the system in response to new information discovered by that system. The output directly supported tactical decisions with ready-to-use information. Time spent in manual clicking and scanning was shifted to thinking about the information the system flagged. Next-generation systems have gained a number of customers who find the new systems more in step with their needs than 50-year-old keyword search.

---

3. Originally published by CMSWatch (now Real Story) in 2004. The last edition which I wrote was the 3rd edition which appeared in late 2006.

This *CyberOSINT* monograph has a single goal: Provide an individual engaged in investigatory work, military operations, and intelligence activities with a collection of technology capability summaries. This volume is a ready reference to the functions of a next-generation information access system and to some of the vendors offering these platforms, solutions, and software today.

The monograph contains an introduction to next-generation information access (NGIA) systems and their creators and vendors. These are companies providing systems that organize and orchestrate collection, analytics, and output into a cohesive intelligence support system. The book consists of information for a person who consumes and acts upon intelligence. The monograph is not aimed at programmers, but the volume contains information that technical professionals will find useful. The key findings from our research and the discussion of future developments provide a framework for thinking about automated systems for numerous applications.

The researchers assisting with this project identified via interviews and secondary research specific companies delivering next-generation intelligence solutions. Some of the companies discussed declined to provide current information about their products and services. We have used open source information to provide a useful but probably basic introduction to what are extremely complex software systems.

The information presented is derived from unclassified sources, interviews we have conducted over the years with information retrieval experts, and knowledge derived from our consulting work with some of the firms included in the report. The research team assisting me consisted of librarians, journalists, and analysts.[4] From an initial pool of about 60 organizations, we selected more than 20 commercial enterprises offering next generation information access systems or what we call "NGIA" systems. The companies selected for inclusion in this monograph meet three criteria the research team formulated:

1 They currently provide high-value information processing solutions, platforms, and systems to law enforcement, security, and intelligence organizations in the US and elsewhere

2 They have state-of-the art technology for identifying and presenting information from large flows of OSINT

3 They make use of artificial intelligence and predictive analytics to generate outputs about otherwise unnoticed facts, trends, and events.

---

4. The members of the CyberOSINT: Next Generation Information Access monograph included Constance Ard, Don Anderson, Whitney Grace, Ric Manning, Anthony Safina, Melanie Sazegar, Stuart Schram, Robert Steele, and Dawn Yankeelov, whom I wish to thank for their work. However, any errors in the monograph are my responsibility.

The cyber OSINT functions for multi-language translation technology and the geographic manipulation of OSINT are so important that we have created chapters addressing these topics and identifying important vendors of these systems. We have included a chapter about the use of OSINT in network devices that protect the perimeter of an organization from threats outside the perimeter as well as inside the perimeter.

Also, the monograph provides a bird's-eye view of the components comprising an NGIA system. The discussions of more than 20 companies, both large and small, offers an overview not available in any other review of automated collection, automated analysis, and automated reports for law enforcement and intelligence professionals.

Much of the information in this volume will have direct applicability to a number of security and competitive intelligence challenges that commercial and not-for-profit organizations face. Most of our examples are tailored to governmental entities engaged in investigations, warfighting, and intelligence activities.

The information in this monograph has been gathered from unclassified, open source materials. Most of the companies innovate rapidly. The screenshots and some of the sample reports are almost certain to be outputs of subsystems that have been upgraded.

This is not an academic research report. We have tried to use plain English to describe the technologies presented in this monograph. The phrase Big Data appears, but we have limited its use. "Automated collection" means that flow of structured and unstructured information from web sites, newsfeeds, Facebook, Twitter, and other sources are "big." We have included a glossary to provide definitions of some words and phrases that require additional commentary; for example, *Monte Carlo method*.

Unlike some reports from consulting firms, the companies profiled in this monograph did not pay to be included, an increasingly common practice for industry surveys and trend analyses. We did seek information from each company profiled, and a handful responded. Most ignored our requests for information.

Finally, this monograph includes a bare minimum of theory, deep historical background, and Booz, Allen & Hamilton-type management discussion. Our purpose is to inform a law enforcement and intelligence professional, not sell consulting or products.

We view manual systems for collection and analysis of unstructured information and structured data as traditional tools. The 10th-century tools of the compass and horse have mostly been superceded by more sophisticated equipment like attack helicopters equipped with modern electronics.

I want to reiterate that the automated collection, analysis, and output systems available today require the involvement of trained professionals. The most advanced systems amplify the work of humans—a type of force multiplier.

This monograph is not a justification for reducing a workforce. If anything, the monograph maps out areas in which a law enforcement, security, or intelligence professional can increase his/her value by learning about these systems. Automated systems are going to play an increasing role in the future. An individual with knowledge of the systems and methods becomes an even more valuable colleague with cyber OSINT expertise.

NGIA systems are not the end of the road for information retrieval. NGIA systems mark a new beginning for research, development, innovation, and high-value software. The challenges identified in the closing chapter are ones that cannot be resolved quickly. Progress is evident, and NGIA systems are and will continue to be a potent tools for law enforcement and intelligence.

Stephen E. Arnold

Harrod's Creek, Kentucky

January 22, 2015