Managing Semantic Big Data for Intelligence

Anne-Claire Boury-Brisset
Defence Research and Development Canada
Québec, Canada
anne-claire.boury-brisset@drdc-rddc.gc.ca

***Abstract*— All-source intelligence production involves the collection and analysis of intelligence data provided in various formats (raw data from sensors, imagery, text-based from human reports, etc.) and distributed across heterogeneous data stores. The advance in sensing technologies, the acquisition of new sensors, and use of mobile devices result in the production of an overwhelming amount of sensed data, that augment the challenges to transform these raw data into useful, actionable intelligence in a timely manner. Leveraging recent advances in data integration, Semantic Web and Big Data technologies, we are adapting key concepts of unified dataspaces and semantic enrichment for the design and implementation of a R&D intelligence data integration platform MIDIS (Multi-Intelligence Data Integration Services). The development of this scalable data integration platform rests on the layered dataspace approach, makes use of recent Big Data technologies and leverages ontological models, and semantic-based analysis services developed for various purposes as part of the semantic layer.**

Keywords—intelligence, data integration, knowledge extraction, ontology, Big Data

I. Introduction

The advance in sensing technologies, the acquisition of new sensors, and use of mobile devices result in the production of an overwhelming amount of sensed data, that augment the challenges to transform these raw data into useful, actionable intelligence in a timely manner. Consequently, intelligence operators and analysts have to deal with ever-increasing amounts of ISR data and information from various sources (SIGINT, IMINT, GEOINT, HUMINT, OSINT, etc.), produced in disparate multiple media formats (raw data sets from sensors, e.g., video, images, sound files, as well as human reports and open source text), and distributed across different systems and data stores.

As part of a research project conducted within the Intelligence and Information Section at Defence Research and development Canada (DRDC) – Valcartier, we are investigating advanced concepts, techniques and technologies in order to provide enhanced capabilities for the management and integration of large-scale heterogeneous information sources and intelligence products made available to intelligence operators and officers in support of the production of intelligence and sense-making activities.

Our ultimate goals are to:

- Provide timely and relevant information to the analyst through intuitive search and discovery mechanisms;
- Provide a framework facilitating the integration of heterogeneous unstructured and structured data, enabling Hard/Soft fusion and preparing for various analytics exploitation.

This paper describes ongoing research for the design and implementation of a prototype for scalable Multi-Intelligence Data Integration Services (MIDIS) in support of these objectives, based on a flexible data integration approach, making use of Semantic Web and Big Data technologies. The paper is organized as follows. In the next section, we present recent work addressing multi-intelligence data integration, followed by a short introduction to Big Data challenges. Section IV describes the proposed architecture for large-scale intelligence data integration and analysis and details the main components of the resulting architecture. Section V provides details about the implementation using Big Data technologies. Section VI provides some conclusions and directions for future work.

II. Multi-Intelligence Data Integration

Intelligence is about data management and processing: 1) data collection from various sources, 2) data analysis for the production of intelligence and 3) dissemination of intelligence products. Intelligence data management nowadays presents the following characteristics:

- Increase of sensor data volume (terabytes to exabytes);
- Heterogeneity: multiple data formats and standards, mix of structured and unstructured;
- Need to quickly acquire and process intelligence information;
- Agility is required to be able to incorporate new data sources;
- Support to data exploitation: each piece of data represents some part of a situation, intelligence data contain entities that must be understood and correlated.

Data integration aims at combining data that reside at distributed, autonomous, and heterogeneous data sources into a single consistent view of the data [7]. Traditional approaches propose either centralized or federated data integration. The centralized approach requires heavy pre-processing through extract, transform load (ETL) processes while the latter can denote performance and complex transformations issues. These approaches have been largely detailed and challenged in the literature, and they have been recently exposed by Singleton [19] as part of a research work in the military domain.

As an alternative to these approaches to cope with large-scale heterogeneous data management, Franklin, Halevy and colleagues [11] proposed the concept of dataspaces as a new abstraction for information management. That is, it promotes a flexible co-existence approach for the incorporation of heterogeneous data into a dataspace, and a description of the concepts of the domain at a higher-level of abstraction. Integration in terms of schema harmonization is realized in a pay-as-you-go approach [12].

Looking for a flexible data integration solution to deal with the ever increasing heterogeneous data sources in the intelligence domain and information fusion, S. Yoakum-Stover proposed a framework to implement this scheme [20, 21]. Based on that approach, D. Salmen and colleagues [16] described their implementation of the approach. It rests on the definition of a data integration framework (DRIF), also called Data Description Framework (DDF) in previous papers, based on a unified data integration model. The idea is to define a simple data representation scheme to encapsulate every piece of data from heterogeneous sources into a unified representation. The elementary constructs are composed of signs, terms, concepts, predicates and statements, the latter being conceptually similar to the Semantic Web Resource Description Framework (RDF) triple composed of subject, predicate, and object.

Based on this unified scheme, the dataspace is organized into several layers, namely:

- Segment 0 contains the external data sources and systems from which relevant data are extracted;
- Segment 1 (unstructured data) represents the data store for artefacts;
- Segment 2 (structured data) is the universal store for data structured according to the unified representation scheme;
- Segment 3 (data models) contains the representation of data models and ontologies to facilitate the mapping and integration of heterogeneous data.

The concepts underlying the unified dataspace have been implemented as part of the US Army's Distributed Common Ground System (DCGS-A) Cloud initiative [17]. Moreover, to address semantic heterogeneity, B. Smith and colleagues [17] propose a strategy for the integration of diverse data through semantic enhancement, by adding a semantic layer to the data (explicitly represented in segment 3).

Leveraging this approach, we are adapting the underlying concepts for the design and implementation of a R&D intelligence data integration platform MIDIS (Multi-Intelligence Data Integration Services) to meet our requirements in support of intelligence. In previous research, our team has developed several intelligence support tools in support of collation and intelligence production, and knowledge-based systems on top of military domain ontologies to meet various analysis requirements [15]. Some of the relevant components from these tools have been incorporated as services as part of a SOA-based Intelligence Science and Technology Integration platform (ISTIP) in development.

The data access component had to be further developed in this platform to provide the ability to dynamically ingest, integrate and manage data from various intelligence sources. Consequently, MIDIS aims at enriching the data access component of the ISTIP platform to provide the set of services needed to ingest multiple intelligence data formats available, transform them into a unified model, and make these data accessible, searchable and exploitable (e.g. data mining) in support of intelligence analysis.

The design and development of MIDIS as a scalable data integration platform rests on the layered dataspace approach and makes use of Big Data technologies. Moreover, we leverage ontological models, and semantic-based analysis services developed for various purposes as part of the semantic layer within the architecture described in section IV.

III. Big Data Challenges

Considering the huge amount of data produced every day in both the commercial and the defense areas, the Big Data paradigm promotes novel approaches and technologies for data capture, storage and analytics to deal with "massive volume of unstructured and structured that cannot be managed and processed with traditional databases and software approaches" [3].

Big Data are initially characterized according to 3 Vs, namely: 1) Volume or scalability: ability to manage increasing volumes of data, for storage and analysis; 2) Variety: heterogeneity of data types, data formats, semantic interpretation; 3) Velocity: timeliness or rate at which the data arrives and time in which it must be acted upon. Additional Vs are sometimes added, to denote the Veracity of data, as well as the Value that can be extracted from Big Data.

The problem of information overload is not new, but it is amplified in the new information era. Big Data challenges encompass most data management processes, i.e. data capture, curation, storage, search, sharing, analysis, and visualization. In our research work, we are interested by Big Data solutions for on-the-fly integration of heterogeneous data from various sources, effective search among heterogeneous possibly inconsistent data sets, while managing data granularity and consistency. Some of these will be discussed later in the paper.

IV. Architecture components

The implementation of the unified dataspace approach points toward Big Data technological solutions, as they provide scalability, elasticity, replication, fault-tolerance, and parallel processing. Next, we present the proposed global architecture for intelligence data integration, its main components (data ingestion process, ontology support and semantic enrichment, search and analytics), and interactions with other reasoning modules.

A. Global architecture: from Collection to Analysis

Figure 1 represents the high-level architecture and data flow, from data collected from heterogeneous data stores, their ingestion into the dataspace, to intelligence analysis by specialized reasoning services. The key components include:

- Data ingestion from heterogeneous sources formats and integration into the unified dataspace segments;
- Ontology-based semantic enrichment;
- Data querying and analytics;
- Interactions with external reasoning modules.

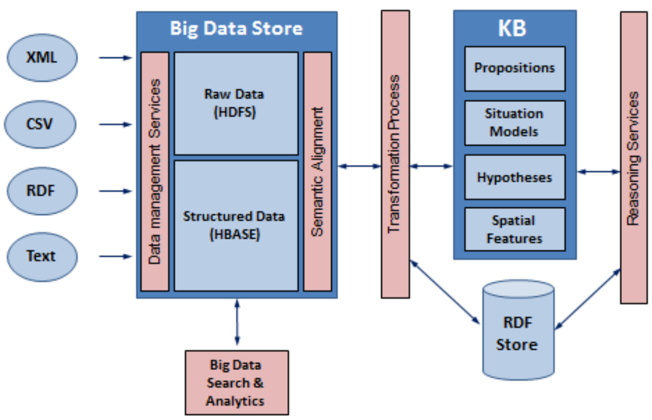

Figure 1: Intelligence data integration and analysis framework

B. Data ingestion

The system ingests intelligence data from representative sources provided in heterogeneous formats, in order to illustrate the integration of a variety of intelligence data as used by intelligence analysts to conduct multi-intelligence all-source analysis. A subset of the considered data sources in this context include:

- Structured data coming from intelligence or operational database, including track data;
- Intelligence reports;
- Imagery database;
- Data from a Content Management System;
- Internet open source (e.g. Twitter).

The data ingestion pipeline is applied to structured and unstructured data (cf. Fig. 2) as follows. Figure 2 illustrates the data flow and transformation process from external data sources, and shows explicitly how data pieces move to different segments.

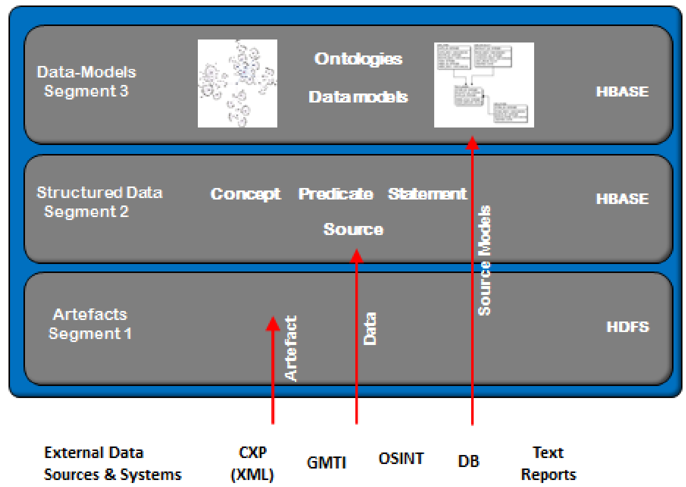

Figure 2: UDS layered architecture and data flow (adapted from Yoakum-Stover, 2012 [22])

1) Structured data

The ingestion pipeline for structured data processes various structured data sources (RDB, CSV, XML, RDF format) in order to populate the UDS in segments 1, 2, 3. The approach makes use of a XML configuration file generic enough to process each data schema provided (e.g. WSDL web service provides the XML schema to be processed). Data files are then parsed to extract data of interest and load them according to UDS constructs, i.e. concepts, predicates, statements into the UDS and reference to the source in segment 2, source model in segment 3, while the imported data source is ingested in segment 1.

2) Unstructured data: annotation and extraction

Unstructured data (e.g. intelligence reports, documents) is processed according to a text analysis pipeline using semantic annotation and knowledge extraction services supported by domain ontologies. Documents are analyzed and semantically annotated using concepts instances from domain ontologies (named entities, people, location, …). Then, knowledge in the form of statements (e.g. X is_located_at Loc) is extracted using pattern matching rules. These processes use the popular GATE platform (General Architecture for Text Engineering) [4] as the underlying natural language processing component. Documents and their annotations are stored in the segment 1 while extracted facts and metadata that provide meta-information about the documents are stored in the segment 2 according to the unified model (structured data). Metadata of interest include data provenance, uncertainty, temporal and spatial information.

In military intelligence context, imagery data sources (images, videos) are currently managed using metadata according to standard agreements (e.g. Stanag 4559) to facilitate information sharing (e.g. coalition operations). The next step in our architecture will be to adapt and enrich the data ingestion process for this type of source, possibly including automated information extraction.

C. *Ontology and semantic enrichment*

The proposed integration approach rests on the exploitation of domain ontologies to facilitate the harmonization of data models in a flexible and incremental manner.

1) Ontology engineering

Ontologies describe flexible and extensible conceptual models that explicitly represent the concepts in a domain of interest and the relationships that exist between them. Ontologies have been considered as an enabler for information integration and have also been exploited in support of information management or reasoning to meet different needs:

- To provide a standardized vocabulary and a taxonomy of the concepts in the domain of interest and facilitate information sharing;
- To support text analysis and semantic annotation;
- To perform federated semantic searches;
- To perform automated reasoning on top of the ontology and business rules;
- As a knowledge base (instances, relationships) to capture information about the domain.

In the military domain, ontologies have been developed for the last decade to meet various requirements: ontologies in support of command and control [13], low-level and high-level information fusion, in particular situation and threat assessment [1], or intelligence analysis [18, 2].

At DRDC, domain ontologies have been developed and exploited in order to fulfill command and control as well as intelligence requirements in different specific application contexts, namely:

- Maritime domain ontology in support of threat analysis and anomaly detection.
- Situation awareness ontologies to support knowledge management and knowledge mapping applications.
- Ontologies related to terrorism and Improvised Explosives Devices (IED) for ontology-based semantic annotation of texts in support of intelligence collation.

In the evolving military context, such as counterinsurgency and counter-terrorism, cyber-warfare, civil-military operations, the human terrain is a key component. The National Geospatial intelligence Agency (NGA) has undertaken the development of human geography data standards and models that define top-level constructs and a set of sub-models encompassing topics of interest such as religion, language, demographics, ethnicity, groups, culture among others. The key high-level concepts are composed of *Feature* to represent temporally persistent real-world phenomenon, *Event* to represent instantaneous or short-duration real-world phenomenon, *Actor* to represent an intentional entity that acts or has the capability of acting as a participant in an event (individuals, groups), and *Information* to collect non-geometric properties of other entity types. Based on these models, we have developed an ontology of human geography to formally represent the entities present in these models, thus enabling automated reasoning upon it. These models provide knowledge to support applications such as the Intelligence Preparation of the Operational Environment, terrain analysis, and social network analysis that require a formal representation of the human terrain elements.

In some of our previously developed ontological models, concepts are derived from the hierarchy structure of the JC3IEDM (Joint Command and Control Information Exchange Data model) and its subsequent MIP Information Model revisited and represented as a UML model. The model decomposes battlespace entities along *Objects* and *Action/Event* high-level concepts. Consequently, key high-level concepts contained in such ontologies comprise: individuals, groups and organizations, events that occur and activities that are conducted in the area of operations, their location, the characterization of the reported information, etc. Ontologies also formally represent the relationships that may exist between these entities. Of course, the spatio-temporal dimension inevitably associated to these concepts has to be modelled accordingly.

Domain ontologies are developed incrementally by adapting recognized multi-stages development methodologies, leveraging as much as possible military models and doctrine documents. Such development approaches promote a modular, layered approach to ontology construction, built on top of foundational or upper-ontologies (e.g. SUMO, BFO, Dolce, etc.) that represent generic concepts, which can be further extended to represent more specific concepts in the domain of interest according to a hierarchical taxonomic structure.

In the intelligence domain, the set of concepts of interest is derived from a thorough analysis of key processes and data sources, e.g. collation and analysis phases, in order to capture the essential entities in the ontological model. While elements of such knowledge are captured in some existing models, it is of interest to develop the corresponding ontological models and integrate them on top of some upper-level ontologies. Looking at the high-level concepts taxonomy of our ontologies, and some existing upper ontologies mentioned above, they present similarities in the high-level decomposition. BFO (Basic Formal Ontology) [13, 18] as well as the UCore Semantic Layer are models that we are leveraging to benefit from prior modeling efforts. We are revisiting and integrating them as part of this work.

Moreover, domain ontologies are being extended as new data sources or applications required additional concepts to be considered, and as the domain evolves (e.g. human terrain, cyber). As mentioned in [18], rigorous management and governance principles have to be applied to ensure consistency and non-redundancy.

Domain ontologies are developed using the OWL language based on Description Logic due to its popularity, interoperability facilitating the reuse of ontology parts, expressiveness and tractability to represent domain knowledge with expressive semantics. Consistency checking tools are used to ensure that the developed ontologies are free of inconsistencies.

2) Semantic Enrichment

Semantic Enrichment (SE) [17] is a process for horizontal data integration based on the use of ontologies to integrate and semantically enhance data models. The enhancement is accomplished by annotating (tagging) the models by the terms of the ontology(ies), thus linking together the various resources in a semantically coherent way.

According to the layered organizational structure of data in the unified dataspace, the suite of ontologies and source data models are part of segment 3. Mappings between terms of the ontologies and labels in the data models are explicitly defined at this level too, so that data models are harmonized using the semantic layer.

Consequently, using this extra semantic layer, additional semantic power (inferencing) can be exploited by query engines, or reasoners (e.g. exploiting "same_as" relations between terms linked by the same concept in the reference ontology).

To fulfill semantic enrichment approach consisting of semantically linking data, unstructured documents are also processed by exploiting the terms and structure of ontologies.

D. Data search and analytics

As mentioned above, this work leverages and extends previous research we have conducted in support of intelligence, e.g. the provision of information management and exploitation services to support the analyst in his activities: semantic search engines, filtering, notification/alert services, etc.

The focus in the present research is to provide scalable solutions for large-scale data management and analysis. Consequently, we are investigating various techniques and solutions that fulfill analysts' increasing needs in terms of:

- Analytics from large data sets: data mining, data/document clustering, data correlation among various data sets, etc.
- Efficient search and retrieval within unstructured and structured data sets.

Multi-intelligence data are ingested into the dataspace segments 1 and 2 as presented above. Consequently, efficient indexing and search techniques and tools have to be proposed both for data in segments 1 (unstructured world) and in segment 2 (structured data). While analytics tools benefit from Big Data technologies (batch distributed processing), the required search tools have to provide real-time performance results. Some techniques are discussed in section V.

E. Interface with intelligence reasoning modules

While MIDIS first aims at integrating intelligence data from heterogeneous data sources for further retrieval and exploitation, it is part of a comprehensive architecture (ISTIP) for the analysis and production of intelligence. Thus, interfaces to facilitate data flow/transformation between the UDS and reasoning components are required (cf. Fig. 1). Consequently, we provide mechanisms and services to export data through a transformation process into appropriate formats to/from existing intelligence analysis modules.

- Intelligence reasoning services make use of various rich data formats required as input by their engine (e.g. rule-based reasoning and/or case-based reasoning), e.g. propositions, situation model, spatial feature, hypotheses structures.
- Data can also be exported as RDF into a graph representation to be used by various reasoning services, e.g. social network analysis algorithms.

Inversely, data produced by the various reasoning modules can be persisted in the dataspace. They are ingested back as new data in the UDS through the appropriate transformation process, thus made discoverable for subsequent processing.

V. Technological aspects

The implementation of our multi-intelligence data integration system leverages emerging Big Data and SOA technologies.

A. Big Data Technologies

To cope with the processing of ultra-large scale data sets, Big Data technologies exploit distributed storage and processing. The open source Apache Hadoop Framework [5] allows for the distributed processing of large data sets across clusters of computers using simple programming models. It provides several components, including the MapReduce distributed data-processing model, Hadoop Distributed File System (HDFS), and HBase [6] distributed table store. These main components and emerging tools are being exploited for the implementation of our integration architecture (Cloudera's platform).

1) Data ingestion

Data ingestion benefits from Hadoop MapReduce distributed processing for large data sets. As presented above, structured data ingestion is done by using a XML configuration file for each data format. Data files are then parsed via MapReduce and loaded into the UDS.

Artefacts data are stored in HDFS in segment 1, structured data are stored in HBase in segment 2, and data models in segment 3 in HBase as well.

Knowledge extraction from textual documents using semantic text analysis services were not initially implemented using parallel processing. We are considering their adaptation into Hadoop environment to benefit from distributed processing of large documents corpus and are also looking at alternate approaches such as those proposed in Lin and colleagues' book [8]. Additional envisioned services for extraction value from textual intelligence reports datasets include cross-document co-referencing in HDFS.

2) *Indexing / Query*

For users (or services) to retrieve relevant information from the HBase UDS in near real-time, we aim at providing efficient indexing and query solutions.

First, considering out of the box query tools, the Hive query engine has demonstrated poor performance. The recent Cloudera Impala query engine is being experimented, the performance is improved due to the fact that it supports direct query on HBase indexes and does not use MapReduce.

Moreover, several input data formats to the UDS will be as RDF triples (metadata extracted from text, imagery data tagging, data extracted from content management systems, etc.). Conceptually, the UDS segment 2 can be considered as a HBase quad store where the fourth element added to the triple refers to the source (named graph). We are looking at techniques to perform efficient queries to retrieve RDF data in this context (e.g. extraction of graphs for Social network analysis).

One interesting approach is provided by Rya [14] that introduces storage methods, indexing schemes, and query processing techniques that scale to billions of RDF triples across multiple nodes, while providing fast and easy access to the data through conventional query mechanisms such as SPARQL. Rya proposes a method of storing triples by indexing triples across three different tables corresponding to the permutations of triple patterns, i.e. (Subject, Predicate, Object), (Predicate, Object, Subject), and (Object, Subject, Predicate). We are experimenting with this approach, and are exploiting OpenRDF Sesame (SPARQL) for HBase [10].

Preliminary tests are being done with various data sources, as well as using the LUBM benchmark dataset [9] to assess the performance and compare with other approaches.

3) *Analytics*

While intelligence analysis requires specialized reasoning tools and human intervention, Big Data Analytics may reveal interesting insights from the analysis of large data, (e.g. predictive/trend analysis) by using appropriate techniques such as data mining. Apache Mahout is one of the first distributed machine-learning open source framework built on top of Hadoop. It is a candidate for data clustering, classification, collaborative filtering, recommendation, or profiling that we are considering in order to demonstrate value-added from data using Big data analytics.

B. *SOA*

Service Oriented Architecture (SOA) has emerged as the predominant paradigm for the building of flexible and scalable architectures in net-centric environments. SOA is an architectural discipline that relies upon the exposure of a collection of loosely-coupled, distributed services which communicate and interoperate via agreed standards across the network. Some benefits are directly based on the principles of service orientation, mainly: services are loosely coupled, autonomous, discoverable, composable and reusable. Consequently, SOA principles offer an appropriate approach to data integration. The services can be composed into higher-level applications to support agile business processes. By augmenting the data services layer, and incorporating integration services as described above, the data integration environment will facilitate access to data and discovery, integration of data from diverse sources, and handling of large volume of data.

The envisioned set of services complements the SOA-based Intelligence Science and Technology Integration platform (ISTIP) in development at DRDC Valcartier. This platform already incorporates a set of data representation schemes and relevant services in support of various intelligence analysis tasks and sense-making activities: the analysis of textual documents, (semantic annotation of text based on domain ontologies, and automated extraction of facts from documents based on pattern matching rules), as well as multiple reasoners (rule-based reasoner, case-based reasoner, multiple hypotheses situation analysis) [15]. Our contribution will augment the platform with additional intelligence data services, using flexible and efficient representation schemes. This will facilitate the linking of data among the various sources, in order to make sense of the large amount of data made available to analysts, and provide improved situational awareness.

VI. Conclusions

In this paper, we have presented the ongoing work that we are conducting for the development of a scalable and flexible intelligence data integration and analysis platform. As part of this initiative, we leverage our previous R&D work using semantic technologies, in particular the suite of ontologies and services that are part of our ISTIP platform. Moreover, we are leveraging a proposed integration approach [22] and adapting it to our needs. We are currently developing data integration components by experimenting with recent Big Data technologies to address scalability and performance.

Big Data technologies represent a shift in terms of programming approach, and their promise produce an increasing interest within the data/information management community. But proposed solutions are still immature, and first experimentations show that they require incremental development and testing stages to improve performance. In our military intelligence context, Big Data performance is critical if these technologies are be used in tactical environments.

While we aim at providing a comprehensive data management and exploitation platform, further research is required to deal with entity resolution, disambiguation, data cleaning, etc. in this context. Recent research proposed in the Big Data world should provide relevant insight.

A data integration platform can be viewed as a prerequisite to multi-sources information fusion. Work within the hard/soft information fusion community addresses similar challenges, and we looked at them from an architecture perspective. The management of data uncertainty should be considered beyond simple metadata when integrating intelligence data from heterogeneous sources.

We are also investigating approaches to the integration and exploitation of internet open sources in support of intelligence analysis, in particular from social media (e.g. twitter).

REFERENCES

[1] Boury-Brisset, A.-C. Ontology-based approach for Information Fusion, in Proceedings of the 6th International Conference on Information Fusion, Cairns, 8-11 July, Australia, pp. 522-529, 2003.

[2] V. Dragos, Developing a core ontology to improve military intelligence analysis, in International Journal of Knowledge-based and Intelligence Engineering Systems, 17, pp.29-36, IOS Press, 2013.

[3] Gartner, Hype cycle for Big Data, Gartner research report, 2010, also published in 2012.

[4] GATE, General Architecture for Text Engineering, http://gate.ac.uk/index.html.

[5] Hadoop. http://hadoop.apache.org/.

[6] HBase. http://hbase.apache.org/.

[7] M. Lenzerini, Data integration from a theoretical perspective, In Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, 2002.

[8] Jimmy Lin and Chris Dyer. Data-Intensive Text Processing with Mapreduce. Morgan and Claypool Publishers, 2013.

[9] Guo, Yuanbo, Pan, Zhengxiang and Heflin, Jeff . LUBM: A Benchmark for OWL Knowledge Base Systems. Web Semantics. 3(2) July 2005.

[10] OpenRDF. http://www.openrdf.org/.

[11] M. Franklin, A. Halevy, and D. Maier. From databases to dataspaces: A new abstraction for information management. SIGMOD Record, 34(4):27-33, December 2005.

[12] S. Jeffery, M. Franklin, and A. Halevy. Pay-as-you-go user feedback in dataspace systems. In Proc. of SIGMOD, 2008.

[13] B. Mandrick, Creating an extensible command and control ontology, in Int. Journal of Intelligent Defence Support Systems, Vol. 4, No. 3, 2011.

[14] R. Punnoose, A. Crainiceanu, and D. Rapp. Rya: a scalable RDF triple store for the clouds. In Proceedings of the 1st International Workshop on Cloud Intelligence (Cloud-I '12). ACM, New York, NY, USA, 2012.

[15] Roy, J. and Auger, A., The Multi-Intelligence Tools Suite - Supporting Research and Development in Information and Knowledge Exploitation, 16th International Command and Control Research and Technology Symposium (ICCRTS), "Collective C2 in Multinational Civil-Military Operations", Québec City, Canada, June 21-23, 2011.

[16] Salmen D., Malyuta T., Hansen A., Cronen S., and Smith B., Integration of Intelligence Data through Semantic Enhancement, in Proceedings of the 6th international conference on Semantic Technology for Intelligence, Defense, and Security (STIDS 2011), George Mason University, Fairfax, Virginia, November 2011.

[17] B. Smith, T. Malyuta, W. S. Mandrick, C. Fu, K. Parent, M. Patel, Horizontal Integration of Warfighter Intelligence Data. A Shared Semantic Resource for the Intelligence Community, Proceedings of the Conference on Semantic Technology in Intelligence, Defense and Security (STIDS), George Mason University, Fairfax, VA, October 23-25, 2012.

[18] B. Smith, T. Malyuta, D. Salmen, W. Mandrick, K. Parent, S. Bardhan, J. Johnson, Ontology for the Intelligence Analyst, CrossTalk: The Journal of Defense Software Engineering, November/December 2012,18-25.

[19] J. Singleton, Data integration: charting a path forward to 2035, Air War college, research report, Feb. 2011.

[20] S. Yoakum-Stover and T. Malyuta, "Unified data integration for Situation Management," IEEE MILCOM 2008.

[21] Yoakum-Stover S., Malyuta T., Antunes N., A Data Integration Framework with Full Spectrum Fusion Capabilities, Sensor and Information Fusion Symposium, Las Vegas, 2009

[22] S. Yoakum-Stover, A. Eick, Breaking the Data Barriers, DGI, London, 2012.