

ISR Data Processing in Military Operations

Ladislav Burita*, Ales Novak

Department of Informatics and Cyber Operations, University of Defence, Brno, 66210, Czech Republic

ARTICLE INFO

Article history:

Received: 17 June, 2020

Accepted: 31 August, 2020

Online: 17 September, 2020

Keywords:

ISR data processing

Business Intelligence

Data Mining

Big Data

ABSTRACT

This paper provides an overview of Intelligence, Surveillance, and Reconnaissance (ISR) data with respect on NATO standards and recommendations; further presents methods, tools, and experiences in ISR data processing in military operations. The steps of the Intelligence cycle and disciplines Business Intelligence (BI), Data Warehousing, Data Mining, and Big Data are presented in the introduction. The literature review is oriented to the analysis of intersections between ISR and BI methods. The next chapter describes the ISR data processing in detail; there are listed structures, formats, standards, and data from the operational point of view. The ISR operational picture is explained, and steps of the ISR data mart is completed. The last part is oriented to Big Data processing; NoSQL, in-memory and streaming databases. The last two chapters are focused on the description of research results in the processing of ISR data. The ISR data mart experiment processes the radio transmission data that consists of detected radio signals. Results are visualized in RapidMiner Studio. The Big Data experiment is realized in Apache Hadoop.

1. Introduction

This review paper is an extension of work originally presented in the 7th International Conference on Military Technologies [1]. Paper summarizes methods, tools, and experiences in ISR (Intelligence, Surveillance, and Reconnaissance) data and information processing in military operations.

The important focus in the paper is oriented on the NATO ISR data standards. The ISR data format includes in authors point of view: the structured data, unstructured (text) data and Big Data.

The ISR data processing mainly uses Business Intelligence (BI) techniques, which include data and text analysis, Data Mining (DM), Data Warehousing (DW) On Line Analytical Processing (OLAP), and Decision Support Systems (DSS).

1.1. The ISR Processes and Data

The Command and Control (C2) of the military units is mostly mentioned as OODA (Observe-Orient-Decide-Act) management cycle [2]. Each phase of the cycle can be supported by appropriate information technologies (IT). The most important phase is orientation, which affects both the creation of the right alternatives for decision and the method of observation and that functions are supported by IT. The Intelligence Cycle (IC) supports the orientation phase. The IC is defined as “The sequence of activities

whereby information is obtained, assembled, converted into intelligence and made available for users.” The IC goal is to solve commanders’ problems related to the operational situation. It consists of four phases [3]: Direction, Collection, Processing and Dissemination.

The ISR process “A coordination through which intelligence collection disciplines, collection capabilities and exploitation activities provide data, information and single-source intelligence to address an information or intelligence requirement, in a deliberate, ad hoc or dynamic time frame in support of operations planning and execution” [4] is a part of data collection in IC.

The ISR activities consist of five phases: Task, Collect, Process, Exploit and Disseminate (TCPED). In the collection phase, by the NATO documents, data are obtained from heterogeneous sources and thus covers multiple intelligence disciplines:

- Human Intelligence (HUMINT) - information that is collected and provided by human sources.
- Acoustic Intelligence (ACINT) - the detection and utilization of acoustic signals or emissions.
- Imagery Intelligence (IMINT) - collection, processing and exploitation of image sequences.
- Measurement and Signature Intelligence (MASINT) – the source of information in scientific and technical area.

*Corresponding Author: Ladislav Burita, ladislav.burita@unob.cz

- Open Source Intelligence (OSINT) – information gathered from publicly available sources.
- Signals Intelligence (SIGINT) - collecting and exploiting electromagnetic signals.

The approach of NATO to ISR area is rapidly changing because of disrupting possibilities in technology and science. In the last years, the community of innovation in NATO has been in cooperation with the NATO Support and Procurement Agency (NSPA), prepared conditions for extensive changes in the AWACS (Airborne Warning and Control System) aerial survey and the method of exchanging information.

NATO is planning the next version of AFSC (NATO Alliance Future Surveillance and Control) as an ISR distributed network systems as opposed to a single platform. This is meaning as combination of technical and human resources that would provide more enhanced surveillance capability opposed a single complex and expensive platforms. [5]

1.2. Business Intelligence

Business Intelligence (BI) is including tools, technologies, and practices data collection, aggregation or integration, processing, analysis, and presentation of in the form of information to support decision-making tasks.

All sources of data including information systems (IS) for decision preparing contain the huge volume of data. It is not possible to obtain the awaiting information for the decision process directly from the data sources and it is unnecessarily time consuming and expensive.

It must be prepared many questions that lead to the desired result. Data are spread in many sources in a various format, and therefore obtaining information is very difficult.

Commanders or directors are able to prepare decision strategy quicker and better by BI; are able to simply prepare or modify company strategies or processes, which leads to competitive advantage in a military area.

BI is an approach that has an impact on the effective decision, and thus increases the quality of the organization and leads to an advantage in warfare.

BI infrastructure includes components that collects, stores, summarizes, aggregate, integrate and analyzes the data produced by an organization or military unit. It is a broad term that encompasses data mining, data warehousing, online analytical processing (OAP), performance benchmarking, descriptive analytics, business process reengineering (BPR) and reporting, see Figure 1.

1.3. The data Warehousing and OLAP

The data warehouse (DW) is prepared as a trusted and integrated data collection for analyze processes with outputs that are directed to the decision process in a company or a military unit. It helps users solve problems and preparation decision in business and military missions.

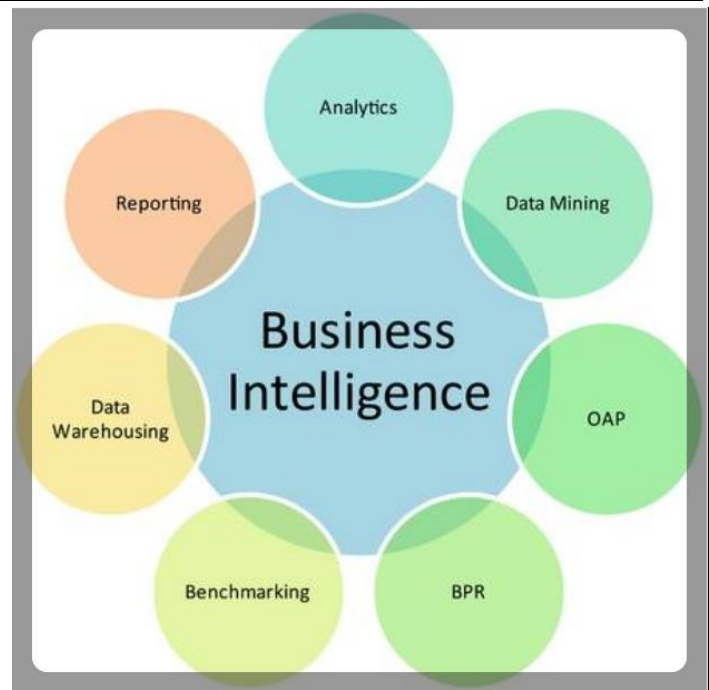


Figure 1: Business Intelligence [6]

Sources for DW are IS in business, Enterprise Resource Planning (ERP) systems, Customer Resource Management (CRM) data sources, and other data files, see Figure 2. The DW is preparing from data sources thru ETL (Extract – Transform – Load) processes.

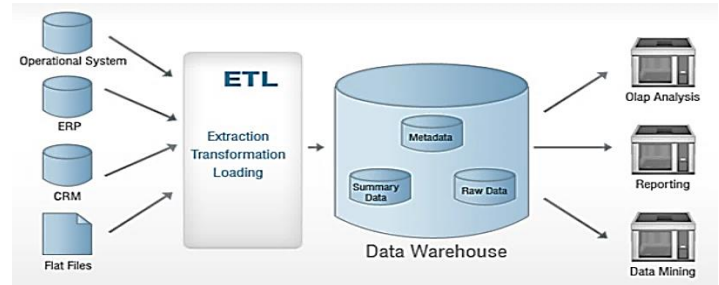


Figure 2: Data Warehousing [7]

The main reason of DW constitution is the possibilities of the IS of other data sources for decision-making support:

- Information systems have not a satisfactory performance in analysis, it takes too long, it is expensive and the historical data are usually missing.
- Data are spread in more inconsistent applications (production, commerce, human, finance, logistics, and other) resources.
- Database structure of the DW data sources are not optimal for decision-making elaboration.

OLAP (or OLAP Cube) is abbreviation of On Line Analytical Processing. It is a multidimensional memory whose dimensions form required view of data structure and many variants of detail view in information. The OLAP cube in gasoline industry consists of dimensions: Time, Product, and Measure, see Figure 3.

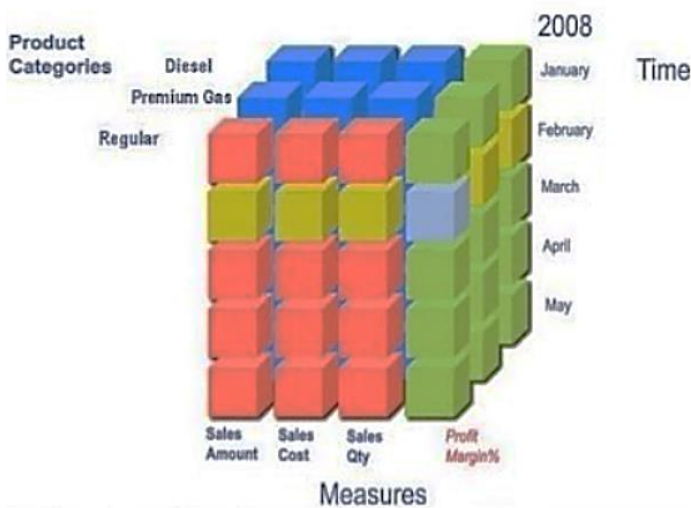


Figure 3: OLAP – the multidimensional memory [8]

Each node of the OLAP consists of product category, time, and sales results (amount, cost, quantity, and profit margin in %). In the prepared data structure, the results can be by managers good monitored from various directions.

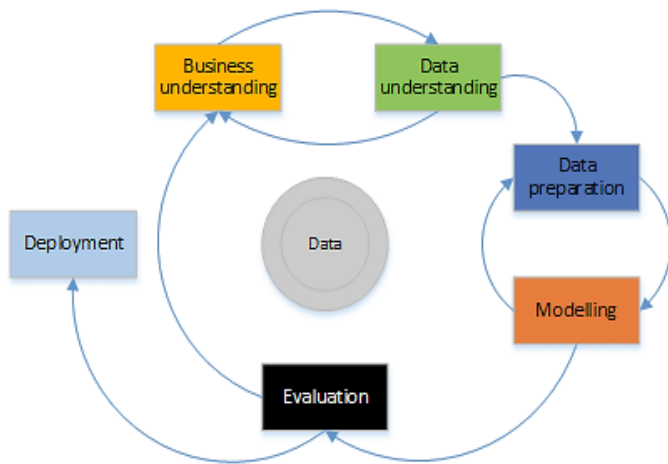


Figure 4: CRISP-DM methodology phases [9]

1.4. Data Mining, Methods and Tools

The mission of Data Mining (DM) is to find information, not known before, from databases. There are used modelling approach and statistical tools to gain knowledge from data. The usually used DM methodology is Cross Industry Standard Process (CRISP) that include processes of data acquisition, processing and transformation by application of DM method, evaluation, visualization, and presentation; see Figure 4.

The effectiveness and quality of the decision process depends on the reliable data; mistakes may lead to bad results. Mistaken data is a result of practical processing and data can be incomplete, inconsistent or should contain problems.

Set of actions in the data processing:

- Cleaning Data - completion the missing values, identify and remove boundary values, eliminate data inconsistency.

- Data Integration – various sources of data should be integrated.
- Data Transformation – aggregation, normalization and reduction makes data useful for analysis.
- Descriptive summarization of data – understanding of structure and properties of the data.
- Discretization - conversion of analogue format into digital form.

The often-used DM methods include decision trees (DT), neural networks (NN), statistical methods (SM), and genetic algorithm (GA). The popular method of DM is statistics: discrimination and regression analysis, and pivot tables.

In DT method is data divided into groups which represent the same class. The goal is to find the optimal tree by the principle that differentiate chosen attributes the best.

Method NN is based on set computational neurons with weighted connections. NN functions are set up by testing data.

GA is used if it is a problem to find the quick solution, and method is likened from the natural environment.

The most popular BI SW in the academic environment are products SAS Institute and IBM SPSS.

SAS is the statistical SW for data management, various kinds of analysis (multivariate, predictive, or statistical) for BI, a policy investigation. SW is able to process data from many sources, perform analysis on them, and makes available a graphical user interface for common user and developed solution using the special language for technical users.

The most extensive SAS solution offers for customer intelligence. It includes parts for marketing, social media and web analytics, is able to profile customers, predict their business steps, and provides optimal communication.

The SAS SW for enterprises and organizations is preparing for risk and compliance analysis, includes risk simulation, scenario building to reduce the risk and development of corporate policies.

The SAS supply-chain SW prepares forecasting for company development, solution of products distribution, inventory and optimal pricing. Exists also the SAS SW for environmental sustainability, economic and social consequences in causal relations between production and its impact on the eco-environment. The SAS has solution for government, industry, trade, and communications. [10]

The SPSS SW suite solves various kind of analysis (statistical, text, predictive), includes solution for artificial intelligence (machine learning algorithms).

SPSS offers open-source connections, integration with Big Data tools, and simply deployment into own applications.

The SPSS SW is prepared as flexible and ease to use for all categories of users (from common to advance).

SW is offered for all sizes and complexity of projects and can improve company or organization efficiency, minimize risk, and find opportunities [11].

1.5. Big Data

Big Data is a term that includes primarily theme with amount of data, its format, generation and changing of data, and processing. The first characteristics of big data is large *volume*. Simple example is possible to take from social media that millions users generate every second data that volume is constantly growing.

The next characteristics is *velocity* of generating and changing data. Simple example is possible to take from Internet of Things (IoT), which sensors generate data about your condition. Sometimes they are data streams, their processing cannot be performed by conventional analytical methods.

The third characteristics is dealing with *variety* of formats. Data can be standardized or unstandardized, formalized or structured (databases, files), unstructured or text (documents), web pages, pictures, schemas, graphs, etc.

The last two characteristics refers to the *veracity* refers to the credibility and reliability of the data, and the *value* is the importance given by company or military organization to use data. In a military environment, the reliability of data, for example when assessing the enemy's situation, is crucial.

The summary of Big Data definition concludes the five “V” specification: volume, velocity, variety, veracity and value. There are special means and techniques for big data processing, an overview of which is at Figure 5. They are divided into three groups, and 15 sub-groups:

1. Computing tools.
2. Storage tools.
3. Support technologies.

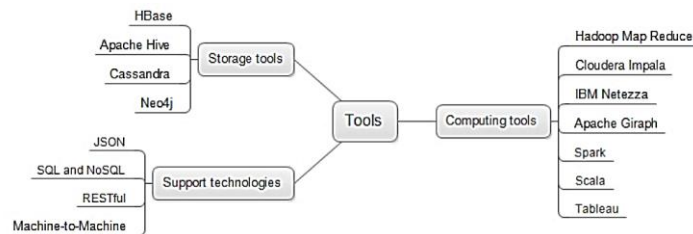


Figure 5: Mind map representation of Big Data tools [12]

As the variety, volume, veracity, value, and velocity of data is undergoing a rapid transformation and aggravating the cognitive burden on the Warfighter, armed forces require technologies to store, process, verify, and disseminate actionable intelligence at pace.

It is widely accepted that decision-making in the future information-dominated battlespace will depend on big data processing using artificial intelligence (AI) methods. It is important to address ways to gain and retain the information advantage through disruptive innovation and big data analytics, addressing cloud computing, the acquisition of data from disparate sources, data validation, visualization, accelerated and autonomous decision-making, and network resilience through cyber defense mechanisms. AI, big data technologies and ML will transform ISR and command and control (C2). [13]

2. The Literature Review

The aim of literary review is to map the interest of authors of articles, indexed on Scopus, in the topic published in the article. Information from the selected articles is oriented about the knowledge of the ISR area in relation to information technologies (IT). Overview of the selected relevant articles is listed in Table 1. The goal of individual queries was to find the relevant papers only in one area of interest. The queries were directed to all fields of the papers (the first version of the literature review, on March 26, 2020). The IT query found papers that correlate the query (BI and DW and DM and BigD) and in complex query was the result connected to ISR area. Only one paper [14] has met a complex selection requirement. The paper deals with the application of big data analytics in meteorological research.

Table 1: Number of relevant papers to the query - the first version

Query	ISR	BI	DW	DM	BigD
Individual	3 981	30 478	36 840	552 791	228 493
IT		619			
Complex	1				
ISR and DM	8 (open access, citated>0)				
ISR and BigD	9 (open access, citated>0)				

After the first version of the literature review was discovered that more than one-half of the found papers have keyword "ISR" matching only in references. Therefore, a second literature review was carried out (on April 2, 2020), where the queries were directed only to fields Title, Abstract, and Key-words of the papers. The overview of the selected relevant articles is listed in Table 2.

Table 2: Number of relevant papers to the query – the second version

Query	ISR	BI	DW	DM	BigD
Individual	1 697	7 503	12 849	169 113	81 292
IT		22			
Complex	0				
ISR and DM	15 (5, citated>0)				
ISR and BigD	12 (5, citated>0)				

The ten selected articles (see Table 3) were further analyzed in detail; see Table 3 that shows the areas of interest. Each paper was described by three themes – characteristics of the paper interest. Among the most, numerous characteristics of the analyzed papers are “military” and “data or information fusion”, these themes integrate orientation and interest of most analyzed papers with connection to the theme “ISR data processing”. This first conclusion of the analysis correlates completely with the focus and aim of the article.

Let us now get focus more closely on the analyzed articles. First, how data or information fusion is assumed in chosen articles and with what requirements is this operation connected? In addition, the second focus: Which artificial intelligence (AI) techniques, agent based, and knowledge approaches were used.

In [15], the author explains information superiority than can be achieved by application of the four RIGHT: information, time,

format, and people. However, it is only right when skilled commanders are able getting information use in praxis. Information superiority is possible only by intensive application of information and communication technologies that spread information in almost real time by remote access regardless of time and space. The condition of good cooperation on the battlefield is information interoperability and security in broaden contexts. That is why NATO is intensively developing standardization in the field of data exchange, which includes the concept Coalition Shared Data (CSD). It supports operational processes in NATO Joint ISR and by using the Intelligence Cycle.

Table 3: Areas of interest in the analyzed papers

Area of interest	Paper										Σ
	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	
Agent-based system		1			1						2
AI-Machine learning							1				1
AI-Neural network									1		1
Air Force						1					1
Cybersecurity										1	1
Data or information fusion	1	1	1	1					1		5
DNN architecture				1							1
GDPR								1			1
HPEC architecture						1					1
Information management	1										1
ISR mission support						1					1
Knowledge approach				1							1
Military	1	1	1		1	1	1		1	1	8
Navy					1					1	2
OSINT and BDA								1			1
Personal security								1			1
QuEST strategy							1				1
STACK model			1								1

Explanation to Table 3:

- AI-Artificial Intelligence
- BDA-Big Data Analytics
- DNN-Dual Node Network
- GDPR-General Data Protection Regulation
- HPEC-High Performance Embedded Computing
- OSINT-Open Source Intelligence
- QuEST-Qualia Exploitation of Sensor Technology
- STACT-Spatial, Temporal, Algorithm, and Cognition

In [16], the author presented elements to achieve integration of information (information fusion). The solution includes a distributed surveillance system and intelligence databases. Data sources comes from electronically situational awareness, controlled by military and civilian security operators and commanders in principle of ISR. The surveillance system works automatically; integrates data from aircraft systems, sensors, and background information into a military situational picture. Components for information fusion are agent-based mechanisms for integration and drawing using different reasoning techniques, knowledge modelling, information extraction and transformations.

The data fusion strategies for ISR missions, similar to previous paper, is presented in [17]. Extraction of semantic information from various set of sensors and its processing is based on new AI algorithms and computational power. Content of the project is bringing challenges for information fusion algorithms for moving targets, combination of image data with different temporal

frequency, getting the contextual information by monitoring of mobile phones. Name of solution is STACK- Spatial, Temporal, Algorithm, and Cognition.

In [18], the author described the distributed fusion systems for integration. Solution includes adaptive methods that improve usefulness data sources. The data processing cycle supports to find needed data, to develop models (situation modelling), and evaluate predictive characteristics of model (impact of prediction to success in mission). The goal of system is to improve availability, quality, variety, and actuality of data for information fusion. Machine learning techniques learn to find typical symptoms for integration data with fusion system and ontology. Extension of the Data Fusion & Resource Management Dual Node Network (DNN) architecture was applicable. The data mining methods are used to find models in data sources with help of automatically learn component behaviors by fusion procedure.

Project Albert is oriented to key areas that traditional modeling and simulation techniques often do not capture satisfactorily and uses two data management concepts, data farming and data mining, to assist in identifying areas of interest. The actual solution in Project includes four agent-based models that allow agents to interact with each other and produce emergent behaviors [19].

The Air Force Research Laboratory is developing new computing architecture, designed to provide high performance embedded computing solution to meet operational and tactical real-time processing intelligence surveillance and reconnaissance missions [20].

The achieved results in use of big data and relevant technologies, introduction of artificial intelligence (machine learning) learning are specified in [21]. The situation understanding is solved in collaboration human and machine using the Qualia Exploitation of Sensor Technology (QuEST). The QuEST is, in comparison to the limits of traditional big data processing, a completely progress and offer a higher-level system in situational awareness preparation.

Analyses, focused to open source intelligence and big data analytics with the emphasis on reconnaissance and how personal security is part of that entity. The main question is how privacy manifests itself as part of both mentioned data sources [22].

The big data from military sources are extensively multimodal and multiple intelligence, where numerous sensors along with human intelligence generate structured sensor and unstructured audio, video and textual ISR data. Data fusion at all levels remains a challenging task. While there are algorithmic stove-piped systems that work well on individual modalities. There is no system that can seamlessly integrate and correlate multi-source data; as textual, hyperspectral, and video content. It is not effective to solve integration of individual data sources gradually; better results can be achieved by integration at a higher level of the model by deep fusion of all data sources, while achieving accuracy of integration and obtaining the basic properties of signals [23].

US Navy is proposing emerging maritime strategy of command, control, communications, computers, intelligence, surveillance and reconnaissance with big data [24].

Summary in the sources of the literature review to the content of the manuscript: the one view is bringing similarities in data description and its processing and other view differences between analyzed papers and this manuscript in detail point of view at ISR data. The manuscript is specifying formats of ISR data as NATO standard (STANAG). ISR products are described and explained in relation to the command and control processes, management cycle and to the intelligence loop.

The common focus is in military area with ISR data processing, including even big data. The ISR data has sources from many various activities and is of different formats. Procedures of data processing are oriented primarily to data integration (fusion) and then to dissemination for military operations support.

In the manuscript are mentioned experiences and achievements of the Czech Armed Forces (CAF) in the described topic. It is also a comprehensive overview of the use of business intelligence tools.

The manuscript presents the results of research during the elaboration of the dissertation, too.

3. ISR Data Processing

3.1. ISR Data Overview

The Intelligence Cycle, the ISR process and its individual steps were introduced in previous chapters. In these individual steps, many heterogeneous sensor data and information are created, processed and managed. Military orders and tasks are issued to military units, civilian agencies and other cooperation organizations. Data are collected on gathering sensors platforms based on those orders. Operators and analysts process and exploit collected data and based on them they create ISR reports or other needed documents. Finally, ISR results are spread to its customers.

All sensor data, information and documents can be processed, stored and managed in many different ways. They are typically stored in the traditional File System or the Relational Database. This type of storage represents an easy way to store the data but in most cases are proprietary data formats used. There is a lack of interoperability because of proprietary data format utilization so it is difficult or even impossible to share proprietary data among different sensor platforms, units or even allies and nations. As a result, operators, analysts and commanders might not be able to understand gathered data so those data are not used in the decision preparation. NATO to reduce this interoperability lack introduced the concept of Coalition Shared Data (CSD).

CSD was standardized in STANAG 4559 as the NATO Standard ISR Interface (NSILI) [25] and its objective is to ensure interoperability for ISR documents exchange. NSILI defines the interface as a standard to search and access in ISR document libraries (IPL). CSD and NSILI are used as a synonym within NATO ISR community. For simplicity, the term NSILI will be used for the standard itself and the term CSD for its physical implementation in this article.

The CSD aims to ensure interoperable sharing of ISR products but NSILI is not designed and intended for user-friendly data querying, searching, reporting, data mining and analysis. NSILI does not specify client behavior; it is focused only on the server-side functionality and machine-to-machine interfaces, such as

other CSDs, sensor platforms and exploitation stations. Moreover, NSILI data model is not universal military data model and therefore it does not cover all relevant military data formats so all ISR data and information cannot be stored in the CSD. NSILI also defines its own dissemination and synchronization mechanism, which affects data accessibility and availability on different nodes.

The operator and analyst needs to search relevant data efficiently regardless of their formats, storage or even CSD implementation details and limits. To achieve this goal, it is proposed to build a data warehouse on top of CSDs, sensor data databases and other military data sources. The data warehouse can help to solve some issues mentioned while other difficulties would remain. Those are connected with strict NSILI data model and streaming nature of some sensors data. The Big Data approach will be utilized to cope with those aspects.

3.2. ISR Data Formats

In the following paragraphs, will be introduced relevant ISR and military data formats, which will be later used as a base for building ISR data warehouse. Many data format standards are applied in a military environment to store ISR specific data. Those data format standards include common military data formats, ISR specific ones as well as common enterprise data formats such as PDF, Office formats, images formats, video formats etc. In the article is focus only on military data formats, both common and ISR specific ones.

3.3. Common Military Data Formats

Common military data formats relevant to ISR data include textual messages, operational pictures, blue force tracks etc. These data are widely used for the exchange and integration of military systems. Relevant data standards are already implemented and adapted in the CAF and its implementation in the CAF has been confirmed on many experiments, trials, etc. Some of those standards cover not only data formats but includes also transport protocol and mechanism of exchange.

Textual messages are included in the APP-11 [26] message catalogue, which is NATO product, used in exercises and missions to exchange text as a structured information between allied military units. The messages are prepared upon the latest technical standard ADatP-3 that defines the rules of organizations and exchange of the messages. The actual version of the APP-11 message catalogue includes four hundreds messages ISR specific ones. All messages can be exchanged using the XML or slash-separated files.

Operational pictures utilize the NATO Vector Graphic (NVG) for coding and common using overlays. The NVG includes an XML data format for the coding battlespace elements into result messages and report for automated exchange of the overlays [27].

Blue Force Tracking (BFT) is the ability to track flawlessly with 100% accuracy the identification and location of own military forces in NATO missions, exercises and operations in almost real time. NATO C2 are equipped with this system and preparation for working with BFT is carefully required. NATO Friendly Force Information (NFFI) and NFFI in Message Text Format (FFI MTF) are two XML based message forms that are used by Blue Force Tracking capability [28].

3.4. ISR Specific Data Formats

The ISR community developed and adopted a wide set of data format standards for individual intelligence disciplines. The set of those ISR data standards is maintained by STANREC 4777 Ed. 2 - NATO ISR Interoperability Architecture (NIIA) [29]. Although NIIA lists about 25 key ISR standards, only the most significant for the CAF are mentioned in this article:

STANAG 4545 – NATO Secondary Imagery Format (NSIF) provides implementation guidance that is designed for the distribution, storage, and interchange of secondary imagery products.

STANAG 4559 – NATO Standard ISR Library Interfaces (NSILI) promotes interoperability of NATO ISR libraries and provides services for the exchange of shared ISR products.

STANAG 4609 – NATO Digital Motion Imagery Standard (NDMIS) guides the consistent implementation of Motion Imagery Standards to achieve interoperability in both the communication and functional use of Motion Imagery Data.

STANAG 4658 – Cooperative Electronic Support Measure Operations (CESMO) exists to enable the warfighter to rapidly share and receive electromagnetic spectrum threat information.

STANAG 5516 – Tactical Data Exchange - Link 16 provides guidelines on how to ensure interoperable use of Link 16 Tactical Data Links (TDL) to disseminate information.

The most important standard from a data integration perspective is already introduced STANAG 4559 NSILI, which will be evaluated in more details. NSILI defines architecture; use cases (publish, ingest, retrieve, synchronization); two service interfaces, modern web service interface and obsolete Common Object Request Broker Architecture (CORBA) interface; metadata model etc. NSILI covers two main functional areas, ISR Product Library and JISR Workflow.

ISR Product Library covers the first set of data managed by NSILI. It is specified as an AEDP-17 NATO STANDARD ISR LIBRARY INTERFACE. There are about 20 different types of ISR documents that are supported by standard AEDP-17.

They have format as ground moving target indication, tactical data link (TDL) reports, geographic area of interest and geospatial vector, intelligence information requirements, images and video, Microsoft office files, request for information (RFI), operational rules, organization structures (ORBAT), system specifications and assignments; chemical, biological, radiological and nuclear information; electronic order of battle, system deployment status, messages and tactical symbols.

AEDP-17 defines the following services:

- The service to metadata of query entries of ISR documents included within CSD servers.
- The publishing service to publish and modify the content of CSD servers.

The metadata set valid to all ISR products is specified by NSILI. This set contains ISR document status, source library, publisher, date and time of modification, security classification

and ability for publication, product type, identifier of mission, used language, comments, related geographical data etc.

NSILI is defining as an additional group of metadata for each ISR document. The additional set of metadata extends the common set of metadata by attributes relevant only to given ISR product type; it is containing logical data model of metadata but it does not prescribe how it should be implemented on the physical level. Query and publish services defines XML based structure of metadata (Catalogue Entry Data Set) which can be exchanged among multiple CSDs.

JISR Workflow is covering the second set of data managed by NSILI. It is specified as an AEDP-19 NATO STANDARD ISR WORKFLOW ARCHITECTURE [30]. AEDP-19 defines Community of Interest (COI) services to manipulate following data entities: organization, requirement, priority, task, request and geographic area of interest (GAOI). COI services utilize underlying Simple Persistence Service (SPS++) to store entities.

SPS++ stores entities as XML artefacts in a content-agnostic way and allows easy synchronization among nodes. On the other hand, SPS++ does not allow efficient search, because the whole entity is stored as one (atomic) XML artefact without any auxiliary attributes.

Both sets of services, ISR product Publish/Query services and JISR Workflow services are intended for operational purposes mainly, not for data analyses. The support for efficient data searching is very limited because

IPL introduces its own specific query language, which enables to search ISR products according to the value of attributes, and its combination but extended data querying is limited. JISR WF services provide even only predefined business services without any extension mechanism available.

3.5. ISR Data from Operational Point of View

In the previous sub-chapter, selected data formats were described from a technology point of view. The following section will focus on its utilization in ISR operations from an operational point of view. The ISR tasks, sensors data, ISR products, and ISR operational pictures will be discussed.

3.6. ISR Tasks

The ISR task is a directive for the appropriate employment of ISR assets. There are software tools utilized to manage ISR tasks and visualize them in different perspectives (time, unit, GAOI) to depict ISR tasks decomposition, dependencies or conflicts. The main goal is to effectively coordinate and manage data gathering activities. [31]

In NATO operations, ISR tasks should be managed within the Information Request Management and Collection Management (IRM&CM) process [32].

IRM&CM process is corresponding with Task and Collect phases of the ISR process and it aims to decompose commanders' tasks in form of Request for Information (RFI) into the hierarchical structure of Priority Information Requirements (PIR), Specific Intelligence Requirements (SIR), Essential Elements of Information (EEI). ISR tasks may be refined into specific (proprietary) orders to allow automated tasking of ISR assets.

The overview lists and plans are created based on this decomposition and it consists of a list of approved and prioritized ISR collection requirements, ISR tasks and dynamic re-tasking priorities. [33]

Those lists and plans include Intelligence Collection Plan (ICP), Collection Exploitation Plan (CXP) and Collection Tasking List (CTL), which are intended to track and monitor the operation of data collection.

It aims to assist in producing, completing and monitoring intelligence requirements, to indicate how each requirement is to be satisfied, to visualize tasks structure and de-conflict tasks.

The general structure of an ISR task includes identification of the task, name of the task, description of the task, associated geographic area, date and time of task validity, sensor or asset responsible for collecting activity according to organization structure (ORBAT), associated PIRs, SIRs, EEIs and other detailed parameters.

There are a lot of technology standards used to store ISR tasks:

- Structure of requirements, tasks and requests defined in AEDP-19, which should be the preferred option.
- Tasks format defined in AEDP-17.
- APP11 messages stored in the file system or the database.
- Intelligence discipline-specific formats, e.g. CESMO requests.
- Proprietary textual or binary formats stored as files.
- Proprietary database structures.
- Office documents (Word, Excel, PDF), and others.

It is not possible to create an overall overview of all tasks due to diversity of used data formats. To solve this issue data mart could be built on the top of multiple ISR tasks storages. Building this data mart is challenging due to the fact of a variety of data formats used. Tens or even hundred APP11 message types, individual disciplines data formats, office documents, AEDP-17 and AEDP-19 XML structures are in a place for the ETL phase. Due to the number of data sources and used data formats, implementation of ETL phase might include hundreds of transformation rules and its maintenance might be expensive. Also, rigid data structures of data mart might not be flexible enough, more flexible Big Data technologies could be utilized to build a data mart and mine the data.

3.7. Sensor Data

Sensor data are the main output of the Collection phase of the ISR process. Sensor data serves as input for the analyst's work. Although intelligence disciplines cover multiple data sources, this article and experiments focus, only on those sensors data formats which are used in the CAF, especially IMINT and COMMS-ESM ones.

COMMS-ESM data formats

Sensor data formats for COMMS-ESM includes detected activities, direction findings and geolocations. Detected activity (DA) is a logical data entity describing detected radio signals and

its parameters. DA is produced by sensor platform based on (multiple) sensors (activity detectors) raw data.

DA consists of physical parameters such as carrier frequency, quality and level of the detected signal, date and time of the detection, transmission time, type of signal, type of the modulation and other technical information depending on the sensor and its capabilities. All parameters are the same within one detected activity, when one of the parameters changes, the new activity is automatically detected and recorded.

Direction finding (DF) is a logical data entity describing direction from an emitter (to the receiver) of the radio signal and its parameters. DF is produced by sensor platform based on (multiple) sensors (direction finders) raw data. DF consists of physical parameters such as direction, focus error, carrier frequency, quality and the level of detected signal, date and time of the detection, transmission time, and other technical information depending on the sensor and its capabilities. DF is associated with one or multiple DAs.

The geolocation is a logical data entity describing a position of the emitter and error ellipse. The geolocation can be produced on the sensor platform or commands post, it can be calculated automatically or set manually, both options are based on the intersection of DFs.

All mentioned entities (DA, DF, geolocation) are standardized in STANAG 4658 CESMO but proprietary formats are widely used on most sensor platforms, as captured physical parameters are always sensor specific. Gathered data are transformed into CESMO format in the case that sensors and assets are running CESMO operation and they are connected into the CESMO network.

CSD does not support those entities (DA, DF, geolocation) as they are in the scope of Electronic Warfare (EW) rather than in the scope of ISR. Due to this fact, EW and SIGINT data are stored in specific (proprietary) databases, which can be utilized for building ISR data mart. Due to the diversity, volume and streaming nature of the data, the Big Data approach can be considered in processing those data.

Imagery data formats

Imagery data formats are intended to store images (photos), video sequences and video streams. Many proprietary and standard-based data formats are used in this area to cover the complexity and diversity of image and video shooting. Imagery data formats contain not only imagery data itself but also image/video metadata such as a sensor position, heading and telemetry, camera technical details (field of view, lens parameters, etc.), camera type (electro-optical, infra-red, thermal, etc.), picture technical parameters (resolution, dimension, bands, encoding, etc.), video technical parameters (resolution, dimension, sampling frequency, encoding, etc.). Imagery analysts can utilize only a few of those metadata; the majority of them is rather technical.

Imagery data and video streams and clips are well supported in NSILI; however, this support is limited only to images compliant to STANAG 4545 NSIF and video clips and stream compliant to STANAG 4609 NDIMS. In the case that other picture or video formats are used, the data should be converted and enriched to

supported formats or it can be stored as a simple document type without taking advantage of imagery and/or video metadata.

STANAG 4545 NSIF is a container for images and its metadata. Imagery data and metadata are combined into one complex binary file. Industry based and proven image formats and codecs such as JPG, JPG2000, not compressed images etc. are utilized for images encoding. Common NSILI metadata set is extended for imagery products with parameters such as imagery category, cloud cover percentage, decompression technique, number of bands, number of rows, number of columns and title.

Common NSILI metadata set is extended for video products with parameters such as average bitrate, category, encoding schema, framerate, number of rows, number of columns, metadata-encoding scheme, MISM level, scanning mode, number of VMTI target reports and VMTI processed. Some of those parameters are very technical ones and they are not suitable for the operator to search and query.

Even in the case that industry data formats are used for imagery data, e.g. JPG captured by the digital camera, they can contain useful metadata stored in EXchangeable Image File Format (EXIF). EXIF is an industry data format for digital photography and enables to enrich raw imagery data by metadata such as geoposition, date, time, camera technical parameters such as pixel dimensions, ISO setting (equivalency of sensitivity of a film to light), aperture, white balance, lens used parameters etc.

Those metadata can be extracted and some of them could be transformed into NSIF or NSILI metadata however, none of those STANAGs support EXIF metadata directly.

STANAG 4609 NDIMS uses industry-based container architecture to store video, audio and metadata components of a video sequence. Industry-based data formats and codecs for video encoding are utilized so STANAG 4609 defines parameters for containers, data formats, codecs and its combination. It also defines a set of metadata using key-length-value (KLV) triples to encode camera position, sensor telemetry, video quality etc. KLV metadata should be extracted when video files or streams are published into CSD.

Not all imagery data are stored in the CSD due to the variety of formats and limited support in NSILI. Imagery data are rather stored directly in the file system, especially in the case that standard JPG pictures or MPEG videos are used. Naming conventions and directory hierarchy are typically used to capture basic metadata as date-time and position.

To build data mart on the top of imagery data, metadata defined in STANAG 4545 NSIF, STANAG 4609 NDIMS and STANAG 4559 NSILI, as well as KLV and EXIF metadata, should be considered in data mart design and they should be supported in the extract phase of ETL.

As was demonstrated in the previous paragraphs, the set of image and video metadata is quite wide and fuzzy which is difficult to maintain in relational databases but Big Data can handle it more appropriately.

Another reason to utilize Big Data technologies is managing a huge volume of multimedia files and geolocation metadata processing.

3.8. ISR Products

ISR products are the output of Processing and Exploitation phases of the ISR process. ISR products include sensor data, ISR reports and other documents created and evaluated by the analyst.

There are about 20 ISR product types with a set of common metadata and ISR product-specific ones defined in the NSILI as was showed in NSILI related chapters. The common structure of ISR report or product consists of the identification of the product, name of the report, associated geographic area, date and time of validity, capturing sensor or asset identification, associated task and other detailed parameters.

There are a lot of technology standards used to store ISR products:

- ISR products format defined in AEDP-17, which should be the preferred option.
- APP11 messages stored in a file system or database.
- Intelligence disciplines specific formats, e.g. COMS-ESM or IMIMT data formats introduced in the previous chapter.
- Proprietary textual or binary formats stored as files.
- Proprietary database structures.
- Office documents (Word, Excel, PDF), and others.

It is not possible to create an overall overview of all ISR products due to diversity of used data formats. To solve this issue data mart could be built on the top of multiple ISR products storages. Building this data mart would be challenging due to the fact of a variety of data formats used, similarly as in the case of ISR tasks data mart. Tens or even hundred APP11 message types, individual disciplines data formats, office documents, AEDP-17 XML structures, NVGs, etc. are in place for the ETL phase. Due to the number of data sources and used data formats, implementation of ETL phase might include hundreds of transformation rules and its maintenance might be expensive. In addition, rigid data structures of a data mart might not be flexible enough, more flexible Big Data technologies could be utilized to build data mart and mine the data.

3.9. ISR Operational Picture

ISR Operational Picture is a specific ISR product, a contribution of the ISR community to Common Operational Picture (COP). COP is [34] a single identical display of relevant operational information (e.g. position of own troops and enemy troops, position and status of important infrastructure such as bridges, roads, etc.) shared by more than one command.

NVG is used for encoding and sharing COPs. NVG standard includes a file format for the encoding of battlespace objects into overlays and protocol for the automated exchange of the overlays. NVG is an XML based format, the elements of NVG represent individual objects and its type (point, line, arrow, etc.), label, position, symbol code and other attributes, see the example in Figure 6.

The symbol code is uniquely defining associated graphic symbol from APP-6 symbology. The code is an alphanumerical sequence (e.g. SFGPUCR---EF***) where each character's

subsequence symbolizes [35] unit identity, dimension, status, function, size of the unit and additional information. There are several encoding standards for military symbology, NATO specific, USA specific and symbology for the dismounted soldier.

```
<?xml version="1.0"?>
<nvg
  xmlns="https://tide.act.nato.int/schemas/2012/10/nvg"
  version="2.0.0">
  <point label="CZE_ISR_BN_HQ" uri="da8eefcc-
1a3f-403d-9d12-e91f60a19693" modifiers="M:53;T:ISR"
symbol="app6c:SFGPUCR---EF***" y="53.1288105381945"
x="18.00605177595984"/>
  <point label="F_EB_coy_A_ISR" uri="22c7cc78-
e3a1-4bb9-81be-f2d4b3cac037" modifiers="M:ISR;T:A"
symbol="app6c:SFGPU-----G***" y="50.20523308421297"
x="12.914633886152924"/>
</nvg>
```

Figure 6: Example of NVG data format

The operational picture represented in NVG format is intended for operational purposes, to visualize a tactical situation in overlays on the map. As its primary usage is visualization, qualitative aspects of the COP such as the size of covered area, count, status and types of units, etc., are not automatically processed or evaluated at all. Those quantitative aspects can be handled by data mart where those values could be extracted during the ETL phase. In fact, it is the task of data mining, because it is about digging hidden information from (source) data. Big Data technologies could be also for managing incoming changes of NVGs due to its streaming nature.

3.10. ISR Data Mart

As was discussed in previous paragraphs, ISR data are stored in multiple CSDs, APP11 messages and files, proprietary databases, proprietary data files, NVG files, office documents, images, videos etc. Although CSD is preferred way to store the ISR data, CSD does not cover all types of ISR products and common military data such as NVGs, NFFIs, APP11s, etc. so there is no central point to collect all ISR related data. Moreover, validation rules applied to the metadata are quite strict in the NSILI, which might conflict with wide and fuzzy nature of a sensor data. Finally yet importantly, as NSILI is focused on the interoperability mainly, it does not cover all requirements of a data analyst, especially requirements related to powerful queries, data analyses and data mining.

Building specific data marts for individual intelligence disciplines was proposed to enable operator and data analyst perform real data analyses on the ISR data. Following steps should be done to design ISR data mart: the identification of relevant data sources, the design of data model for data mart, the definition and implementation of ETL processes, the definition of analytical tasks etc. Data warehouse for all ISR data can be also built on the top of those ISR data marts but its design is out of the scope of this article.

3.11. Data Sources

The possible sources of ISR data presented in previous paragraphs include, but are not limited to:

- CSDs.

- NVG files and servers.
- APP11 messages stored in databases or file system.
- Proprietary databases/file systems for ISR task.
- Proprietary databases/file systems for reports.
- Proprietary ESM-COMMS data databases/file systems.
- Imagery files with NSIF or EXIF metadata.
- Video files with NDISM and/or KLV metadata.
- Office documents.
- CESMO network.
- NFFI network.

3.12. The Data Model for Data Marts

The design of a data model for individual data marts would be a straightforward generalization of logical data models introduced in previous chapters. Common metadata attributes should be added and those could be inspired by NSILI common metadata attributes such as date, time, geolocation, originator, type and security to list the most important ones. APP11 messages would be more challenging as there are about 400 messages types. Common header items could be generalized but individual content of the message should be analyzed with respect of corresponding intelligence discipline. Detail design of individual data marts is out of the scope of the article but some ideas are introduced in the practical part of the article.

There is only a very small subset of common metadata applicable to all individual data marts. The basic idea of its identification was introduced in the previous chapter. Generalization of common metadata applicable to a data warehouse level can be done in the next steps. The data warehouse will then consist of a very small number of common attributes (date, time, position) and a huge number of specific attributes for each database entry, there can be tens or even hundreds of attributes for sensor data. The huge number of attributes can lead to some implementation and performance issues. Columns are used in traditional databases to store attributes so there will be hundreds of mostly empty columns. It may not be effective concerning allocated disk space and memory. Big data technologies, especially column databases are designed and intended for this type of tasks.

3.13. The Extract Task

Extract tasks are to be run periodically; typical business intelligence process runs ETL during the night, which might not be sufficient in the military environment especially on the tactical level. The data extraction interval could be hours or online business intelligence should be employed. Big Data technologies, especially streaming databases are designed and intended for this type of tasks.

Data extraction details will be discussed in the next paragraphs:

CSDs

Data extraction from CSD can be implemented on the top of the AEDP-17 and AEDP-19 services. The query service specified in AEDP-17 uses Catalogue Entry Data Set, which can be used as

an input format for the extraction task. In that case, the extraction task will be universal for each standard based CSD. On the other hand, the web service interface is used so the extraction task does not have to be optimal from a performance perspective. Direct access into the underlying database should be much efficient but it will not be universal as an internal implementation of underlying CSD database differs.

The COI services specified in AEDP-19 to manipulate organization, requirement, priority, task, request and GAOI uses XML structures, which can be used as an input format for the extraction task; there are about 150 entities used. The extraction task will be universal for each standard based implementation as the underlying SPS++ service is content-agnostic. Direct database access is not an option in that case as SPS++ entities are not bound with its type and COI services must be used. On the other hand, AEDP-19 defines its own notification mechanism, which allows spreading out data extraction in the time and thus improving performance.

NSILI defines a synchronization mechanism so the server's topology must be taken into account in the extraction task. It must be specified from which CSDs are data extracted to avoid duplicities and unnecessary extraction of the data. Another approach would be to build a special instance of CSD, which will be synchronized with other CSDs as their copy in read-only mode. The extraction task will be run only on this specific instance so it can utilize direct access into the underlying database and thus provide optimal performance.

NVG

Operational Pictures are stored as NVG files on the file system or in the database and those can be read with the extraction task. The web service interface defined to publish and retrieve NVGs can be also utilized for the extraction task.

The extraction task needs to access NVGs from all the mentioned resources. There could be also built read-only NVG server, which will aggregate NVGs from different NVG sources (servers, files, databases) and notify the extraction task in the case of new or changed NVG.

APP11

An exchange mechanism is not defined for APP11 messages so they can be stored in files, databases or messaging queues. The extraction task needs to access all those storages and to handle the complexity of messages as standard defines more than 400 message types which can be stored in two different formats, XML and slash-separated file.

Proprietary databases and file systems

The standard database connector should be used to extract data from proprietary databases. The definition of the extraction task needs to be done for each individual data source by the source database definition. As the database definition is proprietary there might be legal, technology and other obstacles, which might make the extraction task more complicated or even impossible.

The extraction from the file system is a very similar task, significant difference would be that the parser of the file structure needs to be incorporated into the chain of ETL processes. Parsing of XML or CSV files is mostly trivial or similar to database

extraction but there are more complex and/or binary data formats used.

Imagery and video databases and file systems

Imagery and video files are a good example of complex binary file formats where the complex parser needs to be employed. Common ETL tools mostly do not support binary formats such as NSIF or EXIF so user-specific code needs to be implemented. It needs to be considered that parsing of binary structures might be time-consuming.

CESMO network

The concept of CESMO is based on platforms, which periodically contribute to the sensor network. All assets connected into CESMO network are real sensor systems so the connection of read-only data consumers is not supported from the operational perspective although it is technically possible. For that reason, CESMO should not be connected directly to the extraction task; rather some ESM proxy system should be used. This proxy ESM system would collect, filter and aggregate raw ESM data and store them in the ESM database, which would be a source for the extraction task.

3.14. Transform and Load Tasks

Data extracted from different data sources need to be transformed into data mart data model (the transform task) and then loaded into data mart itself (the load task). Data from multiple data marts can be also processed into the overall ISR data warehouse. Transform and load tasks are standard functionality of data marts and data warehouses, so its analyses and design are out of the scope of the article, only a few notes are mentioned.

Many parameters for those tasks, such as time of jobs execution, frequency and period of jobs, error recovery policies, executed actions etc. should be allowed to be configured by a user.

The transform task is typically covered by mapping tool, which declarative prescribes mapping of input data structures to output structures. This mapping can be defined directly in the source code or industry standard based mapping such as eXtensible Stylesheet Language (XSLT) can be used. Datamart implementations utilize a combination of those approaches and powerful GUI enabling the visual mapping of data structures is available to users.

The duration of the load task might take a lot of time as huge volumes of the data are being stored. Some optimization can be implemented on the database level as well as time-intensity could be also covered on the organizational level. For example, spreading activities during the whole day can be considered.

3.15. Visualization Tools

The final step of a building data mart is the design of the presentational layer for end-user. Most BI tools provide this kind of functionality as an integral part of the toolset. Set of dashboards, tables, graphs are typically available to the end-user. The goal is to enable to visualize underlying data in user friendly and easy to understand manner. Some practical observations and experiences with selected tools are presented in the practical part of the article.

3.16. Data Marts and the Tactical Network Limits

Limits of the tactical network must be taken into account in the case that data mart implementation and topology is designed at the military environment. The tactical network cannot be expected as reliable and it could be limited by low bandwidth, high latency, communication windows and distributed topology; mash-up networks need to be also considered.

The data mart should be implemented only on that organizational level where it makes sense. A higher level of command would be an appropriate level, while sensor level would be probably inappropriate. Anyway, carefully evaluation of a data flow is needed for the proper design of data mart topology.

3.17. The Big Data Tools to Process ISR Data

Some issues remain even in the case of using BI tools for ISR data processing; those issues cover:

- Extract phase is typically run once a day, in a given time.
- Streaming nature of sensor data.
- Wide and fuzzy data sources.
- Multimedia data processing.
- The volume of the data.
- Office documents and other unstructured data.

The concept of Big Data can be used to deal with those issues. The nature of the ISR data and identified issues of its processing satisfies already mentioned characteristic of Big Data described by 5Vs. Following technologies can be utilized to mitigate listed issues: Not Only SQL databases (NoSQL), Map-Reduce algorithm, In-memory computing and Streaming databases.

3.18. NoSQL Databases

NoSQL databases are a wide set of technologies and approaches solving or improving selected issues of relational databases. One of the primary impulses for NoSQL movement was the beginning of massive distributed databases and applications on the internet. NoSQL databases include key-value databases, document databases, column databases, and graph databases to name the most significant ones.

NoSQL databases are employed by Big Data techniques or products as an underlying layer and enabler thus they are evaluated in the following chapter of the article.

3.19. Map Reduce

Map-Reduce represents a universal way to parallelize the computation. Map-Reduce is a combination of two functions, map and reduce. Input data are split into chunks based on key-value pairs. Those chunks are distributed to mapper components, mapper component execute given task, which is written in the map function. The output of the map function is collected and grouped by the controller according to keys in the key-value pair. Then reducer functionality is executed and finally, all values are combined into the result. [36]

Development with Map-Reduce principle means split code into map, reduce functionality, and define rules for distribution input data into chunks. Coordination and synchronization of execution map and reduce tasks, data distribution and collection of results is managed by appropriate Map-Reduce framework.

Apache Hadoop is an open-source implementation of Map-Reduce principle and it is intended to process, store and analyze a huge amount of distributed and non-structured data. Although is Hadoop the key technology for the Big Data, Hadoop and the Big Data are not equal. [37]

Hadoop ecosystem consists of more connected projects including Hadoop Distributed File System (HDFS), MapReduce framework, NoSQL storages Cassandra and HBase and other components. HDFS [38] is a distributed file system intended to utilize common (utility) hardware so it was designed to be fault-tolerant to HW and network faults. MapReduce framework is a computational layer enabling parallel execution of computation in each node. Results of the computation are aggregated and stored in the storage layer (HDFS). All those components form an ecosystem enabling vertical scaling; each layer is implemented in a specialized product, which can be easily replaced.

3.20. In-Memory Databases

In-memory databases offers very fast processing of data, over 100.000 operations per second on a single machine. The volume of processed data is limited by the size of RAM, another downside is that data can get lost between subsequent disk flushes (e. g. due to server crashes or by losing power supply). Due to those facts, in-memory databases are recommended to be used on the strategic or operational level, not on the tactical level, thus In-memory databases are out of the scope of the article [39].

3.21. Streaming Databases

Streaming databases are intended to process the data which are streamed continuously without the option of being available for long periods; so the query processor can only respond to queries that have been previously registered with it and can produce results for the data stream by looking at a portion of the stream available within a certain window. Another use case for streaming databases is a situation when the query is executed in parallel with incoming data processing so returned data does not contain data just received. [40]

An alternative to the streaming database could be streaming platform such as Apache Kafka. The streaming platform allows publish and subscribe to streams of records, similar to a message queue or enterprise messaging system, store streams of records in a fault-tolerant durable way and process streams of records as they occur. [41]

Some ISR data have streaming nature, especially COMMS-ESM data, TDL etc. Even so, streaming databases are out of the scope of this article.

4. Building Multiple ISR Data Marts

Selected ISR data and its main characteristic, BI principles and Big Data technologies were introduced in the previous chapters. In this chapter will be proposed how to utilize BI approach to build the ISR data mart. The universal ISR data warehouse will not be built, only specific data marts for each individual area with a focus on EW data, ISR products, ISR operational pictures and operational diary as an overview of ISR tasks.

Data marts were built with the utilisation of RapidMiner Studio, an open-source software platform for analytics teams. The

tool is used for data preparation, machine learning, and predictive model deployment. To achieve that functionality, the tool contains about 400 operators for data manipulation enabling load data from multiple sources, transform and process them and visualize the resulting output.

4.1. The Data Mart for Electronic Warfare Data

The physical nature of radio signal propagation and other factors such as reflections, terrain conductivity, time of day, weather, etc. make the processing of electronic warfare data difficult and challenging. Another challenge is that radio transmission or broadcasting on fixed frequency or band is not widely used as it is easy to monitor, jam and deception by the enemy so sophisticated methods such as frequency hopping, direct sequence, chirp, time-hopping etc. are used instead. Frequency hopping will be considered in the following experiments.

Frequency hopping is a method of transmitting radio signals by rapidly switching a carrier frequency among many frequencies using a pseudorandom sequence known to emitter and receiver only. In practice frequency hopping means separation of one (logical) radio transmission into a huge amount of very short (2-300 milliseconds) physical transmissions on different frequencies. The activity detector does not know parameters of the pseudorandom sequence so it is not possible to determine to which transmission the detected activity belongs, if any, or if it is burst, interference or disturbance. If the emitter, detector or both are moving, the detected frequency can be changed due to the Doppler Effect. [42]

Due to aspects of the real environment, an activity detector or direction finder will produce a lot of data including bursts, disturbances, interferences, reflected signals, fake signals etc. Captured data may look randomly on the first view so further analysis of data is needed. Physical aspects need to be taken into account for these analyses so an experienced operator is needed. Alternative methods based on data mart implementation will be evaluated to reduce the need for operator skills, one example could be employing histogram to analyse DFs [43].

Table 4: The sample of generated data.

Frequency	Quality	Level	Detection Start	Duration
64566928	80	88	2020-06-13 13:34:03.642	10
65449040	90	91	2020-06-13 13:34:03.652	10
66397251	3	92	2020-06-13 13:34:03.662	10
67477144	93	5	2020-06-13 13:34:03.672	10
64161124	83	92	2020-06-13 13:34:03.682	10

Following part of the experiment focus on analyses of detected activities. The first step of the building data mart is to define the data model as different sensors use different sets of attributes. The proposed data model for detected activities consists of (carrier) frequency, quality (of detected signal), level (of detected signal), date and time of the detection and duration of the transmission. A random set of activities was generated in a proposed logical data format, the sample of generated data is shown in Table 4.

The generated set of activities simulates multiple frequency hopping transmissions on different bands including bursts and

disturbances. Multiple communication scenarios were tested, in the presented example are two frequency hopping broadcasts within one minute frame; one broadcast is in band 63-68 MHz with duration 10 milliseconds of one physical activity and the second one is in the band 66-73 MHz with duration 200 milliseconds of one physical activity.

Both transmissions partially overlap and random disturbances, noise and bursts are added in the band. The different activities/bursts ratio in the sample data were simulated, the number of activities, bursts and its ratio is shown in Table 5. The volume of generated data was limited to about 5000 to 10 000 thousand rows, which is the limit of a free version of RapidMiner Studio.

Those sets were loaded into RapidMiner Studio to be analyzed and visualized. Basic statistical analysis such as average of individual parameters (frequency, duration, quality) are not useful at all but histogram can be used for a basic overview of data distribution for quick identification of two bands with the most significant traffic as demonstrated in Figure 7.

Table 5: Activities/bursts ratio for a different set of data

Activities	Bursts	Sum	Burst/Activity Ratio
3170	32	3202	1,05%
3185	162	3347	5,1%
3148	320	3468	10,2%
3169	1578	4747	49,8%
3163	3243	6406	102,5%

On the other hand, as the number of bursts increases, activities/bursts ratio also increases and as a result, the histogram became much less clear as demonstrated in Figure 8 for 50% activities/bursts ratio.

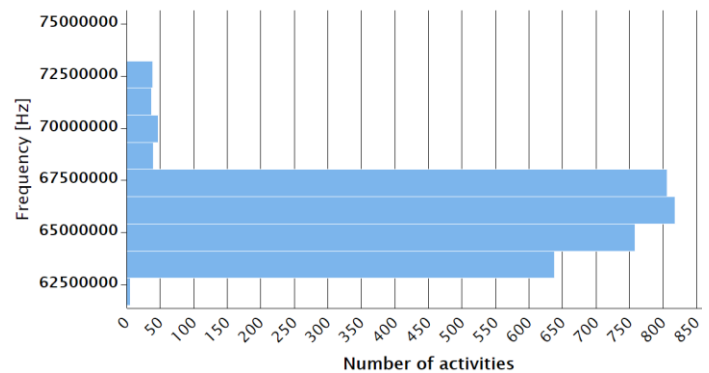


Figure 7: Frequency histogram for sample data with 1% activities/bursts ratio

Distribution of frequencies in the time can be seen on the scatter diagram, examples are shown in Figure 9, 10, and 11 for different activities/bursts ratio. The colour represents the duration of the activity, which can help to analyse data visually.

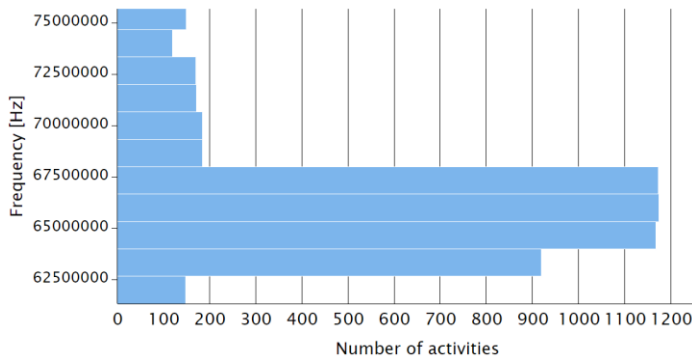


Figure 8 Frequency histogram for sample data with 50% activities/bursts ratio

The data with activities/bursts ratio higher than 10% are quite messy, so it must be filtered in the tool. All activities with quality or level lower than 70 are cut off, the result is in Figure 12 and the resulting picture is much clear to see the relevant activities.

Note: Despite all the data were generated randomly, the output seems to be quite regular and it will be expected that real data are less regular but real measurement data are classified. Similar kind of data analyses and visualization could be done for DF data where azimuth will be used as another dimension for visualization.

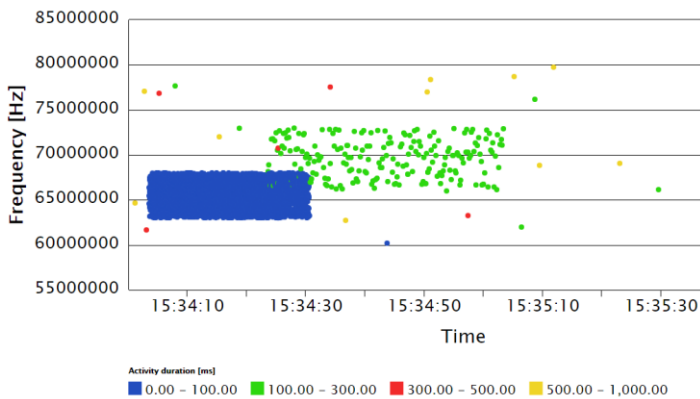


Figure 9: Scatter diagram for the original data with 1% activities/bursts ratio

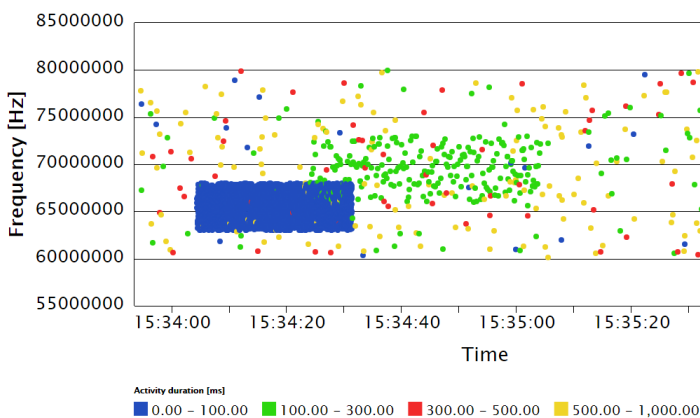


Figure 10: Scatter diagram for the original data with 10% activities/bursts ratio

The main benefit of the presented approach is its easy ability of the graphical presentation and visualization. Although the basic graphical overview of the radio spectrum is an integral part of a

sensor controlling SW, its main purpose is to control the sensor, not to provide advanced data analyses.

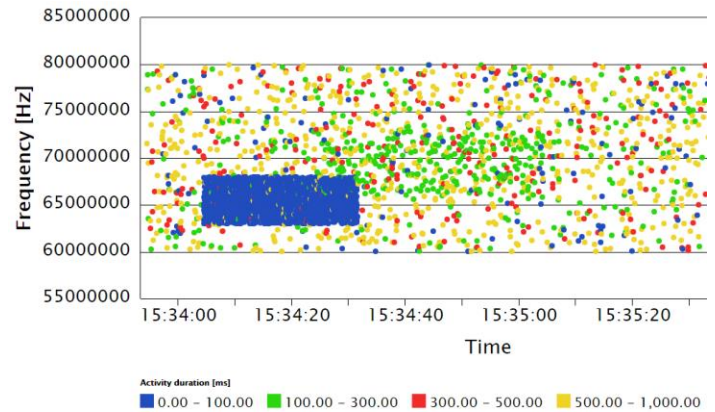


Figure 11: Scatter diagram for the original data with 50% activities/bursts ratio

Moreover, sensor controlling SW is typically tightly coupled with the sensor's data model (and capabilities) so it is hard, or even impossible, to utilize it for processing and visualization of data from different sensors.

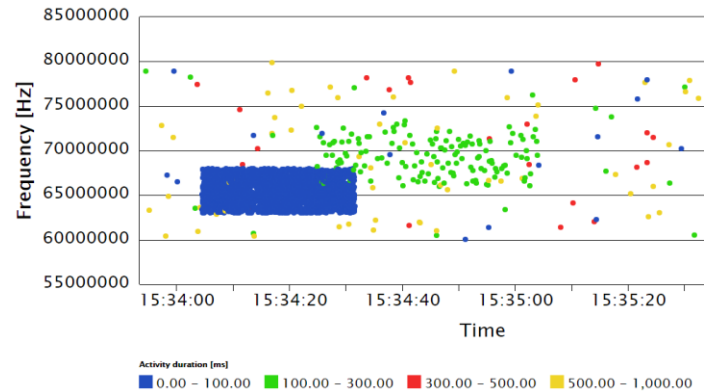


Figure 12: Scatter diagram for the filtered data with 50% activities/bursts ratio

Mentioned disadvantages could be eliminated by the implementation of the data mart and utilisation of its visualisation tools. Those tools enable to load the data from different sensors, to transform them into a unified data model, to clean and filter them and finally to visualize the results. It allows performing much complex data analysis then any sensor controlling SW can with lower effort. Visualisation tools are highly and easily configurable so a different view of the data is easy to achieve. On the other hand, all analyses must be done concerning the physical nature of detected activities and radio waves propagation as well as knowledge of the operational and tactical situation.

4.2. ISR Products Statistics

The common set of metadata for ISR products defined by NSILI includes: ISR product status, publisher, source library, date and time of modification, security (classification, release-ability), ISR product type, mission identifier, language, description, related geographical area (country code or geographic reference box) etc. There was generated a random set of metadata for ISR products in the experiment. Most of the generated ISR products were imagery products, the rest were random products. All mandatory attributes were generated, optional attributes were set optionally. Values of some attributes were set with typos to simulate real data.

The generated set of metadata was load into RapidMiner Studio and processed. In this case, unlike the previous experiment, basic statistics such as the total number of products and the number of different product types could be beneficial, see Figure 13. Note: Number of documents is in logarithmic scale in figure.

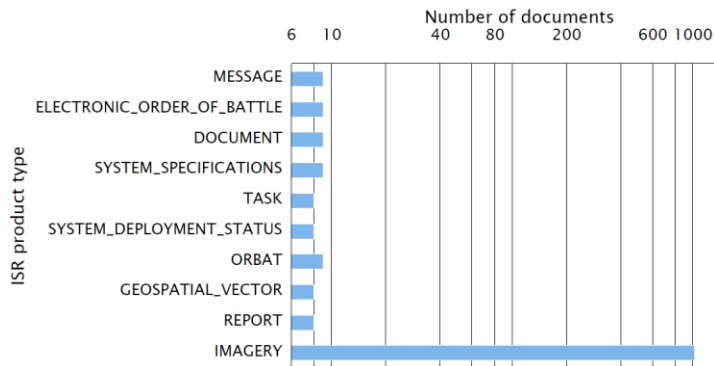


Figure 13: ISR Products types and its number

More advanced statistics were also evaluated, ISR products were grouped by Language and Type, the output is shown in Table 6. It can be easily seen that many imagery products have not set Language attribute, which could be an indicator of the low quality of input data. This observation is an example of how this kind of analyses could easily identify issues in the data and help to improve the process of data collection and processing.

Table 6: The example of grouping ISR products by Language and Type

Language	Type	Number of records
CZE	IMAGERY	511
null	IMAGERY	414
FRA	IMAGERY	37
ITA	IMAGERY	31
ENG	IMAGERY	31
CZE	EOB	7
CZE	DOCUMENT	7
CZE	MESSAGE	6
CZE	REPORT	5
CZE	GEOSPATIAL_VECTOR	5

Queueing three processes generated the report shown in Table 6: data load, aggregation and sort. The chain of data processes is shown in Figure 14, which demonstrate how is data processing managed within RapidMiner Studio.

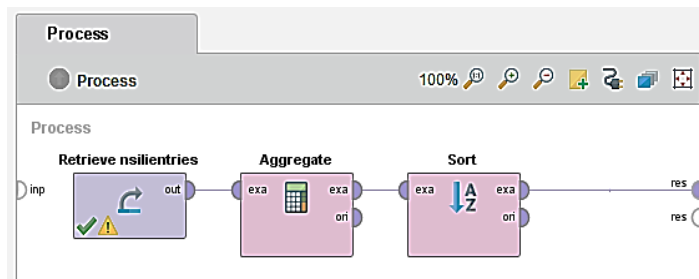


Figure 14: The example of processes design

The data mart built on top of ISR products can be used to generate ISR products statistics. Those statistics and reports to improve collecting tasks and their data quality. The missing data or incorrect values are easy to identify in reports such as in Table 6. The RapidMiner Studio allows drilling down the data to identify the original source of incorrect data. In contrast, NSILI defines the search functionality via its query language only and there is no standard way to see such qualitative and quantitative criteria.

Another benefit would be the improvement of collecting process based on the statistics. The most of ISR products in the shown example were originated from imagery intelligence, based on this simple statistic can commander request data from another intelligence discipline to verify or refute the information by another source.

4.3. The Operational Picture

The operational picture can contain tens or hundreds of battlespace objects such as friendly, enemy and neutral units. In this case, the data model of the data mart would be based on Position, Unit identity, Dimension, Status, Function, Size of the unit, additional modifiers and other attributes. The proposed set of attributes was created as a generalization of the data models of individual military symbology standards. Operational pictures can be load and transform into the data mart, only challenging operation during this step is parsing of military symbol codes. Following analyses, presentation and visualization would be based on the same or similar principles as already presented.

The benefit of BI approach in this scenario is a display of quantitative indicators derived from COP. Those quantitative indicators are not displayed in the COP graphical representation, so their explicit calculation brings new information to the operator. The issue here is that especially the enemy's situation is based on reconnaissance output so it might not correspond to reality.

4.4. The Operational Diary

The operational diary keeps track all activities of the unit and should contain all tasks, requests, and relevant reports. As a side effect, it enables to verify that all RFIs were answered in the given time, the reliability of the sensor can be tracked as well as a response time of the subordinate units, their efficiency etc.

The benefit of BI approach in this scenario is the ability to monitor the performance of units based on automatically tracked tasks and reports. The data model of a data mart would use attributes such as the task and report status, the date and time of creation, the expected time of reporting, the real-time of reporting, collecting platform etc.

The real tactical situation must be taken into account when the operational diary is evaluated. No relevant information might be tracked from tactical reasons or due to the communication infrastructure limitations and issues.

5. Big Data to Process ISR Data

Selected Big Data technologies were evaluated for ISR data processing, especially document databases for metadata storage and Hadoop for ISR data synchronization.

5.1. Document Database as NSILI Metadata Storage

Document databases are an evolution of simple key-value stores, which use the key-value tuples. The key is an identifier of record and the value is content of record that can be both structured

and unstructured. One logical entity can be stored multiple times, each one is optimized for a different type of a query which is very efficient for queries but leads to a rise of the volume of space needed. [44]

Document databases extend this approach with the possibility of structured document storage. Another advantage of a document database is its schema-less nature, which means that structure of stored documents is not validated on the database level so different data structures can be stored within one table.

The experimental CSD implementation with the document database as an underlying layer for AEDP-17 Publish/Query services and AEDP-19 SPS++ services was evaluated and discussed in [45]. The conclusion was that document database usage raised significantly flexibility of the solution as it enables to store any relevant document independently of its content and version. The implementation time and workload were also dramatically lower.

On the other hand, the disadvantage of NoSQL technologies is the fragmentation of its implementation where each particular product is specialized to its specific purpose and/or use case. There is also a lack of standardization so it is very difficult or even impossible to migrate from one NoSQL solution to another.

5.2. Hadoop for ISR Data Synchronization

CSD utilizes two technologies to share and disseminate ISR products and its metadata, CORBA and web service, both of them have some disadvantages.

CORBA as a technology is already obsolete, it is not supported in modern programming languages and it would become obsolete in the future edition of NSILI. Moreover, CORBA protocol is binary so security mechanism based on XML security labelling is not applicable. Real experiences show that CORBA implementation of data synchronization is slow as CORBA is an implementation of Object Request Broker with a lot of overhead functionality from the perspective of CSD.

Web services, in contrast, solve those issues but NSILI does not define all functionality, especially synchronization features are not defined. It is not a problem of web service technology rather of NSILI standard itself.

Although synchronization via Web services is not standardized, it is possible to utilize Web services for synchronization CSDs. Synchronization based on Web services must handle topology of CSDs network, data duplication etc. The use of Hadoop to manage this task is proposed. The basic idea is to export data via Query service to Hadoop File System, then to distribute it to different nodes and finally import the data into CSD via Publish service. Batch export and import need to be implemented which is an implementation detail.

Import or Publish services should handle duplicity. Circle synchronisation is forbidden in NSILI and according to the standard, it is a configuration issue. CSDs synchronisation is a directional point to point and the administrator is responsible for proper configuration and avoiding the circle dependencies. The same rule needs to be satisfied in the case of proposed Hadoop based synchronisation.

Hadoop would be used in the way, that all CSDs has the same data or the same subset of data exported by individual CSDs. One

CSD would export for example only IMINT data and do not export EW data, so all CSD in the network will access IMINT data. The data would be separated to those, which should be synchronised with all participants of the CSD network and those data that should not be synchronised. The set of products can be easily configured by the query definition.

The advantage of Hadoop utilisation would be a much easier configuration of synchronisation network and synchronised products. The map-reduce mechanism is not used in that case, only advantages of a distributed file system are utilised. The synchronisation implemented in this way would be resistant to network or hardware failures which is provided by Hadoop itself. This type of the synchronisation over distributed file system would be applicable for other data distribution as well, including EW data etc. but limits of the tactical network need to be taken into account.

5.3. Map-Reduce for ISR Data

Processing of the ISR data can be seen as Map-Reduce task as data collected on sensor level are reduced and distributed to a higher level on command, where are combined with data from other sources, reduced and so on. Although the idea is similar, human interaction is expected, so the process is not fully automated to be understood as Map-Reduce task.

Based on experience, distributed computing is not ideal for the tactical network. The requirement is that systems should work even in the case that connection is lost, some nodes are temporary or permanently not available etc. Distributed computing and especially MapReduce is a specific approach to computation and only selected tasks can benefit from it, especially those which can be easily parallelised. The infrastructure for parallelisation and data distribution is needed which is not a case on the tactical level. Nevertheless, future research of utilisation MapReduce principles is recommended. Big Data principles are applicable for processing the data of individual intelligence disciplines while Map-Reduce principle seems to be not so beneficial on a tactical level due to infrastructure reasons.

6. Conclusion

The ISR process, principles of BI and Big Data technologies were introduced in this paper. Areas in ISR environment where BI and Big Data approach can be beneficial were identified and a way to build ISR data mart was proposed. Experimental data marts were built and loaded by data preparation (data from practice is not possible to get or it is not available due to security reason), but the possibility of data processing, analyses and presentation remain without problems. The utilisation of the document database for ISR products library implementation and Apache Hadoop for its synchronisation was proposed.

Some other promising Big Data technologies were identified and proposed for evaluation in future works. Those technologies include streaming databases and streaming platforms to process sensor data and MapReduce principles.

Proposed utilisation of Big Data technologies is also applicable to other intelligence disciplines, especially OSINT which is based on research on open sources such as newspapers, internet pages, television broadcast etc. would benefit of all aspects of Big Data technologies. Similarly, IMINT analyses would benefit from presented technologies. In any case, the limits of tactical networks must be taken into account when implementing Big Data technologies.

There were demonstrated an advanced BI business analyses that are not used in ISR systems; it is quite original solution. It was also shown that all proposed analyzes have to take into account the real data from military operations, exercises and missions. Regarding this result is awaiting that the domain experts, operators and the data analyst in military environment have sufficient knowledge not only in monitoring operations, but they are able also prepare appropriate results using BI, AI and other relevant technologies.

Acknowledgement

Mr. Novak was participating of development Integrated SoftWare Module for Command, Control, Communications, Computers, Intelligence, Surveillance, Targeting Acquisition and Reconnaissance (ISWM C4ISTAR) project in 2014-2016 focusing on ICT (Information and Communication Technology) support of JISR (Joint ISR) cycle and interoperability of ISR products through STANAG 4559 NSILI compatible implementation.

Mr. Burita presents the results of the research as a part of the project [46] at the Faculty of Military Technologies of the University of Defence, Department of Informatics, and Cyber Operations.

References

- [1] A. Novak, L. Burita, "Business Intelligence and ISR Data Processing" in Proceedings of the International Conference on Military Technologies 2019 - (ICMT). Brno, Czech Republic: University of Defence, 1-8, 2019.
- [2] OODA Loop [Online]. [Cit. 2020-03-12]. Available at <https://www.toolshero.com/decision-making/ooda-loop/>
- [3] NATO Standardization Office, AJP-2 Allied joint doctrine for intelligence, counter-intelligence and security, 2016.
- [4] NATO Standardization Office, AJP-2.7 Allied joint doctrine for intelligence, surveillance and reconnaissance, Edition A, version 1, 2016.
- [5] The Global Airborne Intelligence and C2 Community's Annual General Meeting, The Hurlingham Club, London, UK, 2020 [Online]. [Cit. 2020-03-26]. <https://www.defenceiq.com/events-airborneisr/>
- [6] Business Intelligence [Online]. [Cit. 2020-03-15]. Available at <https://www.123rf.com/visual/search/9342840>
- [7] Data Warehousing Solution [Online]. [Cit. 2020-03-16]. Available at <http://www.aspiretss.com/services/data-warehousing-solution>
- [8] OLAP Cubes in SQL Server [Online]. [Cit. 2020-03-17]. Available at <https://www.sqlshack.com/olap-cubes-in-sql-server/>
- [9] CRISP-DM methodology phases [Online]. [Cit. 2020-03-18]. Available at https://www.researchgate.net/figure/CRISP-DM-methodology-phases-Adaptation-of-Chapman-P-3_fig1_318309402
- [10] SAS Analytics SW and solutions [Online]. [Cit. 2020-03-20]. Available at https://www.sas.com/en_us/home.html
- [11] IBM SPSS software [Online]. [Cit. 2020-03-22]. Available at <https://www.ibm.com/analytics/spss-statistics-software>
- [12] F. L. F. Almeida, "Benefits, Challenges and Tools of Big Data Management", Journal of Systems Integration, 4(8), 12-20, 2017, DOI: 10.20470/jsi.v8i4.311.
- [13] International conference Big Data for Defence, June 23-25, 2020, London, UK.
- [14] M.K. Hernandez, C. Howard, R. Livingood, and C. Calongne, "Applications of decision tree analytics on semi-structured north Atlantic tropical cyclone forecasts", International Journal of Sociotechnology and Knowledge Development, 11(2), 31-531, 2019. DOI: 10.4018/IJSKD.2019040103.
- [15] B. Essendorfer, A. Hoffmann, J. Sander, and A. Kuwertz, "Integrating coalition shared data in a system architecture for high level information management", in Proceedings of The International Society for Optical Engineering (SPIE): Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies II, Berlin, Germany, 10802, 2018. DOI: 10.1117/12.2501861.
- [16] A. Kuwertz, D. Mühlenberg, J. Sander, and W. Müller, "Applying knowledge-based reasoning for information fusion in intelligence, surveillance, and reconnaissance", Lecture Notes in Electrical Engineering, 501, 119-139, 2018. DOI: 10.1007/978-3-319-90509-9_7.
- [17] Z. Kira, A.R. Wagner, C. Kennedy, J. Zutty, and G. Tuell, "STAC: A comprehensive sensor fusion model for scene characterization", in Proceedings of SPIE, Multisensory, Multisource Information Fusion: Architectures, Algorithms, and Applications, 9498, 949804, 2015. DOI: 10.1117/12.2178494.
- [18] C. Bowman, G. Haith, A. Steinberg, C. Morefield, and M. Morefield, "Affordable non-traditional source data mining for context assessment to improve distributed fusion system robustness", in Proceedings of SPIE, Multisensory, Multisource Information Fusion: Architectures, Algorithms, and Applications, 8756, 875603, 2013. DOI: 10.1117/12.2018386.
- [19] G.E. Home, E.A. Bjorkman, and T. Colton, "Putting agent based modeling to work: Results of the 4th international project albert workshop", in Proceedings of SPIE, 4716, 99-107, 2002. DOI: 10.1117/12.474904.
- [20] M. Barnell, C. Raymond, C. Capraro, and D. Isereau, "Agile Condor: A scalable high performance embedded computing architecture", in 2015 IEEE High Performance Extreme Computing Conference, Waltham, United States, 2015. DOI: 10.1109/HPEC.2015.7322447.
- [21] S. Rogers, J. Culbertson, M. Oxley, H. Clouse, B. Abayowa, J. Patrick, E. Blasch, and J. Trumpfeller, "The QuEST for multi-sensor big data ISR situation understanding", in Proceedings of SPIE: Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR VII, Baltimore, United States, 9831, 2016. DOI: 10.1117/12.229722.
- [22] J. Rajamäki, J. Simola, "How to apply privacy by design in OSINT and Big Data analytics", in 18th European Conference on Information Warfare and Security (ECCWS), Coimbra, Portugal, 2019, 364-371.
- [23] S. Das, L. Jain, and A. Das, "Deep Learning for Military Image Captioning", in 21st International Conference on Information Fusion, Cambridge, United Kingdom, 8455321, 2165-2171, 2018. DOI: 10.23919/ICIF.2018.8455321.
- [24] S.P. Wang, W. Kelly, "IT Professional Conference: Challenges in Information Systems Governance, Gaithersburg, United States, 70293032014, 2014. DOI: 10.1109/ITPRO.2014.7029303.
- [25] NATO Standardization Office (NSO), AEDP-17: NATO Standard ISR Library Interface, Edition A Version 1, Brussels, 2016.
- [26] APP-11 and ADatP-3. What is APP-11 (NATO Message Catalogue) and ADatP-3 (Allied Data Publication 3)? And what is the difference between the two? [Online]. [Cit. 2019-04-14]. Available from: <https://systematic.com/defence/capabilities/c2/interoperability/app-11-and-adatp-3/>
- [27] ADatP-4733 NATO Vector Graphics Specification, NATO Standardization Office, 2017.
- [28] Friendly Force Tracking, US-Army Stand-To [Online]. [Cit. 2019-04-16]. Available from: https://www.army.mil/standto/archive_2016-02-24
- [29] AEDP-02 NATO Intelligence, Surveillance, and Reconnaissance (ISR) Interoperability Architecture (NIIA), Edition 1, Volume 1: Architecture Description.
- [30] NATO Standardization Office (NSO), AEDP-19: NATO Standard ISR Workflow Architecture, Edition A Version 1, Brussels, 2018.
- [31] NATO AJP-2.7 ALLIED JOINT DOCTRINE FOR JOINT INTELLIGENCE, SURVEILLANCE AND RECONNAISSANCE. [Online]. [Cit. 2020-06-01]. Available from: <https://standards.globalspec.com/std/10031247/AJP-2.7>
- [32] Information Requirements Management and Collection Management. [Online]. [Cit. 2020-06-01]. Available from: https://moodle.unob.cz/pluginfile.php/46438/mod_resource/content/1/T3-1%20Information%20Requirements%20Management%20and%20Collection%20Management%20%28lecture%29.pdf
- [33] Intelligence Studies: Types of Intelligence Collection. [Online]. [Cit. 2020-06-01]. Available from: <https://usnwc.libguides.com/c.php?g=494120&p=3381426>
- [34] Common Operational Picture Defined. [Online]. [Cit. 2020-06-01]. Available from: <https://www.coolfiresolutions.com/blog/common-operational-picture-defined/>
- [35] NATO joint military symbology APP-6(C), NATO Standardization Office, 2011.
- [36] What is Map Reduce? [Online]. [Cit. 2020-05-16]. Available from: <https://www.ibm.com/analytics/hadoop/mapreduce>
- [37] What is Apache Hadoop? [Online]. [Cit. 2020-05-16]. Available from: <https://www.ibm.com/analytics/hadoop>
- [38] HDFS Architecture Guide. [Online]. [Cit. 2020-05-16]. Available from: https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html
- [39] In-Memory Database for Time and Mission-Critical Applications. [Online]. [Cit. 2020-05-20]. Available from: https://raima.com/in-memory-dbms-2/?gclid=CjwKCAjwh472BRAGEiwAvHVGm2FqJlqdfL5ljAyoDD7JL8jE_FJu_flf1KJNjN2H90vWnma3qKwtuxoCgVoQAvD_BwE
- [40] What Is a Streaming Database? [Online]. [Cit. 2020-05-20]. Available from: <https://hazelcast.com/glossary/streaming-database/>
- [41] Apache Kafka® is a distributed streaming platform. [Online]. [Cit. 2020-06-01]. Available from: <https://kafka.apache.org/intro>
- [42] Frequency Hopping. [Online]. [Cit. 2020-05-22]. Available from: <https://www.sciencedirect.com/topics/physics-and-astronomy/frequency-hopping>

- [43] „Communication tools PR4G, Retia Company“, Proceedings of the C2 systems in Ministry of Defence Czech Republic, Brno, 2002.
- [44] What is a Document Database? [Online]. [Cit. 2020-06-01]. Available from: <https://www.mongodb.com/document-databases>
- [45] A. Novak, “The Comparison of STANAG 4559 NSILI Compliant Tactical Intelligence Database Implementation in Relational Database and NoSQL Database” in 2018 New Trends in Signal Processing Conference, Demanovska Dolina, Slovakia, 2018, 1-7, DOI: 10.23919/NTSP.2018.8524063.
- [46] ZRO209, Project of departmental research: Development of systems C4I and cyber security, Brno, University of Defence, 2016-2020.