[PO11B18] Analysis of design requirements for the construction of an Open Source Intelligence (OSINT)-Imagery Intelligence(IMINT) integrated database

Yonhong Jeong *, Gayeon Ha
KINAC (Korea Institute of Nuclear Nonproliferation and Control)
* Corresponding author: jyh1404@kinac.re.kr

Introduction

- Since the 2000s, the Democratic People's Republic of Korea (hereafter, North Korea) has advanced its nuclear capabilities through multiple nuclear tests and test launches of long- and medium-range missiles. However, due to the closed nature of its regime, access to sites is limited, and there is insufficient official information, making it difficult to accurately assess the actual state of North Korea's nuclear and missile activities and the level of threat they pose.
- To address these limitations, monitoring approaches combining open-source intelligence (OSINT) and commercial satellite imagery are gaining traction. OSINT is a term that encompasses various forms of publicly available information, including news articles, academic reports, social media posts, and statements from governments and international organizations. Satellite imagery offers on-site evidence, including alterations to structures and land use patterns, as well as vehicle movement data. This facilitates the discernment of nuanced circumstances that textual information alone is unable to reveal.
- This paper presents a comprehensive plan for constructing a big data database (DB) covering North Korea's nuclear tests, missile tests, and security-related activities since the 2000s. The proposed plan encompasses more than merely the collection of OSINT and satellite imagery; it involves the design and implementation of an integrated database. This database is intended to support systematic collection, analysis, and visualization, as well as to provide a platform for the early detection of potential signs. Furthermore, it undertakes a comparative analysis of analogous cases both domestically and internationally, with the aim of elucidating the fundamental considerations and ramifications for the design and operational phases.

Cases of OSINT Utilization

- International Atomic Energy Agency (IAEA): Following North Korea's expulsion of IAEA inspectors in 2009, the IAEA has engaged in the active utilization of open-source intelligence (OSINT) to monitor North Korea's nuclear activities in the absence of on-site access. Consequently, the IAEA has identified indications of operational activity at North Korean nuclear facilities and has incorporated these findings into its annual reports and other documentation.
- United Nations Security Council (UNSC): The United Nations Security Council Panel of Experts on North Korea Sanctions is a specialized body that provides expert advice and analysis to the Security Council on issues related to North Korea sanctions. The Panel of Experts, which is not associated with the United Nations, is responsible for monitoring North Korea's violations of sanctions. This panel has employed public information sources in a proactive manner. The Panel conducts an annual analysis of the illegal diversion of trade or flows of sanctioned goods and issues reports that detail the findings.
- Stimpson Center: The 38 North program at the U.S.-based Stimpson Center is a leading provider of North Korea analysis using commercial satellite imagery. The monitoring of North Korean nuclear and missile facilities, as well as hydroelectric power plants, is a task that is carried out by satellite imagery experts on a consistent basis. 38 North operates a series of public analyses using unclassified satellite images and data in partnership with the U.S. National Geospatial-Intelligence Agency (NGA).
- Center for Strategic and International Studies (CSIS): The Beyond Parallel project at the CSIS utilizes commercial satellite imagery and data visualization to expose North Korea's clandestine military operations and external dynamics. The team has corroborated intelligence agency findings through a meticulous cross-referencing of satellite imagery with GIS mapping techniques, thereby substantiating information previously exclusive to intelligence agencies.
- Royal United Services Institute (RUSI): The RUSI is a prominent institution that focuses on the study of military and strategic affairs. RUSI conducts pioneering research on North Korea's sanctions evasion networks through dedicated OSINT teams, such as Project Sandstone. The identification of illicit activities is facilitated by the cross-analysis of satellite imagery, maritime vessel location data (AIS), and trade company registration records.
- Center for the Advanced Study of the Global Security Landscape (C4ADS): C4ADS, headquartered in the United States, is a prominent think tank that utilizes big data-based OSINT techniques to monitor North Korea's illicit networks. C4ADS conducts analytical activities with the objective of exposing North Korea's maritime smuggling, money laundering, and sanctions evasion tactics. C4ADS employs OSINT, encompassing network analysis and machine learning, to map North Korea's clandestine economic networks (e.g., front companies, overseas labor dispatch, and illicit trade markets). This information is disseminated to UN member states and the private sector to facilitate sanctions enforcement.
- The common features observed in these cases include the integration of various data (satellite, text, geographic information), the provision of visualization and search functions, and the supplementation of areas lacking official information with publicly available data. Specifically, the integration of facility coordinates and satellite images through Geographic Information System (GIS) technology, in conjunction with the facilitation of time-series analysis through an event database, has been demonstrated to be highly advantageous in practical applications.
- Center for Nonproliferation Studies (CNS): The Center for Nonproliferation Studies (CNS) team at the Middlebury Institute of International Studies functions as a private "information detective agency" that is not affiliated with any government. The team utilizes publicly available sources to track North Korea's nuclear and missile programs. A geographic information analysis has been employed to compare terrain features in videos with existing photographs, and past data archives have been utilized to verify these changes. Furthermore, the company utilizes video forensic technology to meticulously verify propaganda videos and photos disseminated by North Korea.

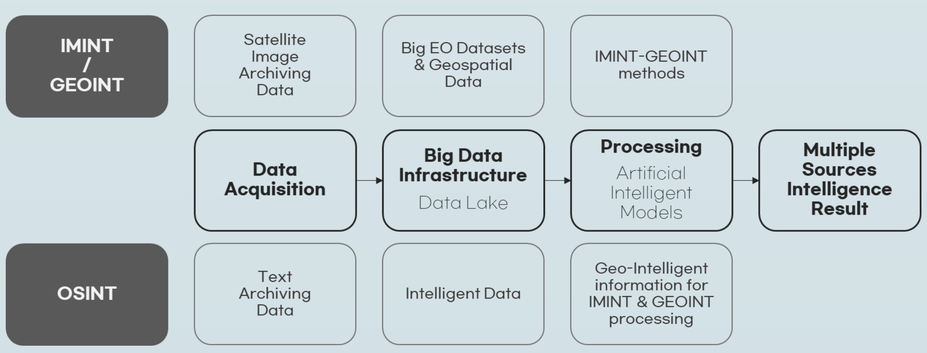

Figure 1. Data Architecture based on IMINT(or GEOINT) that integrates OSINT

Establishment of a big data platform for nuclear activities

- The Korea Institute of Nuclear Nonproliferation and Control (KINAC) is conducting research to develop an integrated big data platform capable of efficiently registering, managing, and searching big data on nuclear activities in neighboring countries, as well as performing AI-based analysis (inference, situation awareness, etc.).
- Conventional approaches to analyzing nuclear activities in neighboring countries depend on the expertise of a select group of specialists. The institute's objective is to establish a big data platform that can assist policymakers by leveraging the expertise of subject matter specialists.
- A database of satellite image data will be constructed, with the same temporal and spatial resolutions employed in the original database. This database will then be combined with a database of publicly available information to establish a comprehensive database for characterizing nuclear activities.
- To address the challenges associated with accessing information related to North Korea, a novel information collection system is under development. This system employs self-expanding crawling technology to expand the relevant knowledge base using a limited number of keywords. Additionally, multilingual-based crawling technology is utilized to overcome source bias and perform comparative analysis of detection trends by source.

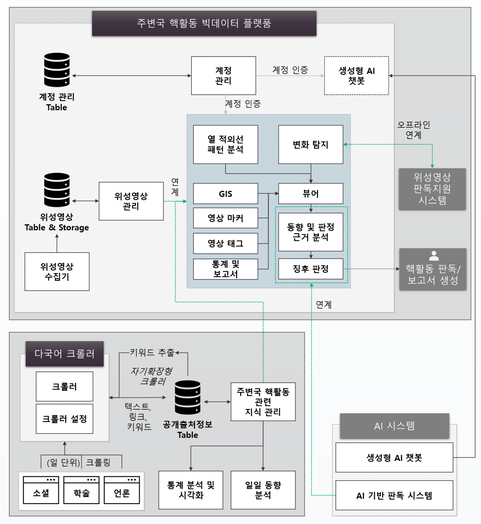

Figure 2. Big Data Platform Design (Draft)

Key Requirements for Database Structure Design

(Integration of Diverse Data Types)

- It is imperative that a unified system be implemented to manage textual OSINT (news, social media, reports), satellite imagery (optical, SAR), and spatiotemporal data (coordinates, capture dates). For instance, satellite images of a nuclear facility, relevant news articles, and the facility's coordinates/history should all be accessible in one place. The International Atomic Energy Agency (IAEA) employs a similar methodology, integrating open-source intelligence (OSINT) and satellite data to assess facility status.

(Efficient Data Storage and Large-Volume Handling)

- Given the typically voluminous nature of satellite imagery files, the adoption of a hybrid structure is strongly recommended. This approach entails the storage of raw images in file systems or cloud services, while metadata is recorded in the database. In light of the potentially voluminous nature of text-based OSINT, the consideration of horizontal scalability (e.g., NoSQL) and robust backup strategies is imperative.

(Fast Search and Querying)

- The system is expected to support complex queries, including keyword searches such as "nuclear test," "ICBM," and "Yongbyon," time-range queries, and geographic filters. This necessitates the implementation of full-text search for textual data, spatial image indexing, and time indexing for chronological queries. As demonstrated by CSIS Beyond Parallel, user interfaces that retrieve data by date or event category are essential.

(Spatiotemporal Data Management)

- Given the inherent correlation between nuclear and missile activities and specific locations and temporal periods, the utilization of a geospatial database, such as PostGIS, is imperative for conducting sophisticated spatial queries and monitoring temporal variations in movement. As with 38 North's digital atlas, overlaying satellite imagery onto maps is crucial for the visualization of facility changes and related activity.

(Data Reliability Assessment and Metadata)

- The reliability of OSINT findings can vary significantly, necessitating the assignment of a reliability score (e.g., utilizing an Admiralty Code) and cross-checking each entry. This process assists analysts in the filtration of questionable data, thereby enabling the prioritization of relevant information.

(Noise Filtering and Verification)

- The interpretation of satellite imagery is often subject to conflicting interpretations. The integration of human expertise with artificial intelligence (AI) algorithms for the detection of disinformation holds promise in minimizing misanalysis.

(Data Updates and Version Control)

- Given the dynamic nature of nuclear facility operations, it is imperative that the database be updated on a scheduled or near-real-time basis. Significant events (e.g., nuclear tests) should be meticulously documented in historical entries to facilitate effective comparison and analysis. Additionally, timelines can be utilized to monitor ongoing changes.

(Security and Access Control)

- The database (DB) may contain sensitive data or costly commercial satellite imagery, necessitating rigorous access control measures. The implementation of encryption or watermarking may be considered. In the context of international collaborations, it is imperative that policy and technical measures be implemented to ensure the appropriate management of varying access levels.

Summary of Technical Implementation Strategies

(Data Storage Methods)

- Use a relational DB (SQL) for structured data on events, facilities, and dates
- Employ a NoSQL (document-based) DB for unstructured, high-volume text (e.g., full news articles).
- Store large raw satellite images in a cloud environment, recording only metadata in the DB.

(Satellite Imagery Processing and GIS Integration)

- Manage imagery types (optical, SAR), resolutions, and cloud cover; preprocess when needed (e.g., cloud masking, radiometric correction).
- Integrate with a GIS server (e.g., Geo-Server) for map-based visualization, allowing users to query specific coordinates and dates for relevant images.

(AI Analysis and Automation)

- Imagery Analysis: Employ object detection (e.g., launchers, vehicles, building expansions), change detection, and anomaly alerts.
- Text Analysis (NLP): Extract keywords from articles and social media, conduct named entity recognition, and perform multilingual translation.
- System Integration: Automate data ingestion so newly acquired satellite images and OSINT are automatically processed and added to the DB.

Combining these elements produces a system capable of continuously tracking, analyzing, and visualizing North Korea's nuclear and missile activities.

Acknowledgements

This work was supported by the Nuclear Safety Research Program through the Korea Foundation Of Nuclear Safety (KoFONS) using the financial resource granted by the Nuclear Safety and Security Commission (NSSC) of the Republic of Korea (No. 00409134).