Transactions of the Korean Nuclear Society Spring Meeting
Jeju, Korea, May 22-23, 2025

Analysis of design requirements for the construction of an open source intelligence (OSINT)-satellite image integrated database

Yonhong Jeong*, Gayeon Ha
Division of Nuclear Nonproliferation Policy, Korea Institute of Nuclear Nonproliferation and Control, 1418 Yuseong-daero, Yuseong-gu, Daejeon 34101, Republic of Korea
**Corresponding author: jyh1404@kinac.re.kr*

Keywords: Nuclear Non-proliferation, Data platform, Methodology of monitoring nuclear activities

1. Introduction

Since the 2000s, North Korea has conducted multiple nuclear tests and medium- to long-range missile launches, steadily advancing its nuclear capabilities. However, because of the regime's closed nature, direct access for on-site inspections is highly restricted, and the dearth of officially verified data makes it difficult to accurately gauge the scale of North Korea's nuclear and missile programs or to assess the level of threat they pose.

In response to these challenges, a monitoring approach that combines open-source intelligence (OSINT) with commercial satellite imagery has gained traction. OSINT comprises publicly available information, including news reports, academic papers, social media content, and announcements from governments or international organizations. Meanwhile, satellite imagery provides direct evidence of changes in infrastructure and land use—such as building construction, ground excavation, and vehicle movements—insights that can be hard to glean from text-based sources alone.

This paper proposes a comprehensive big data database (DB) design to track North Korea's nuclear tests, missile launches, and other security-related events since the 2000s. Beyond merely collecting OSINT and satellite imagery, the aim is to integrate these sources into a unified DB that systematically supports data acquisition, analysis, and visualization, thus enabling early detection of potential warning signals. In addition, this paper reviews comparable domestic and international case studies to draw key lessons for design and implementation.

2. Analysis and Comparison of Similar Cases

Below are representative domestic and international initiatives that use OSINT and satellite imagery to monitor North Korea:

2.1 IAEA (International Atomic Energy Agency)

- After international inspectors withdrew from North Korea, the IAEA started relying on satellite imagery and OSINT to track the status of nuclear facilities such as Yongbyon.
- The IAEA partially compensates for the lack of on-site access by employing GIS to integrate geographic and temporal data.

2.2 38 North (U.S. Think Tank)

- Operates a publicly accessible "digital atlas" that uses commercial satellite images (e.g., from Maxar) and crowdsourced data.
- It serves as an exemplary model of transparency enabled by civilian efforts, allowing the public to observe changes in North Korean terrain and facilities.

2.3 CSIS Beyond Parallel

- Maintains an event database of North Korean nuclear tests, missile launches, and other provocations, supporting chronological and keyword-based searches.
- Demonstrates best practices for integrating multiple datasets to inform policy and academic research.

2.4 RUSI Project Sandstone (U.K. Think Tank)

- It uses open-source information to investigate sanctions violations, including illicit arms trading and shipping routes.
- Shares its findings with the media and relevant authorities to strengthen international sanctions enforcement.

These cases illustrate common strategies such as integrating diverse data types (imagery, text, geospatial), visualization and search tools, and using public information to fill gaps in official data. GIS-based systems are particularly effective, linking facility coordinates to satellite imagery or event databases that enable time-series analysis.

3. Key Requirements for Database Structure Design

3.1 Integration of Diverse Data Types

A single system should handle textual OSINT (news, social media, reports), satellite imagery (optical, SAR), and spatiotemporal data (coordinates, capture dates). For example, satellite images of a nuclear facility, related news articles, and the facility's coordinates/history should all be accessible in one place. The IAEA's approach similarly aggregates OSINT and satellite information to monitor facility status.

3.2 Efficient Data Storage and Large-Volume Handling

A hybrid structure is advisable because satellite imagery files are typically large: raw images can be stored in file systems or cloud services, while only metadata is recorded in the DB. Given the potentially huge volume of text-based OSINT, horizontal scalability (e.g., NoSQL) and robust backup strategies should also be considered.

3.3 Fast Search and Querying

The system should support complex queries, such as keyword searches ("nuclear test," "ICBM," "Yongbyon"), time-range queries, or geographic filters. This requires 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.

3.4 Spatiotemporal Data Management

Because nuclear and missile activities are closely tied to specific locations and times, a geospatial DB (e.g., PostGIS) is necessary for advanced spatial queries and tracking movements over time. Like 38 North's digital atlas, overlaying satellite imagery onto maps is crucial for visualizing facility changes and related activity.

3.5 Data Reliability Assessment and Metadata

OSINT can vary widely in accuracy, so each entry should be assigned a reliability score (e.g., using an Admiralty Code) and cross-checked. This process helps analysts filter out or deprioritize questionable data.

3.6 Noise Filtering and Verification

The same satellite image can sometimes yield conflicting interpretations. Combining human expertise with AI-based disinformation detection can help minimize misanalysis.

3.7 Data Updates and Version Control

Because nuclear facility activities can change quickly, the database should be updated on a scheduled or near-real-time basis. Significant events (e.g., nuclear tests) should be logged with historical entries for easy comparison. Timelines can also be used to track ongoing changes.

3.8 Security and Access Control

The DB may contain sensitive data or expensive commercial satellite imagery, so strict access control is necessary. Encryption or watermarking may be employed. In international collaborations, policy and technical measures should ensure that different access levels are managed appropriately.

4. Summary of Technical Implementation Strategies

4.1 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.

4.2 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., GeoServer) for map-based visualization, allowing users to query specific coordinates and dates for relevant images.

4.3 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.

5. Conclusions

This paper has examined the necessity and the practical approach to developing a monitoring database that integrates open-source intelligence (OSINT) with satellite imagery to track North Korea's nuclear activities. Case studies (IAEA, 38 North, CSIS, etc.) underscore the importance of consolidating multiple data sources, supporting advanced search and visualization, and addressing factors such as scalability, security, and reliability in database design.

A well-designed DB can help detect and deter North Korean efforts to hide nuclear activities, enhance global verification capacities, and contribute to a more transparent, data-driven security environment by allowing for early warning and better-informed policy decisions. Though there are technical and organizational hurdles, commercial satellite technology and OSINT infrastructure improvements now provide significantly greater monitoring capabilities than in the past. Furthermore, collaborating with the IAEA, allied nations, and private research institutions can position this DB as a pivotal resource for future denuclearization verification on the Korean Peninsula.

ACKNOWLEDGEMENT

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).

REFERENCES

[1] 38 North. (2020, 7 월 15 일). *Satellite Imagery of Yongbyon Nuclear Scientific Research Center.* https://www.38north.org/2020/07/yongbyon070120
[2] 38 North. (2022, 4 월 10 일). *Punggye-ri Nuclear Test Site Imagery Analysis.* https://www.38north.org/2022/04/punggye0408
[3] Bermudez, J. S., & Cha, V. (2018). *Undeclared North Korea: Missile Operating Bases Revealed.* Center for Strategic and International Studies. https://beyondparallel.csis.org
[4] Bermudez, J. S., Cha, V., & Kelly, K. (2019). *North Korea's Suspected Thermonuclear Warhead Development.* CSIS Beyond Parallel.
[5] Hanham, M. (2017). Commercial satellite imagery analysis for nonproliferation monitoring. *Nonproliferation Review*, 24(1-2), 91-102.
[6] Hibbs, M. (2017). *Societal verification in nuclear disarmament.* Carnegie Endowment for International Peace.
[7] Lewis, J. (2012). Thoughts on the Alleged Syrian Nuclear Facility. *Arms Control Wonk.* https://www.armscontrolwonk.com
[8] Park, J., Kim, S., & Lee, D. (2021). Applying machine learning to satellite imagery for estimating North Korea's economic activity. *Korean Journal of Remote Sensing*, 37(2), 85-101.
[9] 통일부. (2023). *2023 년 통일부 업무보고 자료.* 대한민국 통일부 공식 웹사이트.
[10] CSIS *Beyond Parallel: North Korea.* https://beyondparallel.csis.org
[11] Jane's Intelligence. *Jane's open-source military database.* https://www.janes.com
[12] IISS. *IISS Strategic Dossier on North Korea.* https://www.iiss.org
[13] NTI. *Country profiles: North Korea.* https://www.nti.org
[14] Bellingcat. *Bellingcat open-source investigations.* https://www.bellingcat.com
[15] PostgreSQL Global Development Group. *PostgreSQL*. https://www.postgresql.org/
[16] PostGIS. *PostGIS*. https://postgis.net/
[17] MongoDB, Inc. *MongoDB*. https://www.mongodb.com/
[18] Elasticsearch. *Elasticsearch*. https://www.elastic.co/elasticsearch/
[19] Apache Hadoop. *Welcome to Apache Hadoop!*. https://hadoop.apache.org/
[20] Apache Spark. *Apache Spark*. https://spark.apache.org/
[21] GeoServer. *GeoServer*. https://geoserver.org/
[22] ArcGIS (Esri). *ArcGIS*. https://www.esri.com/en-us/arcgis/about-arcgis/overview/
[23] QGIS. *QGIS: A Free and Open Source Geographic Information System*. https://qgis.org/