Sean Sobieraj, CISSP

Cybersecurity Data Architect

sean@sean.cc / sean.cc / linkedin.com/in/sobieraj

SUMMARY

Nearly two decades in cybersecurity at CISA, the White House Situation Room, and US-CERT, after starting in network operations in the U.S. Marine Corps. Career centered on giving security data structure: from US-CERT's first machine-readable malware reports, through STIX-based information sharing, to CISA's enterprise data model and Open Cybersecurity Schema Framework (OCSF) telemetry standards. Now building on that work toward semantic modeling and knowledge graphs, where a shared model of security entities and their relationships lets analysts and AI systems reason across sources.

EXPERIENCE

Senior Data Architect

Cybersecurity and Infrastructure Security Agency (CISA) 2022–Present

- Lead the Cybersecurity Division's data management initiative, setting its long-term roadmap, governance structure, and data stewardship model.
- Architected and guided the production deployment of an OCSF-aligned telemetry framework that continuously normalizes more than 100 feeds from over two dozen federal agencies into a common schema (more than three trillion events to date), enabling cross-source discovery and analysis.
- Led the development of a domain-oriented data model of fewer than 10 intuitive domains as a practical alternative to an enterprise model of hundreds of classes, designed to reduce organizational silos.
- Defined agency-wide medallion architecture, tagging, and metadata practices, including governed tags for PII, PCII, TLP, and CUI, plus a framework covering provenance, sharing, and access control across layers.
- Designed core data infrastructure on Databricks, AWS, Alation, and Elasticsearch for batch and streaming workloads, including the division's first shared, reusable pipeline framework and validation rules that keep data trusted from raw to curated layers.
- Identified a gap in the OCSF DNS schema and advanced the fix with external partners into OCSF v1.9.0; developed guidance for using OCSF and STIX together.
- Developed standardized cyber threat intelligence and entity-reference architectures, and integrated indicator and agency-reference data into a threat-hunting proof of concept.
- Proposed a graph-based model to unify visibility across customers, services, and data assets.

DevOps Team Lead, Threat Hunting

Cybersecurity and Infrastructure Security Agency (CISA) 2018–2022

- Established and led a cross-functional development team of federal and contract staff, prototyping analytic tools for threat hunting and introducing Git branching, CI/CD, and automated testing.
- Automated threat intelligence preparation for incident response, replacing hours of manual work per engagement: led development of a MongoDB-backed service that polled six external vendors and DHS's malware triage platform on a schedule and produced Splunk-ready lookup tables for each response kit.
- Built an on-premises data lake with MinIO and Dremio, integrated with cloud object storage, supporting many of the team's analytic tools.

Technical Lead, Capability Engineering

DHS National Cybersecurity and Communications Integration Center (NCCIC) 2014–2018

- Designed and built DHS's malware triage and reporting platform, cutting the time to first findings from days to minutes. It stayed in use for more than a decade and earned the DHS Office of Cybersecurity and Communications Innovation Award.
- Integrated more than 30 open-source, commercial, and custom components (including Cuckoo Sandbox, YARA, PostgreSQL, and Celery) into one web application that automatically drafted an initial findings report for analysts to review and expand, with output in six formats, including STIX XML.

- Created a PostgreSQL repository of eight years of malware report data, enabling correlation across incidents and automated trending and metrics.
- Replicated the platform at other DHS locations and provided it as the proof of concept for what became DHS's Advanced Malware Analysis Center.
- Wrote a Python-STIX marking extension for DHS's Automated Indicator Sharing (AIS) capability, and guided other DHS components on implementing STIX, CybOX, and MAEC.

Duty Officer
White House Situation Room, National Security Council 2013–2014

- Provided continuous situational awareness to the President, Vice President, National Security Advisor, and senior White House officials on rotating 12-hour day and night shifts, monitoring all-source reporting and alerting on breaking crises.
- Wrote and peer-reviewed the President's daily summaries and National Security Council summaries of global events.
- Supported the President's and senior officials' calls with foreign heads of state.

Malware Analysis Team Lead
U.S. Computer Emergency Readiness Team (US-CERT) 2008–2013

- Led malware incident response, coordinating with analytic teams, law enforcement, and intelligence agencies.
- Created US-CERT's first machine-readable malware report format and standardized the report template adopted organization-wide.
- Helped develop the database ontology, XML schema, and TLP guidelines that led to US-CERT's Cyber Indicators Analysis Platform.
- Developed the team's triage utility (Python) and report editor (JavaScript), automating analysis and reporting workflows.

Sergeant, Network Operations
U.S. Marine Corps 2000–2004

- Maintained the Camp Lejeune regional network (1,500 switches, 35,000 users); supervised three technicians and managed more than $500,000 in equipment.
- Built a fiber network from scratch for Joint Task Force 160 in Guantanamo Bay, Cuba, during Operation Enduring Freedom.

EDUCATION AND CERTIFICATION

M.S. Information Security, Purdue University (CERIAS, federal Scholarship for Service), 2008

B.S. Applied Networking and System Administration, summa cum laude, Rochester Institute of Technology, 2006

CISSP, Certified Information Systems Security Professional, (ISC)²

SKILLS

Data architecture: conceptual and domain data modeling, ontology and schema design, medallion architecture, metadata and tagging, data governance, data lineage

Standards: OCSF, STIX, MITRE ATT&CK, ECS

Platforms: Databricks (Unity Catalog, Spark Declarative Pipelines), AWS, Elasticsearch, PostgreSQL, MongoDB, Alation, Dremio, MinIO

Engineering: Python, PySpark, SQL, JavaScript, REST APIs, Docker, Git, CI/CD

Security: malware analysis and reverse engineering, incident response