SANDIA REPORT
SAND2021-7395
Printed June 2021

Sandia National Laboratories

COVID-19, An Exercise in Data Governance at Sandia National Laboratories

Ruth A. Harris
Tracy K. Jones
Macario S. Flores
David Bustamante

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods/

ABSTRACT

In April of 2020, Sandia National Laboratories had an urgent need to identify and manage the data that could be used to create mobile applications, models, reports, and visualizations to assist management in safely bringing the workforce onsite during the COVID-19 pandemic. Multiple divisions volunteered to design and build software solutions; meanwhile, requests for new data sources, including duplicate requests, were inundating Information Technology (IT) and data owners. The Enterprise Data Governance Team was assigned to resolve obtaining and accessing new sources of data in an accelerated timeframe. Through successful collaboration with multiple stakeholders and domain owners across Sandia, the Enterprise Data Governance Team rapidly developed a centralized data strategy and solution for use in safeguarding the Sandia workforce during the COVID-19 pandemic. This foundation enabled teams to successfully develop solutions, including reports for executives and management as well as the data for modeling and scientific analysis.

ACKNOWLEDGEMENTS

Brian N. Post, Senior Manager, Enterprise Information Management, determined critical data management was needed and asserted the Enterprise Data Governance Team as point of contact to field requests and deliver the data solution.

Data Subject Matter Experts (SMEs) were essential for constructing an integrated data solution and ensuring resiliency for multiple uses from applications to reports to data modeling and analysis.

Bernadette S. Edge, Database Engineer, Database/Data Warehouse

Lisa J. Eshelman, Solutions Architect, Facilities, Heat Map

Jeremy B. Myers, Solutions Architect, Sandia Insights (data source for data scientist)

Jennifer M. Rivera, Computer Scientist, Medical Systems

Eric Smyth, Software Systems Engineer, Sandia Insights (data source for data scientist)

Eric T. Trujillo, Solutions Architect, Safeguards & Security

Eileen J. Voth, Solutions Architect, Contact Tracing

Enterprise Data Governance Team:

David Bustamante, Solutions Architect

Macario S. Flores, Software Systems Engineer

Ruth A. Harris, Computer Scientist

Tracy K. Jones, Solutions Architect

CONTENTS

LIST OF FIGURES

This page left blank

EXECUTIVE SUMMARY

Engaging the Enterprise Data Governance Team to manage and architect the enterprise data for COVID-19 got the right data to the appropriate people to aid in safely bringing people onsite and managing the workforce. Chaos was reduced by providing one point of contact via a single email account for all COVID-19 data requests. With one source of truth for the data, reconciliation issues were removed. The foundation was built for a dashboard and reports, which were quickly developed and published for ease of use and secured based on restricted access. This data framework has served well by continuing to evolve and quickly pivot to ingest new data sources as well as provide different views of the data based on new use cases.

ACRONYMS AND DEFINITIONS

Abbreviation	Definition
COVID-19	Novel coronavirus disease 2019, abbreviated as COVID-19 which is the official name given by the World Health Organization on February 11, 2020. According to the CDC, CO stands for corona, VI for virus, and D for disease.
DOE	U.S. Department of Energy
EBI	Enterprise Business Intelligence
GIS	Geographic Information System
NNSA	National Nuclear Security Administration, a semiautonomous agency of DOE
SFO	Sandia Field Office for NNSA
SME	Subject Matter Expert
SNL	Sandia National Laboratories

1. HEADING UP THE CALL FOR ENTERPRISE DATA DURING THE COVID-19 PANDEMIC

In April 2020, Sandia executives and line management had an urgent need to quickly gather data to inform their decisions regarding the well-being of the workforce. The worldwide COVID-19 pandemic was developing, and data collection, analysis and reporting were needed. Requirements were fluid and largely unarticulated. No one had the answer, yet many were valiantly answering the data call and trying to gather the information.

Complexity ensued with no one entity responsible for the data or reporting. Adding to the confusion, different groups generated spreadsheets to report from causing reconciliation issues and inability to validate or trust the data.

The senior manager of Enterprise Information Management asserted the Enterprise Data Governance Team as the point of contact to collect the requests and provide the enterprise data solution.

Adapting and evolving both the process and delivering the data solution for multi-use consumption provided the foundation to easily pivot to address unforeseen needs. The business benefits derived include one source of truth that removed reconciliation issues, can be used many times, is repurposed for new requirements, and provides ability to trend the data.

1.1. The challenge

The Enterprise Data Governance Team was appointed as the primary point of contact to field requests and organize the varying needs in an accelerated timeframe to provide consolidated, trusted data for enterprise use, along with the appropriate approval to access the data. This was the first type of its engagement for this newly formed team (about one year).

1.2. Field the requests

The collection of requests primarily came through email. While some phone conversations were occurring among members of management and others, those resulted in emails so that specifics could be captured and directed to the Team.

1.3. Construct the process

Before designing the data or technical solution, the Team took the time to strategize the approach for this effort recognizing the urgency in terms of timeliness and need for trusted data. There were four elements to define and resolve.

- Identify requests and stakeholders (see Appendix A Stakeholders)
- Determine needs and use cases as source for requirements
- Employ a divide and conquer process
    - Have Enterprise Data Governance Team representation with the primary stakeholders who were active in the process and represented distinct domains (e.g., business units)
- Develop approval method and identify approvers

As there were no clear requirements or primary stakeholders, it was incumbent on the Team to create a process that would identify who would approve the data capture along with the data use cases. The emails and phone conversations were the source from which the needs were determined. These needs were the basis for the requirements.

The Team identified the primary stakeholders some of whom were data owners, policy owners and others who were tasked with specific functions such as leading the data analytics and modeling functions. The Team divided up and worked with the primary stakeholders to both collect additional requirements and represent the status of the data solution. This was essential for the Team to keep up with the dynamic and ever-changing environment and ensure data efficacy.

One major effort was to determine who the approvers were for data acquisition, data use and the users of the data. While there were many who felt the need to be part of the authorization process, most were simply trying to keep their management informed. The data owners and policy owners of the specific data systems and policies provided authorization.

Another activity the Team engaged with was providing enterprise data for modeling and analysis. The Team facilitated the approval and use so that there were no other sources of internal data being used.

1.4. Architect the flexible data framework

Design an integrated solution to provide one source of truth that can be trusted.

- Determine data elements required to support various needs such as external reporting, use in applications and processes like health check, management reporting and most importantly historical data capture
    - The requesters and stakeholders were asking for the current snapshot, yet the team immediately recognized the need for building the temporal data as part of the design
- Seek data sources and assign data SMEs
- Employ security mechanisms to ensure authorized use
    - Secure data access and authorization via email to primary stakeholders
    - Use existing security tools or methods such as specifically defined user groups to construct restricted access
- Leverage data architecture
    - Build onto existing data warehouse created for diverse data integration needs
    - Augment data virtualization to provide views for data modeling and analysis
- Automate data builds and data feeds to ensure timeliness, availability and accuracy
    - Synchronize various data sources
    - Develop data ingestion routines
    - Employ data validation to eliminate errors
- Continuously apply lessons learned and review the status for perspective to evolve the approach

1.5. Develop the data ecosystem

Develop multi-use solution for applications, reports, visualizations and dashboard.

- Leverage the existing Enterprise Business Intelligence (EBI) Portal as the delivery and authorization/authentication mechanisms for reports and visualizations for executives and management
- Design an agile data structure to be used in various scenarios
    - Minimize rework or need to refactor data structures
- Design data visualizations to be filtered and viewed from many different aspects (e.g., number of occupants by building, number of onsite personnel by organization, etc.)

1.6. Deploy adhoc solution

Time is of the essence. Deliver adhoc reports via spreadsheets or other mechanisms built from the enterprise data solution to enable one system to trust. This will negate the inability to reconcile the many local data sources that are typically kept in spreadsheets. Note to the requesters that this adhoc report is temporary and will be suspended when the full solution is deployed.

1.7. Deliver data

Once the ecosystem is complete, deliver the data via productionized and therefore reliable mechanisms such as the dashboard in the EBI Portal for a common access point and user experience. Also, through the data virtualization views for data scientist consumption.

- Leverage existing reporting framework of EBI Portal
    - Dashboard
    - Security mechanism
    - Visualizations
    - Reports
- Augment existing applications, such as the Heat Map application to display building occupancy usage
- Rapidly deploy virtualized views for data scientist modeling and analysis

1.8. Iterate quickly and apply lessons learned

The need to be agile and responsive requires the ability to quickly iterate with any part of the lifecycle and not be constrained to the typical data lifecycle. Recognize that lessons learned will be a continuous part of the process. Constant communication is essential within the team as well as with the requesters and primary stakeholders.

1.9. Sandia's Data Governance Playbook

At Sandia National Laboratories (SNL), constructing the process is key as it is the means that enables "just enough governance" and launches the data development activities while requirements are being flushed out. It allows for agility and a responsive posture for delivering a trusted data solution. This process is encapsulated into a playbook shown in Figure 1-1. The playbook can be used when there is a need for quick, rapid response.

Naturally, this SNL Data Governance Playbook for Rapid Solutions can be tailored to meet specific needs. However, it should be emphasized and encouraged that one cycle does not have to be complete before new requests can be fielded. For example, there may be more data elements identified which can be added to the data sourcing right away and does not have to wait for a planning cycle.

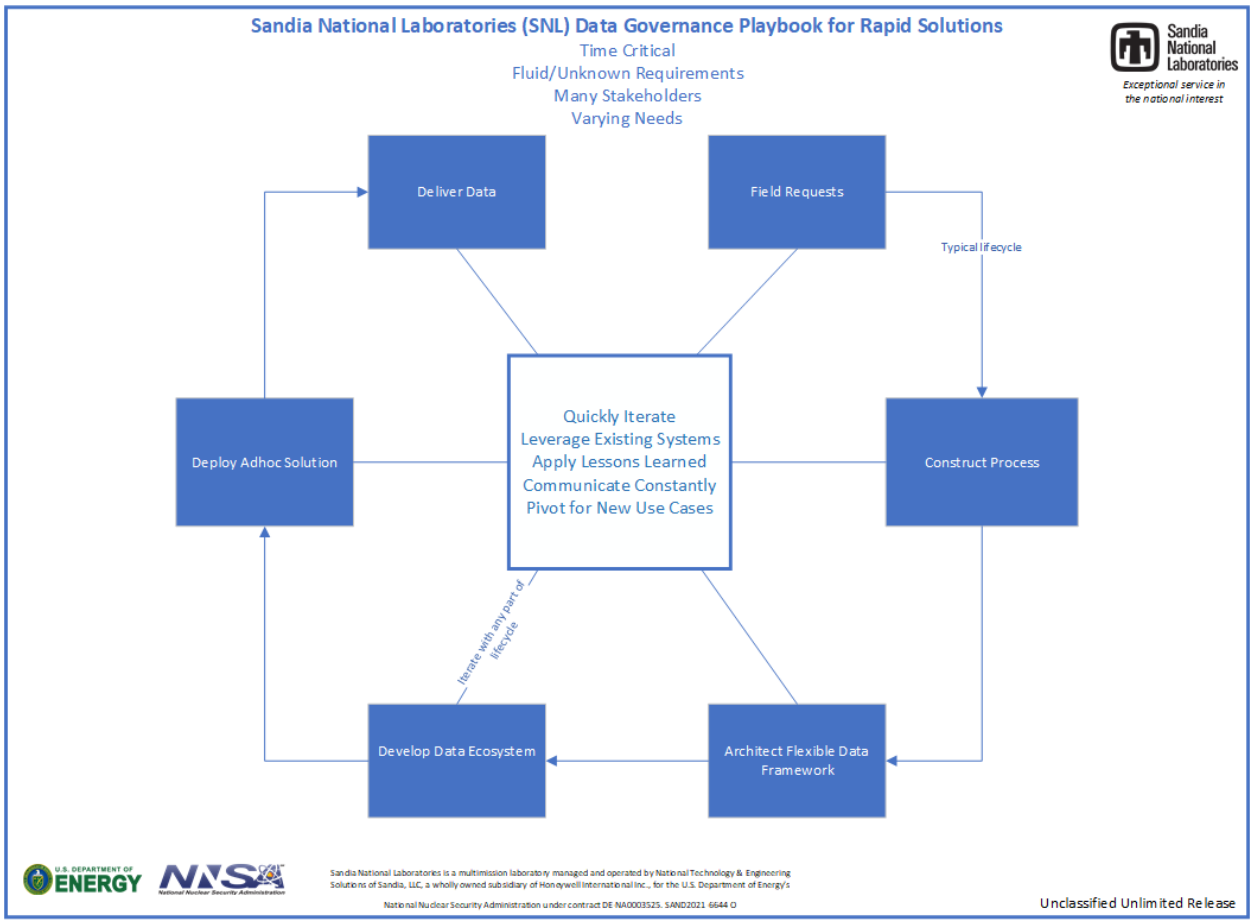

Figure 1-1. SNL Data Governance Playbook for Rapid Solutions

1.10. Resilient results

Success was achieved by employing data governance. The result was a unified iterative approach to data management that ensured data quality was designed from the beginning, time delays were reduced, and the data could be trusted. While it was not anticipated that this would be an ongoing process, this playbook continues to meet the various demands and serve the needs of the requesters and application providers.

There were many aspects of data governance that led to success and continue to meet urgent and changing needs. Foremost, a team of data professionals with the expertise and extensive knowledge of the data provided a cohesive approach from data acquisition to data delivery. Additionally, having a data architecture with flexible data systems and a reporting ecosystem that were quickly leveraged to meet the demand.

Business benefits were derived by using one source of truth that enabled rapid development of applications and reports for current state and trending analysis; and a responsive process that quickly obtained additional data and delivered accurate data.

The COVID-19 enterprise data solution has proven resilient by integrating additional data sources as needed; quickly developing, delivering and producing reports with consistency and reliability; and removing reconciliation issues as well as reducing time delays. The efficacy is demonstrated by meeting new and varying needs of few of which are listed below.

- New use case for Facilities to aid in prioritizing efforts to address air flow in buildings based on occupancy usage which was merged into existing Heat Map application
- New data calls fielded including those from external entities such as Sandia Field Office (SFO)
- New filters developed to add different views of the data based on audience
- Readily facilitate changes in authorized personnel in the varying roles of executives, management, application developers and data scientists

This page left blank

2. CALL TO ACTION

Despite immediate pressures to move fast, employ a forward-looking view to data management considering how the data might be useful and used in future scenarios. Recognize that lessons learned must be incorporated real-time into the design of the process and data for a truly agile, learning data governance approach.

Here are a few questions and actions to consider for success which can also help the team keep a positive attitude during a stressful and demanding time.

- Can this approach and solution be used again?
- When will the next pandemic occur?
- Will historical perspective be beneficial e.g., temporal data capture?
- Establish a small, nimble core data governance team to manage the data calls and coordinate between data providers and data architects
- Employ experienced data professionals with requisite expertise to quickly navigate the various source systems, deliver right solution and guide the SMEs and stakeholders on approach
- Bring along new or less experienced talent if you have the opportunity as this exercise will set the stage for the next success
- Leverage what is in place such as a reporting ecosystem, data architecture and security mechanisms
- Deliver quickly in adhoc methods such as spreadsheets until the automated solution is built
- Identify essential stakeholders who will quickly move the effort along versus those who want to participate and just need to be kept informed
- Communicate frequently
- Be responsive not only to stakeholders but with core team members
- Seek out engagement with core team to ensure all are working on the same objectives and priorities
- Empower the team, help break down barriers, do not wait for clear-cut direction or requirements
- Celebrate success

APPENDIX A. STAKEHOLDERS

The champions and stakeholders for the collection and use of enterprise COVID-19 data are listed by their role and area they represent. Stakeholders were identified by the requests that were fielded primarily via email. Those who were central to the authorization process, a handful of people, recognized the importance of a quick responsive posture to aid in getting the data needed and approval for access and use of the data.

Role	Domain
Champion: Sandia Leadership Team (SLT)	Executive Management
Director	Safeguards & Security
Director	Information Operations
Acting Director	Information Technology Services, Chief Information Officer (CIO)
Director	Cyber Security & Mission Computing, Chief Information Security Officer (CISO)
Director	Medical Services
Chief Privacy Officer	Regulatory and Contract Law
Senior Manager	Enterprise Information Systems
Senior Manager	Enterprise Information Management
Senior Manager	Medical Services
Senior Manager	Executive Chief of Staff
Manager	Executive Strategic Planning
Manager	Ops Research & Computational Analysis
Manager	Data Architecture and Management
Manager	Data Sciences
Manager	Enterprise Business Intelligence
Manager	HR Systems
Manager	Learning & Health Systems
Manager	Physical Systems
Manager	Technical Systems
Manager	Operations CA
Manager	Strategic and Technical Ops
Manager	Strategic Capital Planning

Figure A-1. Stakeholders

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
John D. Zepper	09000, 09700	jdzeppe@sandia.gov
Ruth A. Harris	09730	raharri@sandia.gov
Tracy K. Jones	09730	tkjones@sandia.gov
Brian N. Post	09730	bnpost@sandia.gov
Cynthia M. Huber	09734	cmhuber@sandia.gov
Jeremy B. Myers	09734	jbmyers@sandia.gov
Eric Smyth	09734	esmyth@sandia.gov
David Bustamante	09735	dgbusta@sandia.gov
Macario S. Flores	09735	msflore@sandia.gov
Meara A. Satpathi	09735	masatpa@sandia.gov
Bernadette S. Edge	09738	bsedge@sandia.gov
Michael L. Hagengruber	09738	mlhagen@sandia.gov
Laura M. Charles	09740	lmcharl@sandia.gov
Lori A. Kozlowski	09744	lakozlo@sandia.gov
Jennifer M. Rivera	09744	jjriver@sandia.gov
Sabine A. Boruff	09745	saboruf@sandia.gov
Lisa J. Eshelman	09745	leshel@sandia.gov
Julie Smith	09746	julsmit@sandia.gov
Eileen J. Voth	09746	ejvoth@sandia.gov
Michael L. Montoya	09747	mlmonto@sandia.gov
Eric T. Trujillo	09747	ettruji@sandia.gov
Bonnie L. Hammond	09750	blhammo@sandia.gov
Technical Library	01977	sanddocs@sandia.gov

This page left blank

This page left blank

Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.