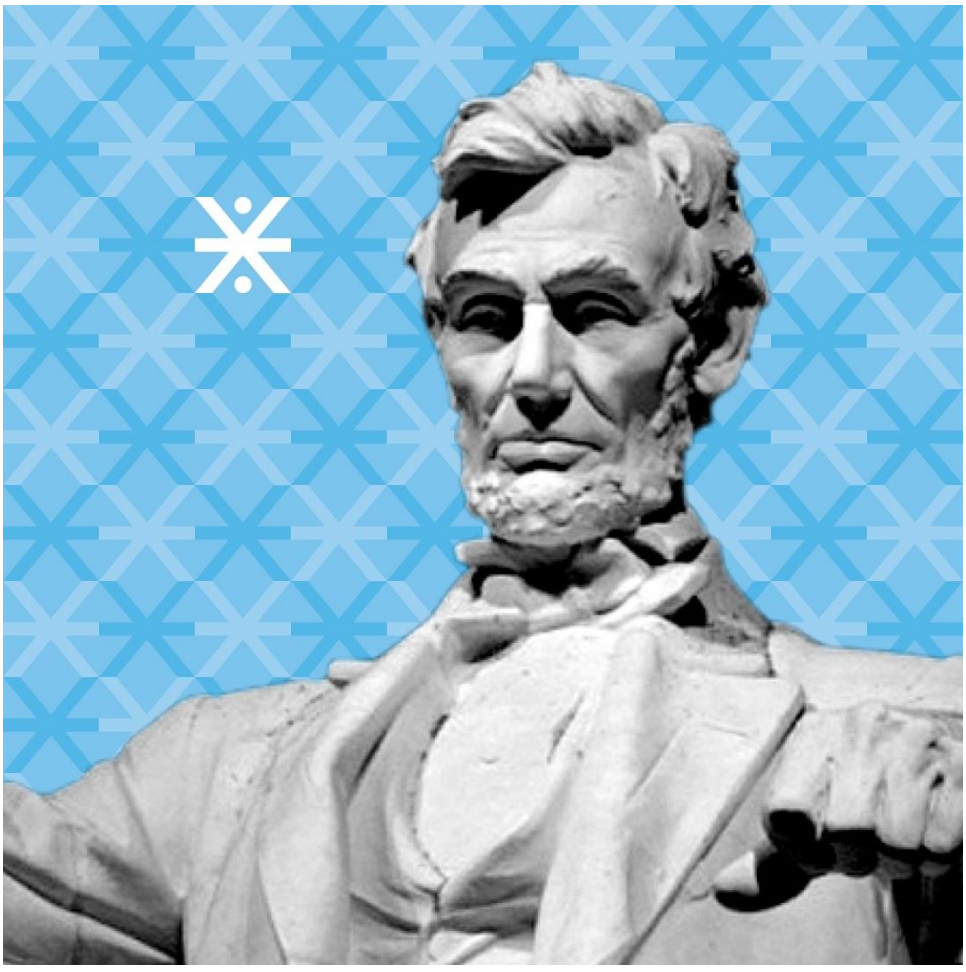
summit

complexity simplified.

Database Development and Integration Capabilities Statement

Submitted to Department of Labor (DOL) Occupational Safety and Health Administration (OSHA)

February 17, 2017



SUBMITTED BY:
Summit Consulting, LLC
DUNS: 134631675

601 New Jersey Ave, NW
Suite 400
Washington, DC 20001
www.summitllc.us

CONTACT:
China Layne
Manager
202.407.8300 Tel
509.984.8943 Fax
china.layne@summitllc.us

About Summit

Summit Consulting, LLC (Summit), headquartered in downtown Washington, DC, is an analytical advisory firm that specializes in applying cutting-edge quantitative techniques to the real-world challenges facing federal agencies and private-sector enterprises. Since being founded in 2003, Summit has grown to a staff nearly 100 professionals with annual revenue over $15m. More than 60% of Summit's staff members hold advanced degrees in economics, statistics, and public policy and have strong data management, statistical programming, and analytical capabilities. Our highly skilled staff members have versatile backgrounds and can apply data management and analytics techniques on a variety of topic areas, including occupational safety and health policies, other worker protection policies, program evaluation, and supporting enforcement programs.

Summit has the expertise to support OSHA's enforcement database efforts:

- **Developing complex and integrated databases,**
- **Developing business intelligence tools,**
- **Expertise with enforcement administrative data, and**
- **Knowledge of OSHA enforcement programs**

Summit has conducted multiple contracts of developing databases to enforcement programs for the U.S. Department of Labor (DOL) and U.S. Department of Housing and Urban Development (HUD), among other Federal agencies. Summit has four specific areas of expertise to bring to supporting OSHA's efforts to build high-quality enforcement databases.

Summit's Relevant Expertise

1. Developing Complex and Integrated Databases

Through our work with DOL and HUD, Summit has gained significant expertise in developing complex and integrated databases to support programs. We highlight two contracts below:

DOL EBSA: Our team developed the **EBSA's Integrated Database** to support the agency's enforcement programs. This database links yearly Form 5500 data into a longitudinal database and integrates it with data from other EBSA offices, including investigative, enforcement, and audit data, as well as external financial information from companies such as Experian and service provider information from BrightScope. The Enforcement Integrated Database allows EBSA staff to see the complete filing, enforcement, audit, and financial status of each retirement plan in EBSA's purview over nearly a decade.

HUD Federal Housing Administration (FHA): Summit supports FHA efforts to estimate the long-term cost of the Mutual Mortgage Insurance Fund. As part of this work, we developed an **integrated database** to support all of the Risk Management and Capital Assessment Projections (RMCAP) reporting and modeling functions. Summit developed and managed the full extract, transform, and load (ETL) process to collect data from FHA databases and source systems on a monthly basis and load them into one integrated database. This database provides FHA staff with full loan level histories of each of FHA's 32 million loans.

2. Developing Business Intelligence Tools to Support Enforcement Programs

For several of our projects, Summit has paired the work of developing complex and integrated databases with the creation of business intelligence (BI) tools that utilize these databases to support our clients' programs. Here are the highlights of three of our BI Tools:

HUD FHA: Our team designed and built a **query tool** that has a user-friendly interface and can be used in FHA's computing environment. The stored data and query capability permits FHA staff to retrieve relevant data, quickly run policy analyses, and easily access and manipulate analysis results. The query tool is also adaptable so that changes and enhancements can be easily accommodated.

DOL Mining Safety and Health Administration (MSHA): Summit created **business intelligence tools** to allow EBSA staff easy access to analysis results, as well as the ability to manipulate the analysis inputs as desired and maintain the analyses models with new, incoming data. These tools helped the underlying analyses progress beyond a single use study to being an ongoing application that supports MSHA's enforcement and programmatic activities.

DOL EBSA: Summit developed a set of business intelligence tools that allow EBSA investigative staff to quickly and easily query relevant information from the Integrated Database to support and inform enforcement activities in the field.

3. Expertise with Enforcement Program Administrative Data

Over our years of working with several agencies, Summit has gained substantial expertise with administrative data for enforcement programs. Throughout these projects, Summit has gained extensive experience with the administrative data of many of DOL's enforcement programs including those at: OSHA, Wage and Hour Division (WHD), EBSA, Office of Workers Compensation Programs (OWCP), MSHA, and Office of Federal Contracting Compliance Programs (OFCCP).

- Summit has provided statistical, research, and evaluation support to EBSA's Office of Enforcement using the agency's administrative data.
- In support of DOL Chief Evaluation Office (CEO)'s Data Analytics Unit, Summit conducted quick-turnaround analyses for enforcement programs across DOL agencies.
- Our team has worked with CEO on the Administrative Data Research and Analysis (ADRA) initiative, an effort to assist DOL agencies with leveraging their administrative data for programmatic evaluation, research, and analysis.

Through these and other projects, Summit has developed a proven ability to quickly understand and use new administrative data sources.

4. Knowledge of OSHA Enforcement Programs

Summit has gained significant experience with, and knowledge of, OSHA enforcement. Summit has:

- Conducted an evaluation of the impact of OSHA's Site Specific Targeting (SST) program to determine whether this targeted enforcement program leads to improved compliance and worksite safety and health outcomes.
- Designed and conducted an evaluation of the outcomes of the Federal Agency Targeting Inspection (FEDTARG) program to assess whether the program is associated with better safety and health outcomes for Federal worksites.

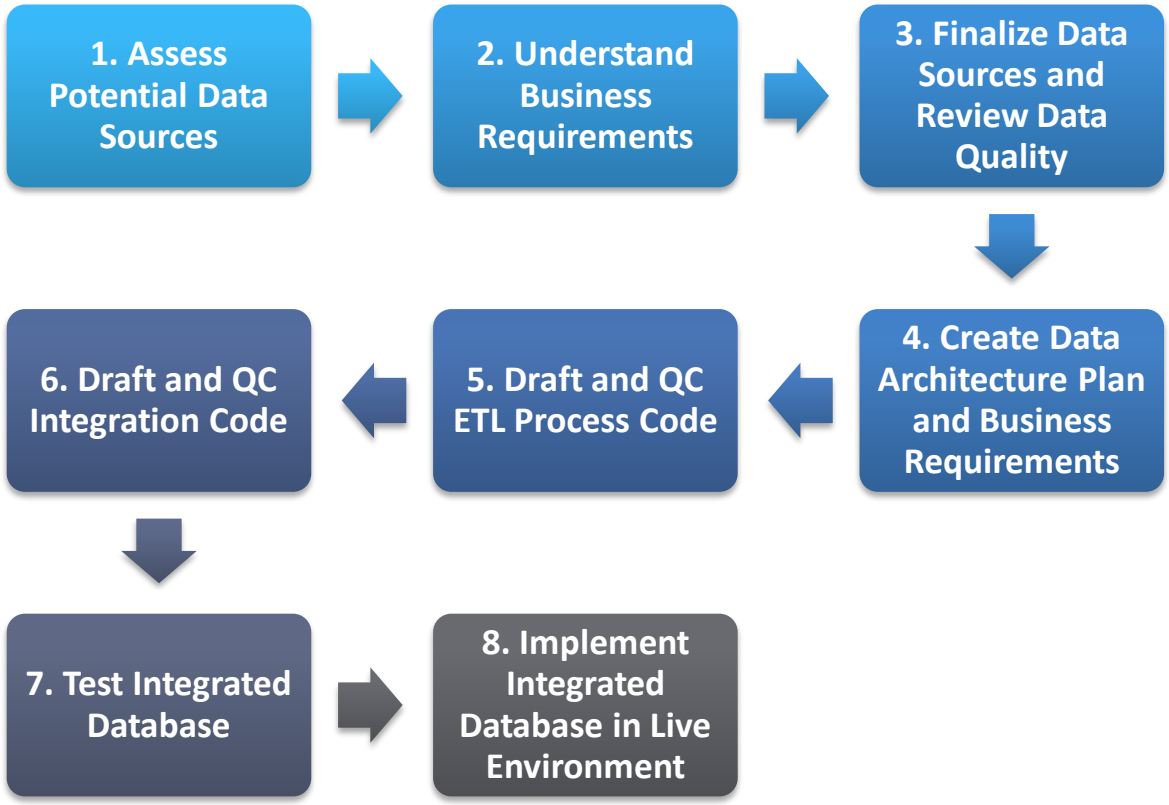
- Designed and conducted an analysis to determine which company, investigation, and enforcement characteristics are most associated with companies contesting citations of safety and health violations.

Summit's expertise and experience positions us to seamlessly apply our data management and analytics skills to help OSHA develop high-quality, fully integrated databases that can provide the data necessary to support the agency's enforcement programs.

Summit's Database Development and Integration Process

Summit uses a proven process for developing complex databases that are high-quality and fully-integrated. In this section, we discuss each step of our database development and integration process. In the following figure, we illustrate the development process.

Figure 1: Summit's Database Development and Integration Process



1. Assess Potential Data Sources

The first step is assessing potential data sources. We first work to identify *current* data sources; their content and usages. We also work to identify *new* data sources; outlining their content, potential, usages, and feasibility for integration. To identify all of these data sources, Summit leverages our knowledge of agency specific data sources and general data experience. Our team also holds working group meetings with agency staff and other program stakeholders. We ask the stakeholders which information sources, databases, and data elements from the set of identified potential data sources are

important for their office in meeting their strategic missions. Data source assessment is a critical process to verify that the database developers and database users have the same assumptions and knowledge about the underlying data sources that will be used to build the new database. Summit's data source assessment helps us identify the primary data sources and perform the data quality control reviews.

2. Understand Business Requirements

Summit collects and synthesizes the business requirements for the integrated database. We gather this information through working group meetings with agency staff and other program stakeholders. During these meetings, our team learns the agency's desired behavior for the database with respect to how the database values, creates, stores, uses, archives, and deletes data. The business requirements ensure that the stakeholders have buy-in on the utility and functionality of the database. The requirements will also define which items are high and low priority for inclusion and development in the database and inform stakeholders on the process for how the database will be constructed. The business requirements help Summit tailor the database to the specific needs of each client and make a highly functional and desired product for the client.

3. Finalize Data Sources and Review Data Quality

Summit recommends the data sources that will serve as primary sources. Our team reviews the primary data sources' documentation and contents and business requirements to determine what should be included in the ETL. Working with the agency's IT department, Summit begins to assess data quality of these primary databases along two dimensions:

Intra-database quality control review: Summit checks for outliers, calculation errors, missing or null values, and contradictions in business rules. The agency's IT department reviews and validates our assumptions in defining quality bounds.

Inter-database quality control review: Summit assesses the consistency of information included across databases that may be integrated.

4. Create the Data Architecture and Business Requirements Plans

Summit uses information gained from the intra- and inter-database quality control reviews to develop the data architecture plan for the integrated database. We follow several steps in developing the Data Architecture Plan:

- Drafting architecture diagrams to describe the proposed data hierarchy and structure of the integrated database and the relationships of the component databases to each other.
- Identifying which information can and cannot be linked across the individual data sources, and
- Developing strategies for integrating individual data sources into an integrated prototype database.

Summit documents the decision rules and accountability framework of the integrated database in the Business Requirements (BR) Plan. This plan details the expected behavior of the integrated database with respect to how data is valued, created, storage, uses, archived, and deleted. Specifically, the BR plan carefully describes the data standardization process, in terms of manipulating and de-conflicting independent data sources prior to integration, to ensure consistent definitions in the integrated database. Likewise, the plan establishes rules for data usage and data definitions based on the

requirements provided by stakeholders. Working with the agency's IT department, Summit identifies and documents in the BR plan the location, management and administration of the integrated database.

Throughout this step of the database development, Summit collaborates closely with the agency's IT department, to ensure the prototype database meets the agency's current privacy, compliance, and security requirements.

5. Draft and QC ETL Process Code

Summit develops statistical programs to extract the individual data sources from their original locations, apply any necessary data transformations for the integrated database, and load the data into our integrated database platform. The ETL Process code conforms to the guidelines outlined in our Data Architecture and Business Requirements Plans. Summit conducts a full quality control review of the ETL Process code during this development process.

6. Draft and QC Integration Code

Summit develops statistical programs to construct a dataset that accumulates and integrates data elements from the identified data sources. This prototype integrated database will provide information at several levels of analysis, as useful for the agency. The Integration code conforms to the guidelines outlined in our Data Architecture and Business Requirements Plans. Summit conducts a full quality control review of the integration code during this development process.

7. Test Integrated Database

Following the rules presented in the Data Architecture Plan, Summit begins accessing and integrating the individual databases to create the integrated database, using the ETL Process and Integration programs developed in the previous two steps. We perform the integration of data in a test environment, using protocols to check the consistencies of data mappings prior to building an integrated database. Our team segments the integrated database into a training and test data set during this testing phase, to facilitate a comprehensive and systematic analysis of the integration process.

Using the training data set, Summit performs quality checks to identify potential quality issues within the data, such as validating variable values, calculations and length, and identifying outliers, interpretation errors, or potentially erroneous values. Our team also compares summary statistics from the integrated dataset against the individual, original data set results to assess potential loss of information. After analyzing the integrated data, Summit presents these summary statistics to stakeholders and works closely with the agency to modify or expand business rules, given the results of the data integration tests.

8. Implement Integrated Database in Live Environment

Summit implements the integrated database in a live environment only after fully reviewing the integrated database in our testing environment. Through this testing, we ensure that integrated database behaves according to documented business requirements *and* that these requirements reflect the agency's expectations for the database.

Select Experience

In this section, we describe the specific projects at DOL and HUD that best highlight Summit's expertise with database development and enforcement administrative data. In the following table, we map the work of these projects to our specific expertise areas.

Table 1: Summit's Qualifications for Database Development

Projects	Expertise Areas			
	Developing Complex Databases	Developing Business Intelligence Tools	Enforcement Program Administrative Data	OSHA Enforcement Programs
DOL EBSA – Data Integration	✦	✦	✦	
HUD FHA – Actuarial Review of Mutual Mortgage Insurance Fund	✦	✦		
DOL CEO - Administrative Data Research and Analysis		✦	✦	✦
DOL CEO - Analytical and Evaluation Support			✦	✦

Data Integration To Support EBSA Enforcement	
Agency	Employee Benefits Security Administration (EBSA), Department of Labor (DOL)
Project Officer	Keith Bergstresser; West Garrett
Period of Performance	9/2012 – 9/2016; 9/2016 – 9/2017

Project Description:
Since fiscal year 2012, Summit has provided statistical and economic analysis within the Office of Policy and Research (OPR) at DOL EBSA. EBSA's objective for this project are to:

1. Estimate a predictive model that can risk rank employee-sponsored benefit plans based on their propensity to violate the Employee Retirement Income Security Act (ERISA) and
2. Develop an integrated database that can support the predictive model as well as other statistical analyses for EBSA's enforcement activities.

To develop the predictive model, Summit assists EBSA in the refinement of the current risk assessment process. Summit applies economic theory and advanced statistical methods to analyze the statistical relationship of a variety of employee benefit plan indicators to the plan's propensity to violate ERISA. Summit analyzes and tests a slate of existing and new indicators to include in the model, including plan financial characteristics, EBSA investigative activities, and indicators taken from the plan's Form 5500 filings. This outcome-based assessment process provides investigators with an effective, efficient method for targeting those employee benefit plans potentially at risk of non-compliance with ERISA.

To develop the Integrated Database (ID), Summit linked multiple years of Form 5500 filing data together and connected this data backbone with investigative and enforcement activity data from the Office of Enforcement (OE), audit data from the Office of Chief Accountant (OCA) and company financial information from external sources such as Experian and BrightScope. This work required understanding the business requirements for the database and the advantages, disadvantages, and requirements of the potential data sources for the ID. Summit designed the database to reflect these various requirements, abilities, and limitations. As part of this work, Summit extensively reviewed and improved the quality of the data to be integrated. Summit drafted and fully tested extraction, transform, and loading (ETL) process code and integration code to fulfill the database design. Finally, Summit implemented the ID in a live environment, i.e., EBSA's data systems, which required ensuring the ID was compatible with EBSA's security requirements. As implemented, the Integrated Database provides a readily accessible source of a range of plan, firm, and program data to support EBSA enforcement activities and staff in the field.

Actuarial Review of FHA's Mutual Mortgage Insurance Fund	
Agency	Federal Housing Administration (FHA), U.S. Department of Housing and Urban Development (HUD)
Project Officer	Shawn R. Jones
Period of Performance	4/13 – 1/16

Project Description:
FHA engaged the Summit Team to provide econometric and cash flow modeling to evaluate the Mutual Mortgage insurance Fund (MMIF) for single family forward mortgages and Home Equity Conversion Mortgages (HECM). The Summit Team developed a suite of models using SAS and SQL to estimate the economic net value of the Forward and HECM MMIF portfolios. The Forward model estimates the probability of claim and non-claim terminations, while the HECM model estimates terminations due to death, move-out, refinance, and default. These forecasts, in conjunction with estimates of loss given default, are used to compute economic value (EV) of the portfolios. After developing the models, the Summit Team provided rigorous advisory support for the models' review and approval by OMB. Furthermore, the Team conducted an independent actuarial review of the MMIF and assisted with the production of FHA's Annual Report to Congress.

As a first step to creating these models, Summit analyzed FHA information systems and created the Risk Management and Capital Assessment Projections (RMCAP) Database. The Summit Team developed a database to support all RMCAP reporting and modeling functions using an AWS Redshift columnar database cluster. Summit developed and managed the full extract, transform, and load (ETL) process to collect data from FHA databases and source systems on a monthly basis and load them into the Redshift database cluster. Using the FHA source data, Summit developed custom database tables which contained loan level histories of each of FHA's 32 million loans (for Forward and HECM) including origination characteristics, delinquency periods, termination/claim events, and associated cash flows to and from FHA.

Administrative Data Research and Analysis (ADRA) 2013 and 2015	
Agency	Chief Evaluation Office (CEO), Department of Labor (DOL)
Project Officer	Molly Irwin and Celeste J. Richie; Janet Javar
Period of Performance	5/7/2013 – 5/6/2016; 9/30/15 – 9/29/17

Project Description:
Through the ADRA contract, Summit conducts evaluation, research, and analyses on worker protection programs for agencies across DOL utilizing the agencies' administrative data. For these projects, Summit accesses the administrative data, reviews and cleans the data, links the administrative data with external data sources as necessary, and designs evaluation and analyses that appropriately utilize the data. The research projects culminate in technical reports or research briefs, as well as briefings or presentations.

Select Projects:

OHSA SST Evaluation:
Design
The Occupational Safety and Health Administration's (OSHA's) Site-Specific Targeting (SST) Program aims to maximize safety and health by focusing OSHA's enforcement actions (inspections and high-rate letters) on worksites with the highest rates of injury and illness. The SST program impact evaluation uses two different evaluation designs, an experimental randomized control trial (RCT) and a quasi-experimental regression discontinuity design (RDD), to assess the effectiveness of these enforcement actions - inspections and high-rate letters - in improving workplace health and safety outcomes. The RDD impact analysis takes advantage of the thresholds set by OSHA to assign worksites to various enforcement actions to construct a comparison group of worksites that are similar to the treatment worksites in all ways except their exposure to OSHA enforcement actions.

Data
The data used in the RCT and RDD evaluations come from a number of different administrative data sources, including OSHA Data Initiative (ODI) survey responses, OSHA Information Systems (OIS) data on inspections, injury/illness logs, and the agency's high-rate letter lists. Summit used a probabilistic record linkage algorithm to link worksite records from the various datasets. For the RDD evaluation, Summit created a longitudinal dataset of ODI survey responses from 1998 to 2011 linked with inspections data and high-rate letter lists.

Deliverables
The final deliverables of the evaluation are (1) a report on the results all experimental and quasi experimental analyses and (2) a public use data file of all analyses conducted.

Federal Agency Targeting Inspection (FEDTARG) Program Study:
Summit conducted an exploratory analysis of the Federal Agency Targeting Inspection (FEDTARG) program for the Chief Evaluation Office (CEO) using Federal Employees Compensation Act (FECA) data from the Office of Workers Compensation Programs (OWCP). The FEDTARG analysis examined the activities of the program (types of Federal worksites inspected by OSHA), the outputs (violations and standards cited during inspections), and the outcomes (before and after injury and illness counts) of the program.

Summit used complex data merging techniques to link data on inspected Federal worksites from OSHA's Enforcement Data (OIS and IMIS databases) with data on lost time cases (LTCs) and fatalities from the Office of Workers Compensation Programs (OWCP) FECA database into a single analysis file. Summit used descriptive statistics to present the types of worksites that were most commonly inspected through the FEDTARG program and used negative binomial regression methods and logistic

regressions to control for worksite characteristics (region, department, industry, union status, and worksite size). These regression methods were used to determine how worksite characteristics correlated with LTC counts, number of violations, specific OSHA standards violations, and violation types. Summit also used negative binomial regression to make before-and-after comparisons to determine the relationship between FEDTARG inspections and LTCs. Summit produced a final report on the methodology used in the analysis, the results of the analysis, and corresponding public use files and restricted use files.

Chief Evaluation Office Analytical and Evaluation Support	
Agency	Chief Evaluation Office (CEO), Department of Labor (DOL)
Project Officer	Jonathan Simonetta and Scott Gibbons
Period of Performance	10/1/2013 – 9/30/2014

Project Description:
Summit provided technical support to the Chief Evaluation Office (CEO) on all analytical and evaluation activities undertaken by the agency. For analytical support services, Summit provided a team of researchers and statistical analysts to assist CEO in designing, conducting, and reporting quick turnaround evaluation and statistical analyses for a variety of Department of Labor (DOL) agencies.

These projects have involved working with a variety of administrative data sources including: MSHA personnel data, WHD performance target data, Equal Employment Opportunity (EEO) Job-1 Form data, and enforcement data from OSHA, MSHA, WHD, OFCCP, OWCP, and EBSA. The methods for these projects have included descriptive and correlational statistical analysis, as well as statistical modeling. The projects generally resulted in deliverables such as memos, briefings, and analysis tools.

Among the specific projects Summit conducted for CEO include:

- Examining which firm characteristics and inspection process characteristics are significantly correlated with higher rates of firms' contesting Occupational Safety and Health Administration (OSHA) violations,
- Creating a model to assess the risk to the Mining Safety and Health Administration (MSHA) of retirements in its workforce,
- Developing a Microsoft Excel based tool to update, refine, and automate the process the Wage and Hour Division (WHD) uses for producing and aligning forecasts of national and district level performance targets,
- Developing and testing a new way for the Office of Federal Contracting Compliance Programs (OFCCP) to use data from the EEO Job-1 Form to better target Federal contractors that may have discriminatory pay practices, and
- Linking data from an internal list of large Federal contractors to publicly available enforcement data across DOL agencies to identify Federal contractors who have violated any Federal labor regulations or laws for the Office of the Assistant Secretary of Policy (OASP).

Select Staff

Albert J. Lee, Ph.D., Summit Founding Principal and Lead Economist

Albert J. Lee is Lead Economist and Founding Principal at Summit Consulting, LLC. Dr. Lee directs Summit's work of data integration, statistical analysis, program evaluation, and enforcement program support for DOL EBSA. He is an expert in econometric modeling and estimation, statistical sampling, and quantitative evaluation methods. He has more than 19 years of consulting experience leading program evaluation and applied statistical analysis engagements for state and Federal agencies. Dr. Lee received his Ph.D. in economics from UCLA.

Oswaldo Urdapilleta, Ph.D., Director

Dr. Urdapilleta directs Summit's labor evaluation, research, and data analytics work for Federal and state clients. He has over 20 years of experience as a labor researcher, health economist, and program evaluator. He has lead research and evaluation projects ranging from rigorous mixed methods impact evaluations, data analysis, compliance monitoring, surveys, and technical support for a variety of Federal, state, and non-governmental organization (NGO) clients. Dr. Urdapilleta has a Ph.D. in public policy from the University of North Carolina at Chapel Hill.

Summit's leadership brings years of experience:

- **Managing complex database development projects,**
- **Leading advanced data analytics studies, and**
- **Supporting DOL worker protection programs.**

China Layne, Ph.D., Manager, Data Analytics and Research

Dr. Layne manages the evaluation, research, and data analytics projects at DOL CEO and other DOL agencies. She has more than 5 years of experience leading evaluation, advanced statistical analysis, and research on a range of labor issues. She has specific expertise in using administrative data and large scale survey data for research and evaluation. Her experience focuses on worker protection programs, occupations, and regional labor markets. Dr. Layne has a Ph.D. in sociology from the State University of New York at Albany.

Yuwen Dai, Manager

Ms. Dai currently manages the Data Integration project for DOL EBSA. She has more than 5 years of experience leading various projects on business requirements collection and design, database development and system implementation, various analytical business analyses, model development and valuation, and fair lending compliance analysis. She has specific expertise in implementing advanced analyses and data systems in SAS, Stata, and Excel on large data sets that contained millions of data records. Ms. Dai holds an M.S. in applied statistics and from Ohio State University.