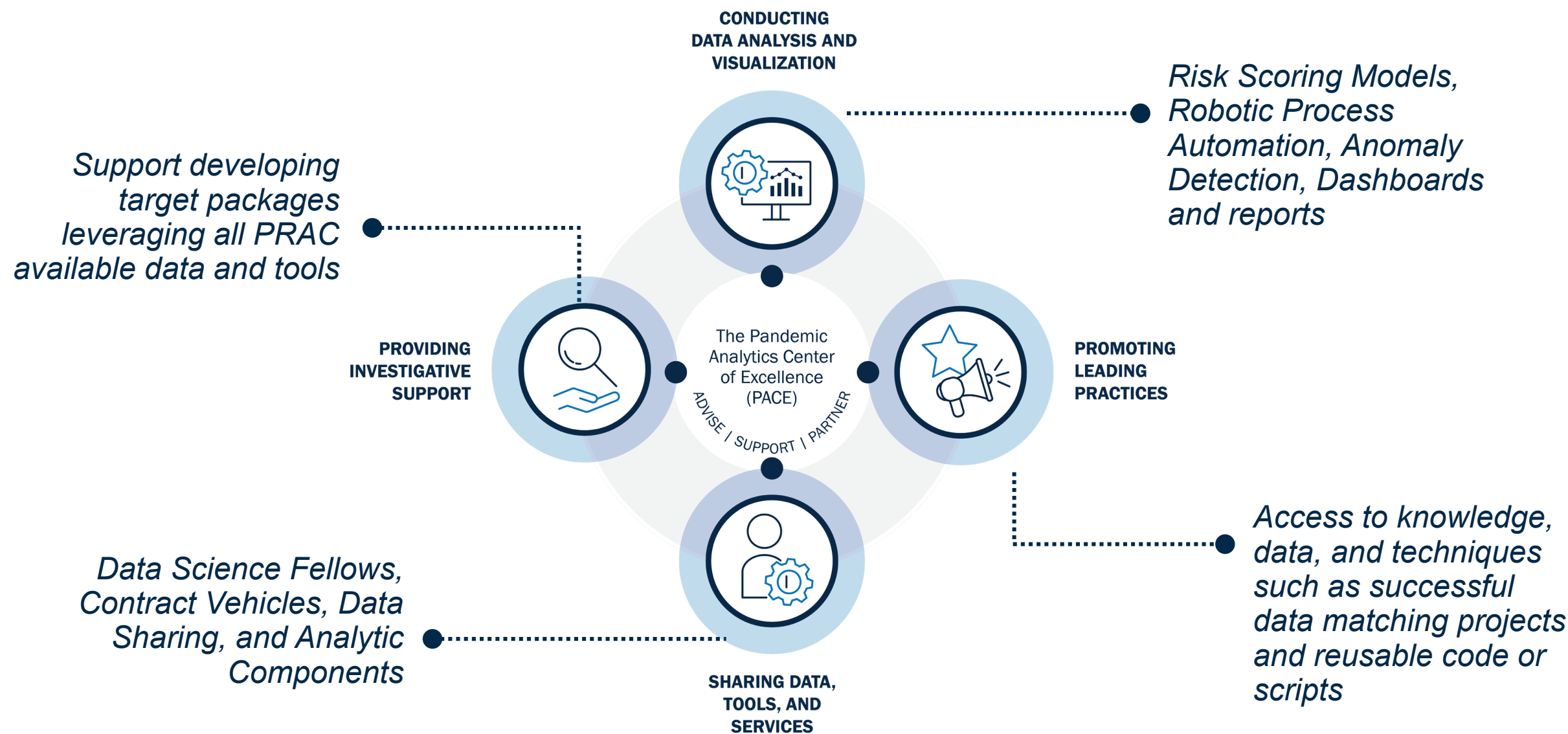# 24th Biennial Virtual Forum of Government Auditors
Pandemic Analytics Center of Excellence

Pandemic Response Accountability Committee (PRAC)
May 17, 2022

Pandemic Analytics Center of Excellence (PACE)

CONDUCTING DATA ANALYSIS AND VISUALIZATION

Risk Scoring Models, Robotic Process Automation, Anomaly Detection, Dashboards and reports

Support developing target packages leveraging all PRAC available data and tools

PROVIDING INVESTIGATIVE SUPPORT

The Pandemic Analytics Center of Excellence (PACE)

ADVISE | SUPPORT | PARTNER

PROMOTING LEADING PRACTICES

Access to knowledge, data, and techniques such as successful data matching projects and reusable code or scripts

Data Science Fellows, Contract Vehicles, Data Sharing, and Analytic Components

SHARING DATA, TOOLS, AND SERVICES

Examples of Current Projects

Hotline Risk Scoring Model

Develop a risk scoring model to help an OIG address the dramatic increase in fraud complaints submitted to the complaint tip line in relation to pandemic spending

Recipient Risk Scoring Model

Develop a risk scoring model to help an OIG better identify high risk Prime and sub-recipients for oversight activity
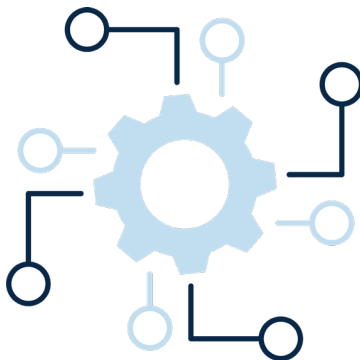
Robotic Process Automation

Utilize business logic and structured inputs to automate tasks associated with the Quarterly Report Reviews for a COVID-19 program

Social Network Analysis

Quickly build out networks using AI to reveal actionable insights on targets, groups and topic using social media and dark web data

Analytic Use Case - Hotline Risk Analytics

Hotline Risk Analytics Project Overview

Project Background

Built a risk scoring model for SBA OIG given the explosion (800 to 190k) of complaints related to SBA programs, as it was no longer feasible to manually review and triage complaints.

Business Challenges

- **Manual review** was impractical
- Most valuable data existed in narrative form
- Structured **data fields are missing or incomplete**
- Data is coming in from multiple channels

Innovation in Action

Ensemble Entity Resolution Techniques
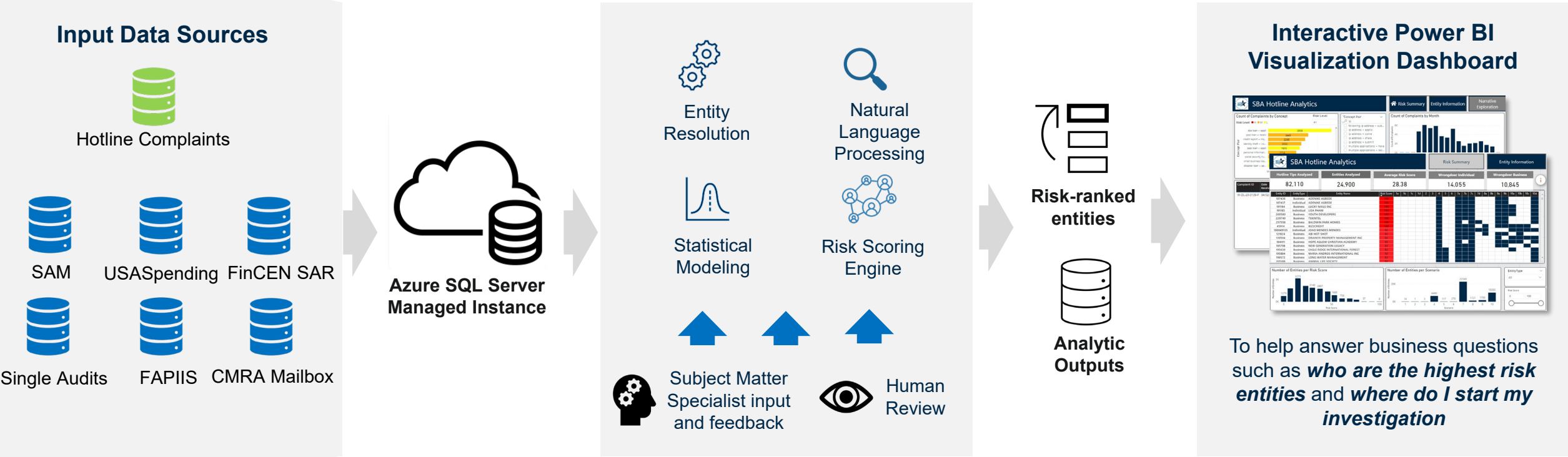
Concept Pair Extraction and Risk Review

Identity Theft Machine Learning Model

Solution Overview

Data Understanding and Preparation

Input Data Sources

Hotline Complaints

SAM   USASpending   FinCEN SAR

Single Audits   FAPIIS   CMRA Mailbox

Azure SQL Server Managed Instance

Analytic Modeling

Entity Resolution

Natural Language Processing

Statistical Modeling

Risk Scoring Engine

Subject Matter Specialist input and feedback

Human Review

Risk-ranked entities

Analytic Outputs

Insights Generation

Interactive Power BI Visualization Dashboard

To help answer business questions such as *who are the highest risk entities* and *where do I start my investigation*

Hotline Analytics Dashboard Demo

Collaborate to Elevate

Streamline OIG's intake processes
- Reduced manual review in **vetting of complaints**
- Accelerated the overall investigative process

Innovate hotline efficiencies across the IG community
- **Replicate the hotline analytics process** through joint projects with other OIG partners
- **Integrate with case management system** to further streamline the complaint review workflow

Share best practices & lessons learned
- Apply appropriate NLP techniques to help **elevate hotline complaint analytics capabilities** at other OIGs, such as risk concept extraction and ensemble entity resolution
- **Utilize self-identified fraud risk categories** to develop machine learning models to glean additional insights from unstructured data

Establish code-sharing procedure
- Code sharing with other OIGs can help accelerate the hotline complaint analysis:
  - Entity resolution code
  - Utility functions such as address standardization, entity name standardization, string similarity comparison, and string tokenization and resorting

Contact Information

Hong (Jo) Zhou

Hong.Zhou@cigie.gov

Sara Dendas

Sara.Dendas@cigie.gov

Thank you!