

From Flat Files to Repository: Lessons from Building a Data Warehouse

Stefanie Lopez-Howard, Statistical Analysis Center Director, Criminal Justice Coordinating Council

Samuel Gonzales, Research and Evaluation Program Director, Criminal Justice Coordinating Council



Georgia Data Innovation Hub
Carl Vinson Institute of Government
UNIVERSITY OF GEORGIA

From Flat Files to Repository: lessons from building a data warehouse

Stefanie Lopez -Howard,
Statistical Analysis Center
Director

Samuel Gonzales
Research and Evaluation
Program Director



Road Map

- I. About the Statistical Analysis Center
- II. Where we Started
- III. The Eight-Year Journey to Repository
 - I. Timeline
 - II. Fund Sources
 - III. Human Resources and Expertise
- IV. The Current State of Play
- V. A View of the Future

About the Statistical Analysis Center

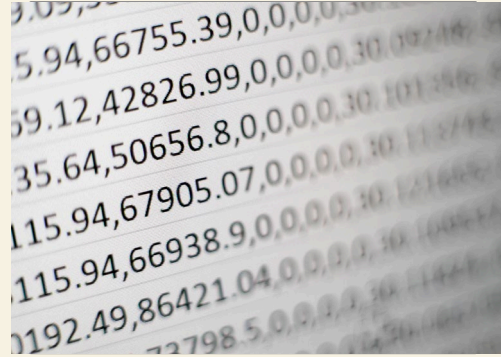


1. Established by executive order as a CJCC function in 1992
2. Supports all the data collection and analysis for the Justice and Compliance Division and the Victim Assistance Division
3. Fulfills CJCC's statutory authority to: serve as a criminal justice data clearinghouse; and provide policy relevant criminal justice research
4. Small but mighty team of 11 analysts and DBAs

Statistical Analysis Center Core Functions



Research and
Evaluation



Data Collection for
Grant Management
and Reporting



Business
Intelligence and
Analytics



Where we Started



Data Collection Sources

Interviewer Server
Administrator from IBM

Emailed Excel documents
from external partners



Data Extraction Method

Interviewer Server
Administrator data
download interface

Download and save
emailed files from
partners



Pain Points

Wide, unweildy extracted
data files that needed
reshaping

Multiple county and
agency name iterations



Nonstandardized Naming Conventions

Evolving and changing
variable names over time
made data cleaning
cumbersome

2015
Slalom first round of discovery to develop data sharing and information management strategy

2015-2016
Slalom builds data delivery and workflow portal in SharePoint for CJCC and Partners

2016-2017
Slalom provides CJCC with data repository architecture and BI tooling options discovery

2018
Grant for CJCC to develop Justice Reinvestment Initiative data portal

2018
CJCC contracts with Data Warehouse Architect and Business Analyst

2019-May 2022
Develop star schema, extraction queries, and ETL for repository

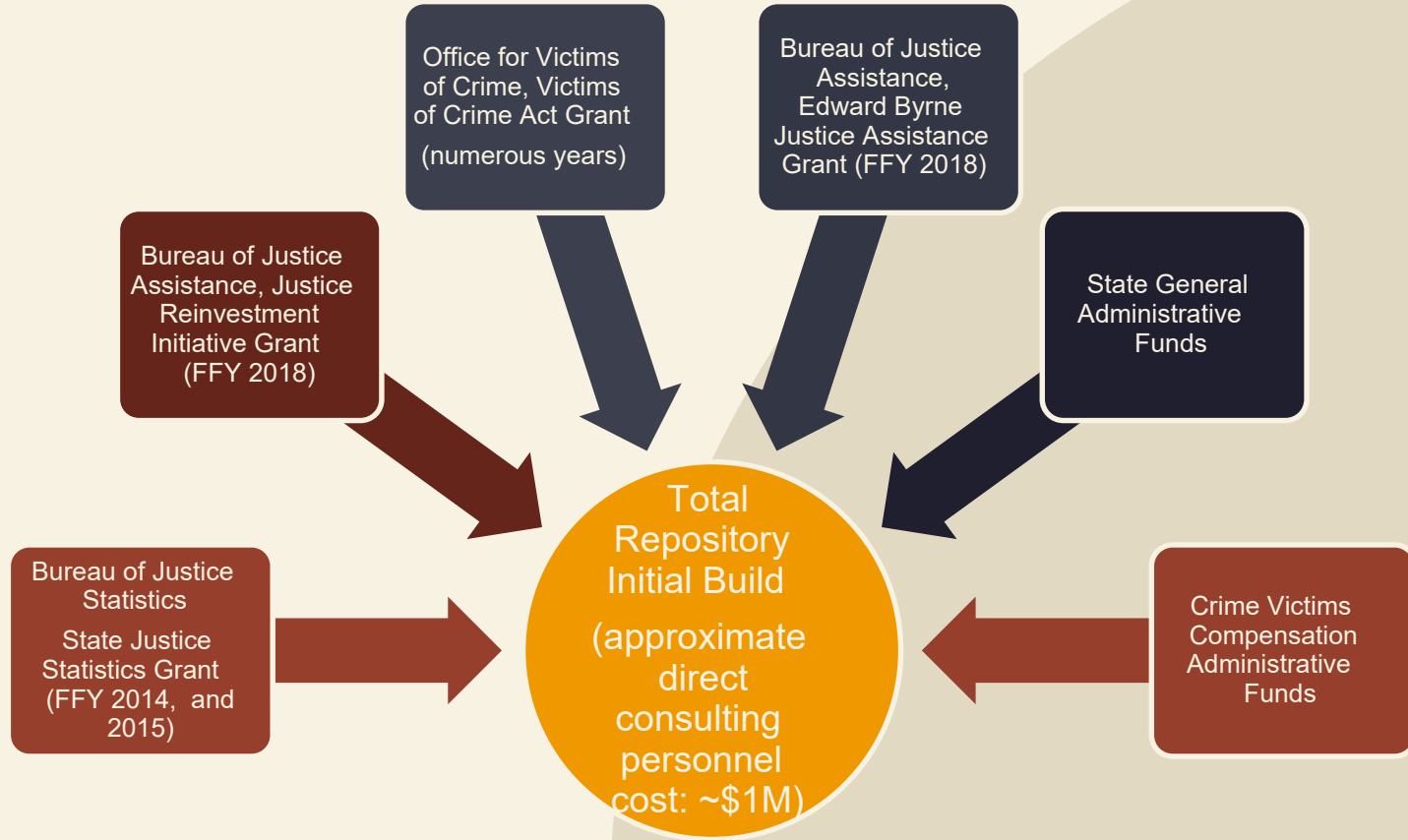
2021
Develop data sharing MOU for external agencies

May 2022 -Present
Load repository and validate data files

2023
Demo sample dashboards and develop process for public -facing portal

The Eight Year Journey to Repository

FundSources



Repository Development Personnel

01.

Independent Contract Business Analyst

Contracted with a business analyst through Covendis

02.

Slalom Consulting Teams

Business Analyst,
Datawarehouse architect,
SharePoint Solutions Architect

03.

Independent Contract Datawarehouse Architect

Contracted with data warehouse architect to develop schemas and data ingestion processes

04.

ETL Developer

Contracted with ETL Developer through Covendis to build stored procedures and Azure Data Factory workflows

FinalArchitecture

- Pipelines Built for **Nine** Different Datasets
- Extract data from CJCC-owned SQL servers, emailed flat files, Azure File Storage
- Data warehouse consists of two repositories
 - Repository for all fact and dimensions tables
 - Repository for views to do reporting
- Microsoft PowerBI Dashboard Interface
- Public vs. Internal Agency Data Portals

Final Repository Tool Set

Repository Database

- AzureSQL Managed Instance

ETL Tool

- Azure Data Factory
- SQL Stored Procedures

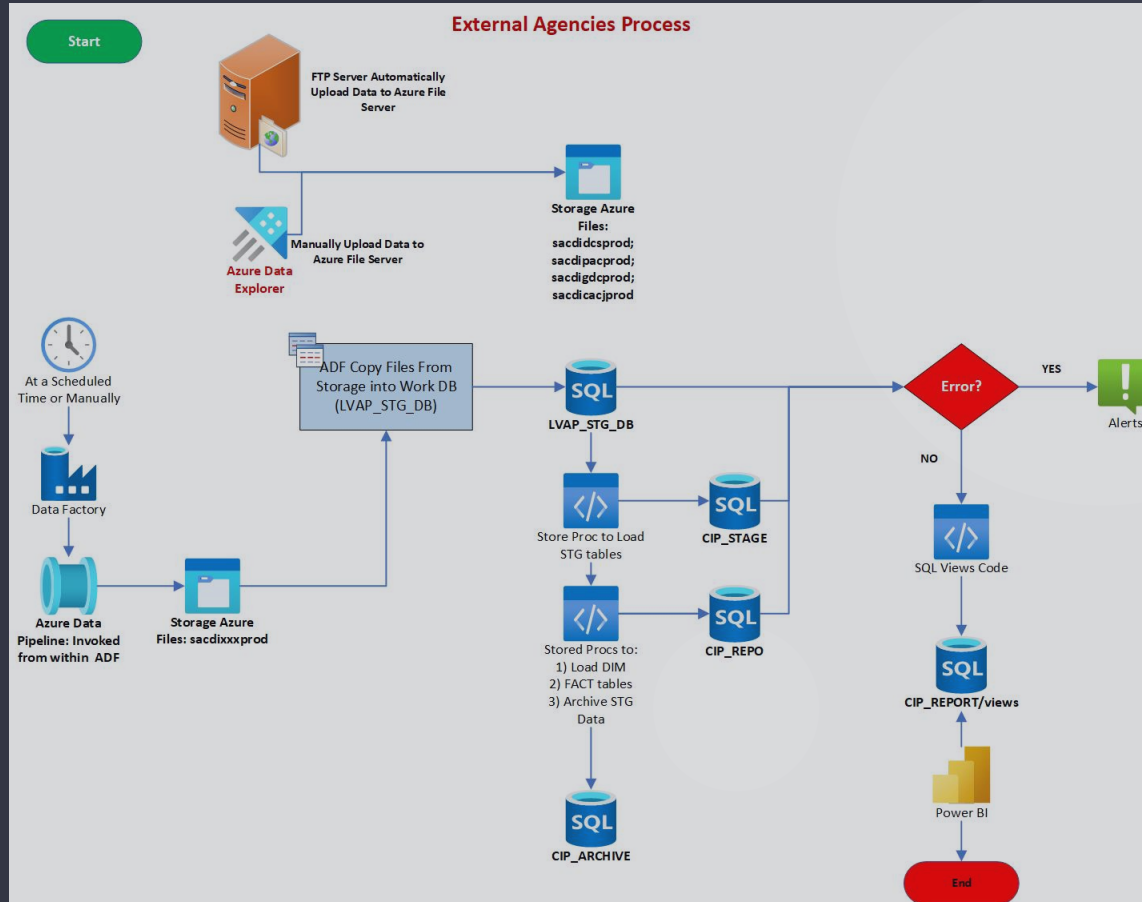
Partner File Staging

- Azure File Storage
- AzLogicApp extracts emailed files and delivers to Azure File Storage

Business Intelligence Tool

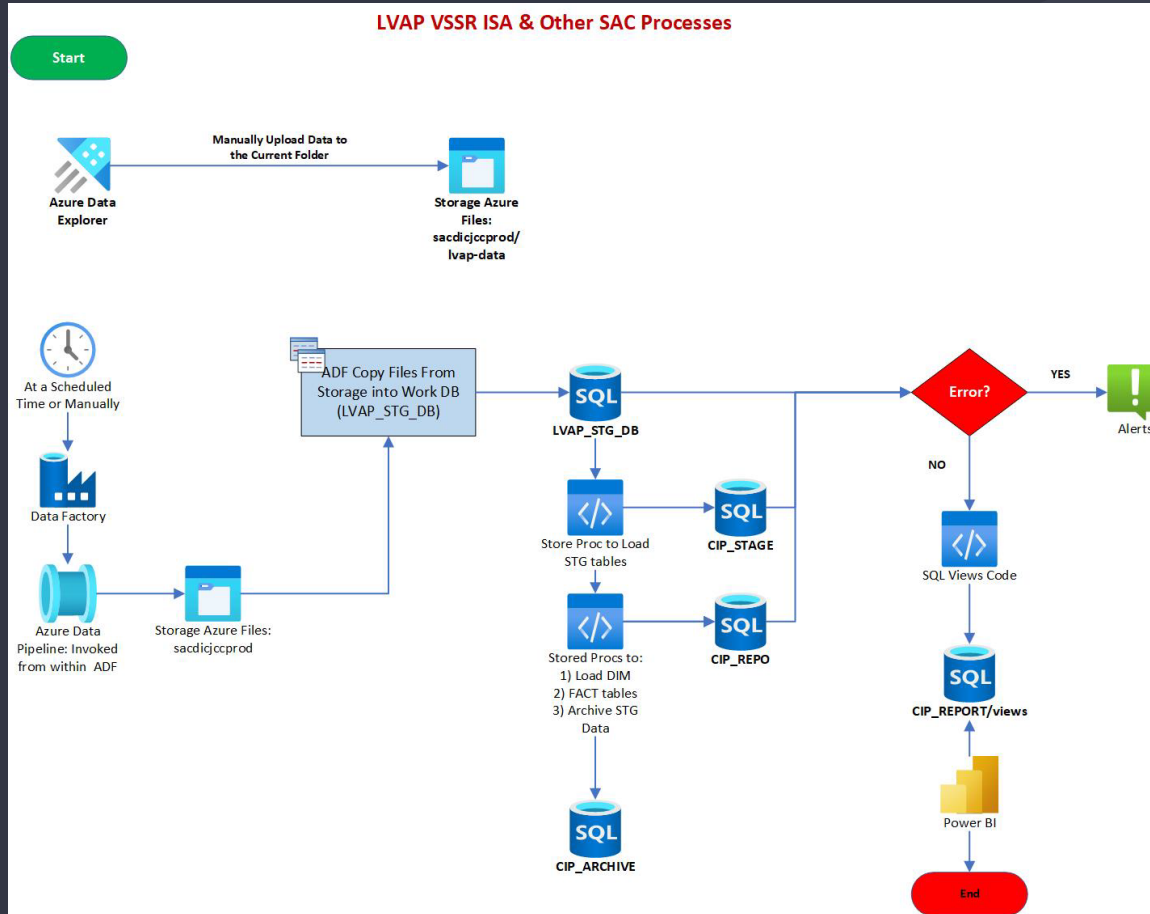
- Microsoft PowerBI

External Agency Upload Data Process

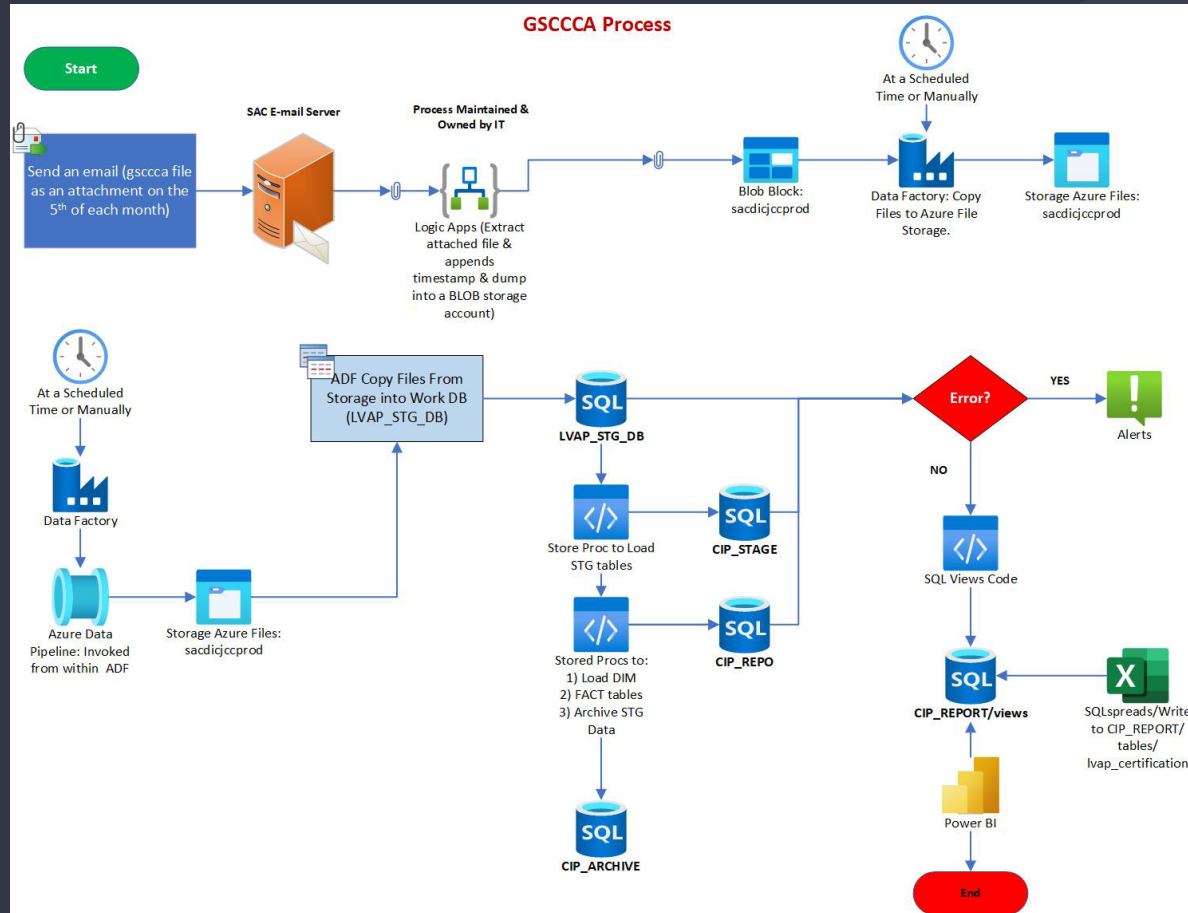


CJC-~~C~~Owned SQL-Server Data Process

LVAP VSSR ISA & Other SAC Processes



Emailed Flat File Data Process



Repository Development **Lessons Learned**



Analyst SME's need to be available to provide feedback on use cases and reporting needs



Development without a large team takes a long time



Mastering data is a labor intensive, quasi - manual process without data management software



Cost differential between Azure SQL Managed Instance vs. Azure SQL Server; cost of cloud computing, etc.

Use Case for the Data Repository for Operational Needs

Local Victims Assistance Program (LVAP or 5%)

- Is a 5% fine collected by courts to fund victim s services in their county

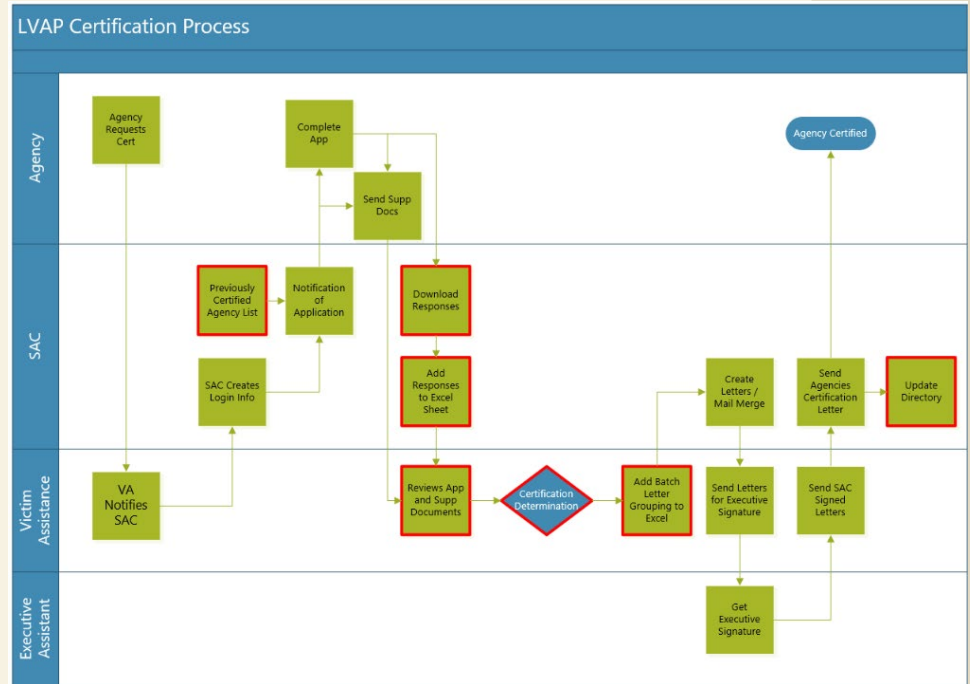
CJCC LVAP responsibilities outlined in statute:

- Certify victim s servie agencies for LVAP eligibility
- Monitor how much LVAP funding certified agencies receive and how many victims are serve with those funds
- Monitor the collection and distribution of LVAP funds by county adm inistrative agencies

We had a real issue with keeping up with certificaiton applications!

Using the Data Repository for Operational Needs

Through an analysis of the LVAP agency certification process it was clear that sharing certification application information and outcomes with our internal and external partners was extremely cumbersome.



Using the Data Repository for Operational Needs

Overview / Demo of LVAP certification data and documents

- LVAP ISA data extract
- LVAP agency document shared with colleagues
- LVAP certified agency guide
- SQL Spreads certification documentation
- LVAP agency certification visualization
- LVAP county non-reporters

Overview / Demo of future uses of the repository used for operational needs

- Human Trafficking Data Clearinghouse

THANKS!



Facebook.com
/gacjcc



instagram.c
om/ ga_cjcc



@gacjcc



Criminal Justice
Coordinating Council

Stefanie Lopez -Howard
slhoward@cjcc.ga.gov

Samuel Gonzales
sgonzales@cjcc.ga.gov

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**