



NC eLink: Completing the Picture to Improve Constituent Services

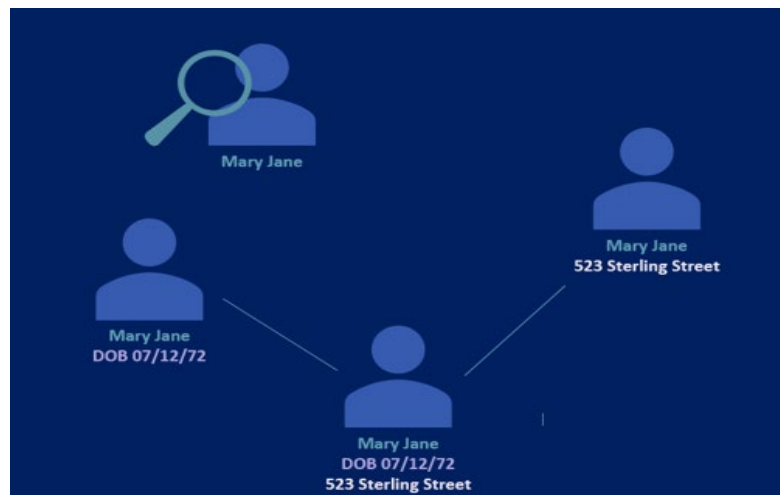
INITIATED 2020
COMPLETED 2022

Carol Burroughs
Director/Chief Data Officer
North Carolina Government Data Analytics Center
919-754-6921
carol.burroughs@nc.gov

EXECUTIVE SUMMARY

Government exists to serve its constituents, but if we do not know who we are serving and what they are receiving, we cannot determine if the services we are providing are effective. To ensure that North Carolina's constituents have access to all the services for which they are eligible and that the services are effective, the state needs to have a complete picture of an individual and the services that person currently receives.

Agencies collect data about the same individuals and entities in different ways and for different purposes. State agencies and departments collect data in administrative systems that are at various stages of maturity, making it difficult and time-intensive to share data across platforms. The North Carolina Government Data Analytics Center (GDAC), within the N.C. Department of Information Technology (NCDIT), works with agencies to provide the data connections across these administrative systems to support informed decisions and measure the effectiveness of government services. With appropriate data interoperability and entity resolution, the state can create a complete picture by matching records across multiple systems.



Working with agencies across North Carolina state government, GDAC developed NC eLink – an enterprise utility to link, store, and maintain unique entity IDs for an individual using data from multiple state and federal agencies. This utility is available to state agencies, offering a standardized linking process across systems and reducing the challenges and duplication of effort involved in building individual matching programs.

The initial proof-of-concept implementation of NC eLink involved a collaborative effort among four agencies relying on four administrative systems. Today, GDAC works with more than 10 agencies to integrate more than 50 administrative systems through the NC eLink utility. NC eLink has helped state agencies work together and share data to better serve our constituents. Currently, the NC eLink utility is being used to identify fraud, protect citizens, and connect longitudinal data that informs educational and workforce policy.

NC eLink presents countless opportunities for agencies to collaborate, sharing data to ensure that the state's constituents are protected from fraud and abuse and that the state's policies and services are effective.

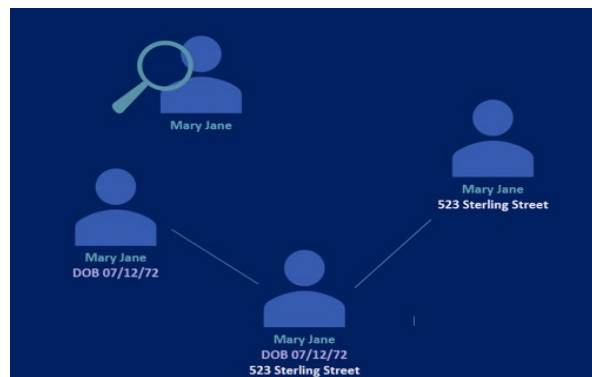
IDEA

Government exists to serve its constituents, but if we do not know who we are serving and what they are receiving, we cannot determine if the services we are providing are effective. These are questions we can answer with data that we currently collect – but in pieces and parts across multiple agencies and administrative systems, leading to duplicative records in some cases and missing information in others. Agencies come to the N.C. Government Data Analytics Center (GDAC), within the N.C. Department of Information Technology, for help with this analysis and to answer these questions.

To ensure that the state’s constituents have access to all the services for which they are eligible and that these services are effective, the state needs to have a complete picture of an individual and the services that person currently receives. This information exists in the vast amount of data that agencies collect specific to the programs and services they provide, but the picture is fragmented because no one system collects all the information or shares the same identifier to connect records. Agencies collect data about the same individuals and entities in different ways and for different purposes. State agencies and departments collect data in administrative systems that are at various stages of maturity, making it difficult and time-intensive to share data across platforms. GDAC works with agencies to provide data connections across these administrative systems to support informed decisions and measure the effectiveness of government services. With appropriate data interoperability and entity resolution, the state can create a complete picture by matching records across multiple systems.

Entity resolution is the process of connecting data points distributed across multiple systems and agencies into a trusted single view of entities. It creates a meaningful and holistic view of data across enterprises connecting with real-world people, places, and organizations. It logically groups records from different sources to build a complete picture of an individual that enables us to:

- Recognize when two records relate to the same identity despite being described differently
- Recognize when two records do not relate to the same identity despite being described similarly



Working with agencies across state government, GDAC developed NC eLink – an enterprise utility to link, store, and maintain unique entity IDs for an individual using data from multiple state and federal agencies. This utility is available to agencies, offering a standardized linking process across systems and reducing the challenges and duplication of effort involved in building individual matching programs.

GDAC has worked with agencies for many years to link data sources that did not have a common identifier. Prior to NC eLink, the matching process was replicated for each project, which was costly and did not always provide consistent outcomes. Based on these experiences with agencies, GDAC developed this standard utility product that helps link data sources easily and consistently with the creation of a unique entity ID. NC eLink provides participating agencies with a crosswalk table of entities and their associated data record identifiers to select, aggregate, and/or summarize data records associated with entities of interest.

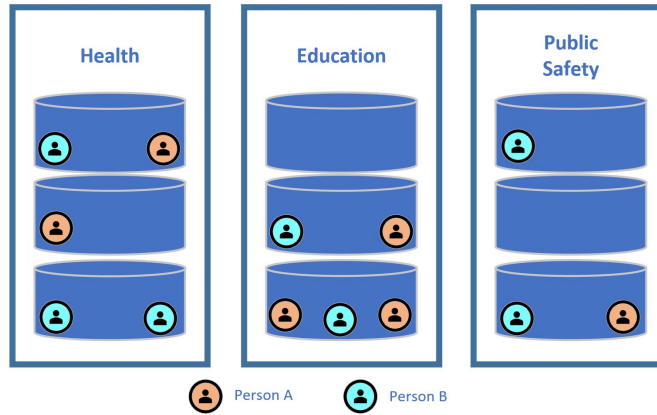
IMPLEMENTATION

NC eLink uses data from multiple state and federal agencies and allows the state to connect, store, and create unique entity IDs for an individual. NC eLink is a utility that integrates standardized information from agreed-upon GDAC data

sources for use in entity resolution and data matching. The initial proof-of-concept implementation was a collaborative effort among four agencies relying on four administrative systems. Today, GDAC works with more than 10 agencies to integrate more than 50 administrative systems through the NC eLink utility.

NC eLink ingests a subset of attributes that identify individuals from various administrative systems. Attributes include unique personal identifiers for each system, as well as details such as an individual’s name (first, middle, last), date of birth, race, gender, Social Security number, or driver’s license number. Personally identifiable information (PII) is used only for matching individuals across systems, without any service-related attributes. A crosswalk table with a master person ID for each unique individual is generated and used to link individual records across systems. This crosswalk enables the linking of individuals through the master person index.

Each blue drum in the image below is an administrative system. The same person is found across multiple administrative systems, sometimes with more than one ID in a single system.



The NC eLink utility helps match these records across systems and connect them to a single master person ID, as shown in the table below.

NC eLink Master Person Index Repository									
Master Person ID	Health system 1	Health system 2	Health system 3	Education system 1	Education system 2	Education system 3	Public safety system 1	Public safety system 2	Public safety system 3
000001	A	A			A	A1 A2			A
000002	B		B1 B2		B	B	B		B

The basic process for entity resolution through NC eLink is as follows:

1. Data preparation and data quality check identify issues and determine the cleansing and standardization.
2. Data linkage with entity resolution identifies a link between two sources, which might then link to other records.
3. Clustering applies linkage rules to logically link and group data to perform entity resolution.
4. Scoring looks at the density of clusters and the confidence in the matching of an individual entity.
5. Entity management includes identity management and lineage.

Data Preparation/Data Quality: Exploratory data analysis and data quality checks are performed on the agreed upon data to identify any issues and determine if or how data should be cleansed and standardized. The NC eLink system includes a suite of data cleansing and standardization processes through which data are cleansed, parsed, and validated. The table below outlines the operations that occur by data type.

Data Preparation Operations by Data Element Group

Data Element Group	Parsing	Cleansing	Standardization	Normalization	Match Code
Name	•	•			•
Demographic			•		
Personal Identification		•	•		
Date		•	•	•	
Address	•	•	•	•	•
Contact	•	•	•		

Some of the data cleansing and standardization operations carried out in this step include:

- Parsing names and standardization
- Removing special characters and punctuation from character variables
- Validating and standardizing personal identification numbers such as Social Security numbers into valid 9-digit values
- Eliminating redundant data and organizing data entries to ensure they appear across all fields and records through data normalization

Data Linkage with Entity Resolution: The next step in the process is to perform entity resolution, which logically groups records using personally identifiable attributes that uniquely identify an individual from different sources to build a more complete picture of an entity. These attributes include a personal identifier unique to the source system, along with certain individual specific attributes.

This step identifies element-based links between data sources. Not all sources have to have the same elements in common for matching. Through this process we can identify a link between two sources, which might then link to other records in other sources. NC eLink leverages more than 23 linkage rule sets to accommodate multiple independent entity-resolution processes. The linkages are joined into clusters of records, where tolerance and weighting build a confidence score that can be used to align with business needs.

Clustering: Clustering is a method by which data elements that share similar characteristics are combined to form groups or clusters. The system provides a default method for clustering linked records, which involves identifying connected components among linked records. The system also allows for an optional custom clustering methodology, providing flexibility in how to conduct clustering to produce a master person ID.

Scoring: Scoring takes place on two levels: the first assesses how tightly records are clustered, and the second assesses the validity of the matching for an entity. Cluster scoring provides a single measure for assessing the validity of a cluster and a way to compare clusters. This scoring helps to understand how well the clusters are linked together. The entity validity score expresses the confidence level of a match and is calculated at the cluster level using the collection of source records in the cluster. In other words, it considers what percentage of records have the most consistent individual key attributes, such as name, date of birth and Social Security number, to quantify variance. Which of these scores is most useful depends on the business use case and is weighted accordingly. The cluster score is used more often for aggregated views; validity scores are used more often when trying to confirm services received by an individual.

Entity Management: After the initial clustering is complete, an incremental update process is followed for any new records that are received. If the entity already exists in the master data sets, then the entity ID that already defines that individual will be applied to the new record, and the record will be moved to the master data sets. If the individual is not in the master data sets, then the record will go through the entity resolution process described above. The system tracks and maintains a complete audit history of the entity ID changes between runs.

IMPACT

The NC eLink utility has been used in a variety of scenarios, including criminal justice, fraud detection, and education.

Criminal Justice Activities

State Bureau of Investigation Sex Offender Registration Compliance – Sex offenders are required to register and verify their address and identity in person with a law enforcement every six months. North Carolina developed an alerting utility that identifies those individuals who have not verified their status with their local sheriff's office. North Carolina took the validation one step further and matched the National Sex Offender Registry across all states to identify individuals who may have moved into North Carolina and forgotten to register. If an out-of-state individual has a North Carolina presence within our administrative systems or is a known North Carolina sex offender and has not recently updated an address change, the model triggers an alert for law enforcement investigations.

Realized value: This program has been in place since 2021 and brought 595 sex offenders into compliance.

Fraud and Compliance Activities

To support fraud modeling, we first needed to understand a person's identity. To gain this knowledge, GDAC developed NC eLink to support the linking of a person across the state's administrative systems. This foundational component positioned us to understand certain patterns of an individual's identity across many systems.

With a knowledge of a person's identity, models have been developed to support the identification of anomalous behaviors and trigger alerts of suspected fraudulent activity.

N.C. Department of Commerce Unemployment Insurance Claimant Program Integrity – Fraud analytic models support unemployment claimant identity theft or claimant fraud and unemployment compliance alerting and overpayments. Compliance analysis ranges from simple eligibility verifications that match claims to deceased, incarceration, and payroll records to more complex algorithmic matching that identifies suspected identity theft and compliance and overpayment analysis.

During the COVID-19 pandemic, unemployment claims reached an unprecedented high. The N.C. Division of Employment Security, in partnership with GDAC, was able to pivot quickly and identify new fraud models on the existing platform. Risk scoring was developed to manage the high volume of alerts generated, allowing automated directing of claims with high scores to secondary automated checks and/or human intervention and review. Integration with the unemployment case management system set triggers to hold payments until claimant verification.

To manage the volume requiring manual review by unemployment investigators, suspicious claims were grouped together by rank and like attributes to improve efficiency and support adjudication of claimants' request. In 2022, using artificial intelligence (AI), predictive models were developed based on the patterns found to block claims in advance of initial processing. In 2023, this AI utility was made available via API interface to automate the identity verification during the claimant intake process.

Realized value: During 2022, the fraud modeling flagged 102,397 claimants for investigation, which resulted in alerting 60,305 claimants with suspicious activity associated with identity theft.

N.C. Department of Commerce Unemployment Insurance Business Tax Compliance – Business unemployment taxes collections support the payment of unemployment claims. With NC eLink as a foundation, we can identify misclassified employees, changes in business ownership/names made to avoid unemployment tax experience rating, and fictitious businesses created for the sole purpose of collecting unemployment benefits through fictitious claimants.

Realized value: Since January 2022, 294 businesses with suspected misclassified employees have been identified. The analytics also found fictitious businesses related to 832 fictitious claims for unemployment benefits associated with fictitious businesses in 2022.

N.C. Department of Revenue Tax Compliance and Identity Theft – Advanced analytics have also been applied to support the reduction of tax refund fraud, payment of fraudulent returns, and business tax compliance. Each year, new and improved models are developed based upon experience gained in analysis.

Realized value: Through advanced automation on the audit candidate selection process, the N.C. Department of Revenue estimates a personnel efficiency gain of \$5.3 million annually. Advanced analytic compliance modeling realized tax revenue assessments of \$16.4 million for fiscal year 21-22 and an additional \$11.8 million from July 1, 2022, to January 31, 2023.

Longitudinal Education Data Analysis

North Carolina Longitudinal Data System – North Carolina is at the forefront of a national movement to leverage longitudinal data systems to inform policy decisions. The N.C. Longitudinal Data System (NCLDS) is a "system of systems" that uses NC eLink to match data across early childhood, K-12, postsecondary education, and workforce from its partner agencies – N.C. Department of Commerce, N.C. Community College System, N.C. Department of Health and Human Services, N.C. Independent Colleges and Universities, N.C. Department of Public Instruction, and the University of North Carolina System. The NCLDS provides data contributors, practitioners, and trusted research partners with secure access to cross-sector, longitudinal, and linked record-level data to help them address some of the state's most pressing policy questions while maintaining the privacy

of individuals in the data sets. These linked records provide a better understanding of the opportunities and challenges that North Carolinians experience as they transition from early childhood, through the education system, and into the labor market.

Looking Forward

Utilities such as NC eLink present countless opportunities for agencies to collaborate and share data to ensure that the state's constituents have access to all the services for which they are eligible and that the services are effective. When agencies work with NC eLink, they begin to complete the picture of an individual and the services they currently receive.

With records consistently matched across agency systems, the state can establish a constituent engagement platform that provides a portal (single point of access) where state programs and services are aggregated and enable constituent user profiles – anonymous and authenticated (logged in utilizing a constituent identity and access management solution). Such a platform will make it easier for residents, businesses, and visitors to learn about, apply for, and consume services from the state and will significantly improve the constituent experience while consuming the state's digital services. With this platform, North Carolina's residents, businesses, and visitors will be able to:

- Establish an account and enter their information once, using one identity across all digital service offerings. The constituent will have control of their data, determining what data they want to share across agencies (i.e., the ability to opt-in/opt-out in sharing their data).
- Use [NC.gov](https://www.nc.gov) as a single point of entry to access state government services and programs. Constituents will no longer have to know where to go to receive different government services. Siloes of government will be broken down or eliminated.
- See all the programs and services a person is currently consuming (e.g., Supplemental Nutrition Assistance Program, Temporary Assistance for Needy Families, driver's license, fishing license), as well as programs and services that they may be eligible for based on their profile. For example, if a veteran is browsing for services, the portal will show applicable services. If a single mother who just lost her job is looking for services, the portal will show applicable services (e.g., unemployment claims, Special Supplemental Nutrition Program for Women, Infants, and Children).

As we connect records to make it easier for constituents to access services, we continue to develop and expand our data management practices that ensure the proper authorized use and access of data. Data sharing is guided by the principle of access to the minimum amount of data necessary to perform the task and, to the extent possible, ensuring deidentification or aggregation rules are applied. Utilities such as NC eLink enable data connection rather than compilation, which makes it easier for agencies to access only the pieces they need when they need them, rather than receiving bulk downloads of all an individual's data and PII to rematch locally.