

DATA-DRIVEN APPROACH TO IMPROVE EFFICIENCY AND EFFECTIVENESS OF SOCIAL PROGRAMS



Supplemental Nutrition Assistance Program (SNAP) and other social **programs** like Temporary Assistance for Needy Families (TANF) require data from multiple agencies (interfaces) to determine eligibility, verify evidence gathered during applications and re-certifications, and to deliver benefits electronically. For example, data from the Social Security Administration (SSA) and Department of Motor Vehicles (DMV) is used for verification of an individual who has filed an application to receive social program benefits. Similarly, SNAP's Electronic Benefits Transfer (EBT) is managed via pre-paid cards through a public company - Fidelity Information Services (FIS). When dealing with multiple data sources from

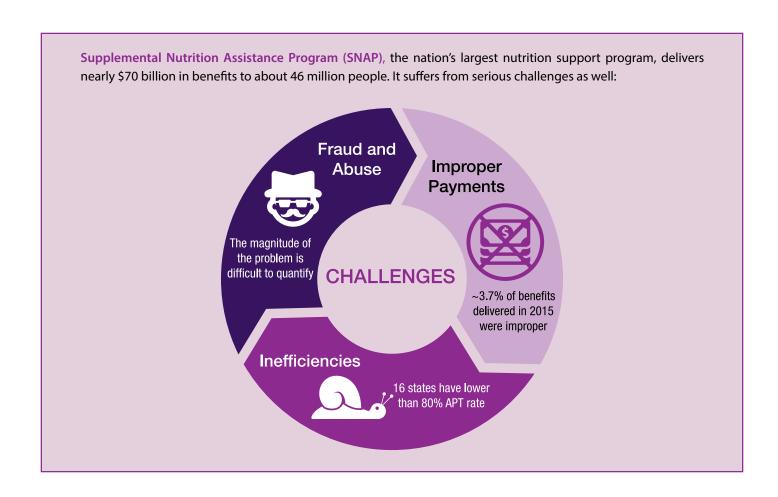
different agencies, a number of factors can impact operations and service delivery:

Accuracy of data – Using in-accurate data to determine eligibility and payments could place undue burdens on applicants and payments to recipients and lead to an increase in appeals workload, which in turn reduces programs' efficiency.

Timeliness of data – The timeliness of data can be an important factor in identifying and preventing improper payments. For example, the SNAP program could improve its ability to identify and prevent improper payments to claimants who have returned to work by obtaining wage data from sources other than quarterly wage data that it receives from employers.

Completeness of data – SNAP/TANF programs usually use a limited subset of the total elements/attributes of the data that is sent to them by an agency. While there are redundancies in some of the data elements that are captured by different agencies, using the entire dataset for verification could improve program integrity.

New data sources – SNAP recipients can traffic benefits by selling their EBT cards to others, exchanging the EBT card and the corresponding Personal Identification Number (PIN) for cash or non-food goods or services. This kind of fraudulent activity shows up on social media and e-commerce sites.



Challenges in managing and analyzing multi-source data

The biggest challenges in managing data from different sources are usually technical. Typical issues include:



Legacy systems that process verification and re-verification on a case-by-case basis delay an applicant's access to benefits.



The XML format for data transfer between agencies makes data retrieval expensive.



Transformation and storage of interface data elements, generally in a highly normalized form of BLOB and CLOB, into relational databases can be very expensive.

Re-think multi-source data ingestion and analysis with Infosys Information Platform

The Infosys Information Platform (IIP) solution can address these technical issues and improve efficiency and effectiveness of social programs like SNAP and TANF.



IIP has the ability to provide batch processing of data for verification / re-verification that could save operational and administrative resources used for case-by-case or manual database searches.



IIP is based on a Hadoop ecosystem which makes the data lake scalable from a storage perspective.



IIP can ingest data (structured, semi structured or unstructed, batch or real time) from various external sources such as database / data stream / File Servers / JMS queue through the data ingestion layer which is called the "Data Source."

Health and Human Services agencies can use IIP to drive their data to insights to action journey

Key benefits of IIP

IIP, based on open source big data frameworks such as Apache Spark, Sqoop, Kafka, Yarn, Hive, and Hadoop, addresses critical operational problems related to data. The platform has been developed on an industrial open source architecture and includes a set of unique adapters, tools and services.

IIP's design approach is to cleanly separate the open source components from the

capabilities built by Infosys and its partners.
This allows IIP to rapidly adapt to any
new capabilities and changes in the open
source layer. For example, if a better NoSQL
capability is available, or a better in-memory
distributed computing paradigm becomes
available, IIP can layer itself on top of that.

The IIP tool set abstracts the disparate open source technologies and provides an easy to use interface for ingesting data into the Hadoop/ Hive data lake. IIP also

includes a one click installer for automated deployments.

Since IIP leverages open source technologies, total cost of ownership is lower compared to licensed products. IIP utilizes commodity hardware to run big data solutions reducing infrastructure cost and meet the needs of on demand infrastructure. IIP also uses in-memory computing paradigm leading to impressive performance-to-price ratio.

*APT = # cases timely processed/#of cases subject to the timeliness measure;

A case is considered processed timely if the household has an opportunity to participate within seven days of the application date for expedited service cases and within 30 days of the application date for regular processing case.



Infosyspublicservices.com