

# Dirección de Desarrollo Digital

Unidad de Científicos de  
Datos



**El futuro  
es de todos**

**DNP**  
Departamento  
Nacional de Planeación



**El futuro  
es de todos**

**DNP**  
Departamento  
Nacional de Planeación

## **GESTIÓN DE BASES DE DATOS DE INGRESO SOLIDARIO**

### **Entidad**

Departamento Nacional de Planeación

- Dirección de Desarrollo Digital.
- Dirección de Desarrollo Social

### **Sector**

Planeación

### **Lenguaje**

Python

### **Fuente de datos**

SISBEN IV, Operadores móviles,  
gestores culturales, Censo  
Indígena, SIMAT

### **Presentación**

De acuerdo con las necesidades surgidas a partir de la llegada de la enfermedad respiratoria COVID-19 y las necesidades económicas y sociales de proveer de un sustento básico a población vulnerable y afectados económicos a las medidas de cuarentena para contener el virus, surgió la medida impulsada por el Estado de Ingreso Solidario para suplir dicha necesidad. Por tal motivo y con base en la información suministrada por la Dirección de Desarrollo Digital, surgió la necesidad de comunicar a aquellos beneficiarios no bancarizados que contaban con dicho beneficio por medio de mensaje de texto. De tal modo, fue necesario la consolidación de diferentes bases de datos, así como su revisión de calidad y gestión con el fin de enviar dicha información a los operadores móviles para poder enviar la notificación a los beneficiarios.

*In accordance with the needs that arose from the arrival of the respiratory disease COVID-19 and the economic and social needs to provide basic sustenance to the vulnerable population and those economically affected by the quarantine measures to contain the virus, the measure promoted by the State of Solidarity Income to meet this need arose. For this reason and based on the information provided by the Directorate of Digital Development, the need arose to communicate to those unbanked beneficiaries who had such benefit by text message. Thus, it was necessary to consolidate different databases, as well as to review their quality and management in order to send such information to mobile operators so that they could notify the beneficiaries.*

### **Objetivo general**

Consolidar y gestionar las bases de datos de no bancarizados suministradas por la Dirección de Desarrollo Social y por los operadores móviles y de acuerdo con sus instrucciones, con el fin de dividir y redireccionar las tablas resultantes a los operadores y poder ser enviada la notificación a los beneficiarios.

### **Objetivos específicos**

1. Consolidar tablas provenientes del cruce entre datos del DNP y los operadores móviles (en una etapa previa donde la UCD no intervino) con el fin de identificar los operadores móviles de los beneficiarios, así como su tecnología celular.
2. Identificar inconsistencias en los datos resultantes de cada cruce.
3. Para aquellos beneficiarios donde no se encontraron en ningún operador móvil, realizar el cruce con los datos de Gestores Culturales, Censo Indígena y SIMAT con el fin de abarcar la mayor población posible para la notificación.



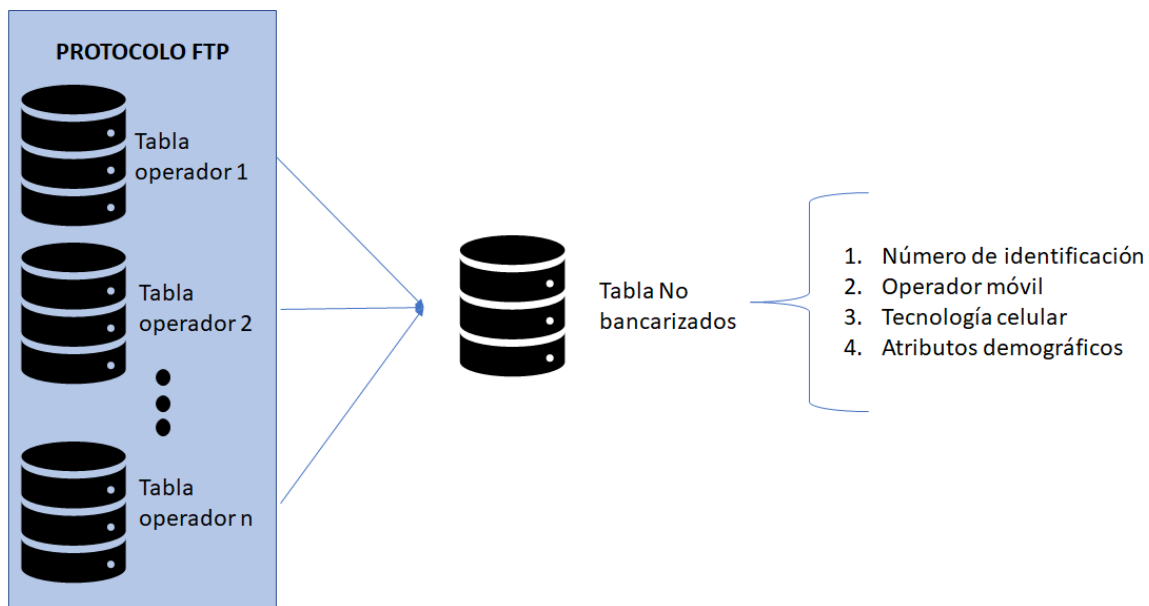
4. Redireccionar la información a los diferentes operadores móviles para enviar el mensaje a los beneficiarios.

### Metodología

Para el desarrollo de este proyecto, se realizaron los diferentes cruces de las tablas basándose en metodologías de álgebra relacional, identificando cada tabla y sus relaciones en el marco de crear una tabla consolidada que sea sencilla y fácil de utilizar para identificar a cada beneficiario, su operador móvil, tecnología celular y y/o teléfono de contacto.

Con base en un cruce previo hecho por los operadores móviles, se consolida la información en una sola tabla trayendo los datos de operador, tecnología y teléfono de contacto (este último se agrega si este es suministrado por el operador móvil). La información es compartida en un principio por el protocolo FTP disponible en el DNP en formatos txt, csv o xlsx, La Figura 1 muestra el flujo de datos para la primera consolidación de datos.

Figura 1 Mapa de consolidación de datos



Con base en la tabla resultante, se revisó la calidad de la información, por tal motivo se miraron, a partir de expresiones regulares, coincidencias poco frecuentes en el número de documento (como lo es la repetición de un mismo número de manera consecutiva en un mismo documento de identidad, números consecutivos en el mismo documento de identidad, documentos de identidad muy largos o cortos y números de documento repetidos).

Con base en la tabla consolidada y de acuerdo con una nueva tabla enviada por la Dirección de Desarrollo Social y luego de una validación de calidad hecha por la Registraduría Nacional del Estado Civil, se consolidó una nueva tabla que contiene la información de las TELCOS e información de contacto proveniente de Gestores

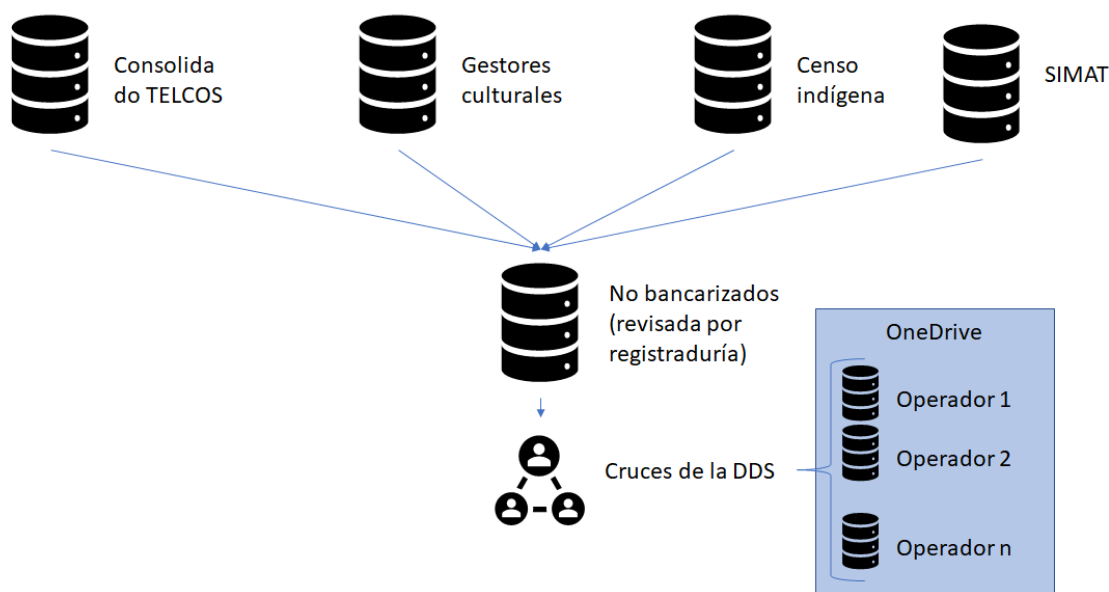


# El futuro es de todos

DNP  
Departamento  
Nacional de Planeación

Culturales, Censo de Indígenas y SIMAT para luego ser remitida a la Dirección de Desarrollo Social y hacer un nuevo cruce con el canal de desembolso del beneficio. La Figura 2 muestra el flujo de datos desde la consolidación de TELCOS (etapa anterior) hasta el redireccionamiento de los datos de nuevo a las TELCOS.

Figura 2 Flujo de datos luego de revisión de registraduría



A partir de los datos consolidados de TELCOS nuevas fuentes de datos (Gestores culturales, Censo indígena y SIMAT) se consolidó la información con base en la tabla enviada por la Registraduría teniendo en cuenta los siguientes criterios.

1. Los JOIN se hicieron a partir del número de documento
2. Se identificaron inconsistencias en el cruce a partir de la similitud en los nombres de cada tabla y se les asignó un score de similitud (dicha similitud se hizo a partir de la distancia *levenshtein*. Esta similitud se hizo en cada tabla teniendo como referencia la base revisada por la Registraduría.
3. Se identificaron inconsistencias en los números de documento de identidad de cada base (tales como números repetidos y consecutivos dentro de un mismo número de documento, documentos repetidos y longitud del documento).

Los datos luego de dicha consolidación y validación fueron enviados a la DDS para agregar información correspondiente al canal bancario para el desembolso. Una vez cruzada la tabla, esta fue remitida a la UCD ya dividida por operador para ser compartida a los operadores móviles (a través de una carpeta compartida de OneDrive de acuerdo con las recomendaciones de la OTSI) para finalmente sea enviado el mensaje de texto a los beneficiarios con la notificación de que cuentan con dicho beneficio.



**El futuro  
es de todos**

**DNP**  
Departamento  
Nacional de Planeación

### **Conclusiones y recomendaciones**

1. Se consolidaron los datos provenientes de los cruces de las TELCOS con información suministrada por el DNP
2. Se identificaron inconsistencias a partir de métricas de análisis de texto y expresiones regulares.
3. Se consolidó una única tabla con base en información proveniente de externos (Gestores culturales, Censo Indígena y SIMAT)
4. Se redireccionó las tablas resultantes a la DDS para luego ser enviadas a los operadores móviles y enviar los mensajes de texto a los beneficiarios
5. Se recomienda tener un sistema de flujo de información e intercambio de datos más robusto para futuros proyectos de dicha naturaleza.

### **Socialización**

El proyecto se ha socializado con la Dirección de Desarrollo Social