

 <p>El futuro es de todos</p> <p>Unidad para la atención y reparación integral a las víctimas</p>	<b>GUÍA PARA LA ESTANDARIZACIÓN Y MIGRACIÓN DE BASES DE DATOS</b>		Código: 520,06,15-1
	PROCESO: GESTIÓN DE LA INFORMACIÓN		Versión: 3
	PROCEDIMIENTO: INSTRUMENTALIZACIÓN DE LA INFORMACIÓN		Fecha: 15/10/2021 Página <b>1</b> de <b>7</b>

## 1. OBJETIVO:

Dar un marco de referencia y estandarizar las actividades necesarias para llevar a cabo la disposición de fuentes.

## 2. ALCANCE:

El proceso inicia con la recepción de las fuentes de información al interior de la SRNI y termina con la disposición de esta para su consulta.

## 3. DEFINICIONES:

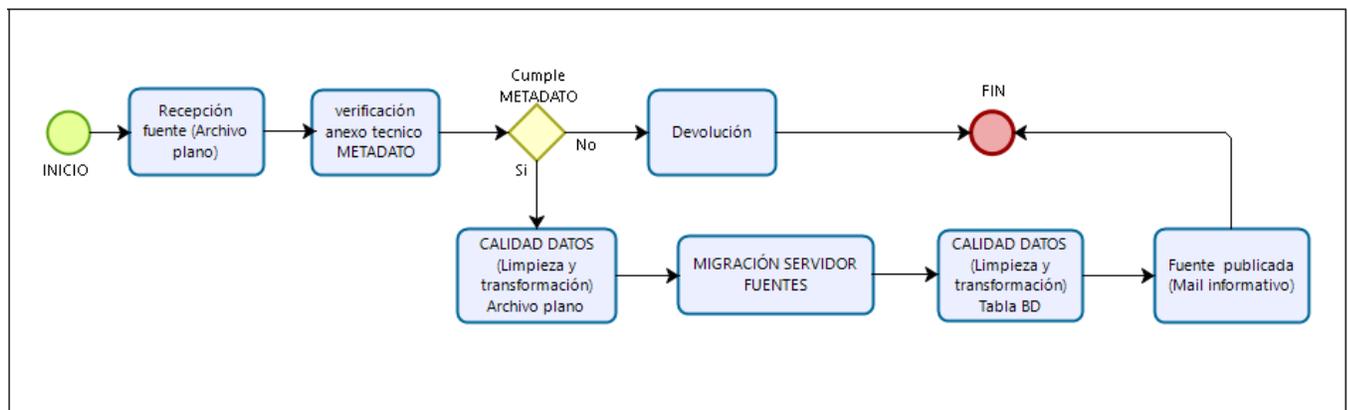
**METADATO:** Caracterización de una fuente de datos, que describe el contenido de variables dispuestas en la salida de datos proporcionada.

**ESTANDARIZACIÓN:** Que los textos sean textos, los valores numéricos o fecha correspondan.

**HOMOLOGACIÓN:** transformación del valor contenido en un campo o columna a valores estándar con el fin de unificar las salidas de los servicios de consulta P.E. Los nombres concatenados en un campo se transforman a nombre1, nombre2 apellido1 y apellido2.

## 4. FLUJO DEL PROCESO:

El flujo de trabajo está enmarcado según el siguiente esquema:



## 5. ACTIVIDADES:

Generalidades: Este documento describe las acciones generadas para disposición de información de los datos que recibe la Subdirección Red Nacional de Información, a partir

 <b>El futuro es de todos</b> Unidad para la atención y reparación integral a las víctimas	<b>GUÍA PARA LA ESTANDARIZACIÓN Y MIGRACIÓN DE BASES DE DATOS</b>		Código: 520,06,15-1
	PROCESO: GESTIÓN DE LA INFORMACIÓN		Versión: 3
	PROCEDIMIENTO: INSTRUMENTALIZACIÓN DE LA INFORMACIÓN		Fecha: 15/10/2021 Página <b>2</b> de <b>7</b>

de los acuerdos de intercambio de información realizados con entidades que conforman el SNARIV.

### **5.1 Inscripción fuentes de intercambio**

Las fuentes de intercambio son documentadas de acuerdo con los anexos técnicos establecidos con las entidades que proporcionaran periódicamente la información. En la herramienta de inventarios se documenta el origen de la información y su metadato respectivo.

### **5.2 Recepción de fuentes**

Una vez descrito el metadato de la fuente y de acuerdo con la periodicidad establecida para la recepción de la misma, los colaboradores designados del grupo de Articulación y Fortalecimiento o del grupo de Sistema de Información geográfica, encargados de la gestión del intercambio de la información con las entidades, diligencian en la herramienta de inventarios el archivo plano de la fuente, a partir de allí se generan una serie de tareas para disponerlas en la Base de Datos Esquema de Fuentes, el cual es un insumo para los sistemas de información y aplicaciones que requieran disponer o consultar la información.

La información del archivo plano y el metadato.

### **5.3 Verificación del Anexo Técnico METADATO**

Consiste en comparar que la información dispuesta en el archivo corresponda con la descripción previa realizada en la herramienta de inventarios de su metadato.

### **5.4 Migración al servidor de fuentes**

Una vez la data ha sido validada y organizada, se procede con la migración, la cual consiste en preparar la Tabla en el esquema de BD que contendrá la información dispuesta en el archivo proporcionado por la entidad, generar los archivos de comando y control correspondientes para la migración del archivo.

- Nombre de la tabla en esquema de Fuentes:

El nombre de las tablas en el esquema fuentes busca estandarizar, organizar y clasificar cada una de las fuentes de información que ingresan a la Red Nacional de Información.

Descripción nombre de tabla

- a) El nombre de la tabla debe iniciar con el prefijo "DB"
- b) El nombre de la entidad en un campo no mayor a 4 caracteres

 <b>El futuro es de todos</b> Unidad para la atención y reparación integral a las víctimas	<b>GUÍA PARA LA ESTANDARIZACIÓN Y MIGRACIÓN DE BASES DE DATOS</b>		Código: 520,06,15-1
	PROCESO: GESTIÓN DE LA INFORMACIÓN		Versión: 3
	PROCEDIMIENTO: INSTRUMENTALIZACIÓN DE LA INFORMACIÓN		Fecha: 15/10/2021 Página <b>3</b> de <b>7</b>

- c) El nombre de la fuente debe ser un campo no mayor a 5 caracteres
- d) La definición del metadato debe ser un campo no mayor a 4 caracteres numéricos
- e) La fecha del corte debe ser no mayor a 8 caracteres donde se debe mantener estructura de fecha DDMMYYYY

Nota: la notación recomendada para La definición del nombre de la tabla es "Snake Case" la cual debe estar separado por un "\_" y no debe contener espacios en su estructura

Ejemplo:

PREFIJO INICIAL	NOMBRE DE LA ENTIDAD	NOMBRE DE LA FUENTE	ID META DATO	FECHA DE CORTE	NOMBRE DE TABLA
DB	UNIDAD ADMINISTRATIVA ESPECIAL DE ATENCIÓN Y REPARACIÓN INTEGRAL A LAS VÍCTIMAS	SUBDIRECCIÓN DE PREVENCIÓN Y ATENCIÓN DE EMERGENCIAS SPAE	4309	31/10/2018	DB_UAEA_SPAES_4309_31102018

## 5.5 Alistamiento de textos planos

Posterior a la definición de los nombres de la tabla se procederá a realizar un alistamiento del texto plano, de tal manera que se ajuste a los parámetros definidos en el diccionario de datos (limpiar saltos de línea, tabulación, espacios dobles, definición del tipo de dato, el largo del campo, definir encabezado de la columna).

Para archivos de menos de un millón de registros se puede usar las siguientes funciones en Excel para agilizar el proceso de depuración.

- Función calcular el tamaño del campo:

```
Function textoMasLargo (Rango As Range) As Integer
Dim curcell As Object
Dim largo As Integer
largo = 0
On Error Resume Next
For Each curcell In Rango
If Len(curcell) > largo Then largo = Len(curcell)
Next
textoMasLargo = largo
End Function
```

- Macro para limpiar tabulación en la celda dentro de un mismo rango.

 <b>El futuro es de todos</b> Unidad para la atención y reparación integral a las víctimas	<b>GUÍA PARA LA ESTANDARIZACIÓN Y MIGRACIÓN DE BASES DE DATOS</b>		Código: 520,06,15-1
	PROCESO: GESTIÓN DE LA INFORMACIÓN		Versión: 3
	PROCEDIMIENTO: INSTRUMENTALIZACIÓN DE LA INFORMACIÓN		Fecha: 15/10/2021 Página <b>4</b> de <b>7</b>

```

Sub LimpiarBase()
Dim TheCell As Range
On Error Resume Next
For Each TheCell In ActiveSheet.UsedRange
With TheCell
If .HasFormula = False Then
.Value = Application.WorksheetFunction.Clean(.Value)
End If
End With
Next TheCell
End Sub

```

## 5.6 Cargue de data al esquema

A continuación, se describe brevemente el proceso de cargue de información que llega mediante textos planos por algún medio digital (DVD, FTP, USB) el cual se registra en la herramienta de inventarios, para luego subirse en un motor de base de datos, a través de un utilitario de inserción de datos.

- Verificar que la estructura del diccionario de datos concuerde con el texto plano: Se realiza una comparación entre la cabecera del texto plano con el diccionario de datos, en caso de que no corresponda se notifica a través de la herramienta de inventarios y se devuelve al equipo de gestión.
- Definición de los nombres de las columnas: Se realiza la definición de cada columna teniendo en cuenta lo siguiente:
  - a. No debe tener más de 30 caracteres.
  - b. No debe tener ningún carácter especial ni espacios en su estructura.
  - c. Si desea usar separadores se recomienda usar “\_”
- Creación de la tabla en el esquema de fuentes:

```

CREATE TABLE DB_DPS_FEST_XX_31032018
(
  FIX_CONS_UARIV NUMBER,
  PER_ID NUMBER,
  ANNO_INF NUMBER,
  CODIGO_DANE NUMBER,
  TIPO_DOCUMENTO NUMBER,
  NRO_DOCUMENTO VARCHAR2(200),
  APELLIDO1 VARCHAR2(200),
  APELLIDO2 VARCHAR2(200),
  NOMBRE1 VARCHAR2(200),
  NOMBRE2 VARCHAR2(200),
  DIRECCION_RESIDENCIA VARCHAR2(200),
  TEL VARCHAR2(200),
  RES_DEPTO NUMBER,
  RES_MUN NUMBER
)

```

## 5.7 Configuración y definición del fichero de comando cmd\_mig.cmd

 <b>El futuro es de todos</b> Unidad para la atención y reparación integral a las víctimas	<b>GUÍA PARA LA ESTANDARIZACIÓN Y MIGRACIÓN DE BASES DE DATOS</b>		Código: 520,06,15-1
	PROCESO: GESTIÓN DE LA INFORMACIÓN		Versión: 3
	PROCEDIMIENTO: INSTRUMENTALIZACIÓN DE LA INFORMACIÓN		Fecha: 15/10/2021 Página <b>5</b> de <b>7</b>

- Creación del fichero Comando cmd\_mig.cmd

```
sqlldr      userid=usuario/clave@fuentes      CONTROL=ctl_mig.ctl      skip=1
SILENT=(FEEDBACK)pause
```

- Definición de parámetros archivo Control.ctl

```
log='log.txt' errors=n SILENT=(FEEDBACK)
```

Userid	Especifica la cadena de conexión al servidor, incluye usuario y clave
Control	Indica la ubicación del archivo de control creado
Log	Ubicación donde se guardará el archivo de registro del cargue
errors	(Opcional) Número máximo de errores permitidos antes que se aborte el cargue. Si no se especifica toma como valor por defecto 50
SILENT=(FEEDBACK)	(Opcional, pero recomendada) Con esta opción no se muestran los mensajes cada vez que la herramienta hace un commit, acelerando el proceso

## 5.8 Definición del archivo control.ctl

**Color amarillo:** En este campo se diligencia el nombre del archivo plano que se va a migrar.

**Color Rojo:** En este campo se diligencia el nombre de la tabla que recibirá la información contenida en el archivo plano.

**Color Verde:** En este campo se define el separador de campos que tiene el archivo plano.

**Color Azul:** En este campo se definen los campos de la tabla con sus respectivas propiedades de carga.

```
OPTIONS (DIRECT=TRUE, ERRORS=300, rows=500)
UNRECOVERABLE load data CHARACTERSET WE8ISO8859P1
```

 <p>El futuro es de todos</p> <p>Unidad para la atención y reparación integral a las víctimas</p>	<b>GUÍA PARA LA ESTANDARIZACIÓN Y MIGRACIÓN DE BASES DE DATOS</b>		Código: 520,06,15-1
	PROCESO: GESTIÓN DE LA INFORMACIÓN		Versión: 3
	PROCEDIMIENTO: INSTRUMENTALIZACIÓN DE LA INFORMACIÓN		Fecha: 15/10/2021 Página <b>6</b> de <b>7</b>

```
infile 'TEXTO_PLANO.csv' "str '\r\n'"
append
into table DB_DPS_FEST_XX_31032018
fields terminated by ';'
OPTIONALLY ENCLOSED BY '"' AND '"'
trailing nullcols
(
FIX_CONS_UARIV sequence(MAX, 1),
PER_ID "REGEXP_SUBSTR(:PER_ID, '[0-9]+')",
ANNO_INF "REGEXP_SUBSTR(:ANNO_INF, '[0-9]+')",
CODIGO_DANE "REGEXP_SUBSTR(:CODIGO_DANE, '[0-9]+')",
TIPO_DOCUMENTO "REGEXP_SUBSTR(:TIPO_DOCUMENTO, '[0-9]+')",
NRO_DOCUMENTO "UPPER(:NRO_DOCUMENTO)",
APELLIDO1 "UPPER(:APELLIDO1)",
APELLIDO2 "UPPER(:APELLIDO2)",
NOMBRE1 "UPPER(:NOMBRE1)",
NOMBRE2 "UPPER(:NOMBRE2)",
DIRECCION_RESIDENCIA "UPPER(:DIRECCION_RESIDENCIA)",
TEL "UPPER(:TEL)",
RES_DEPTO "REGEXP_SUBSTR(:RES_DEPTO, '[0-9]+')"
)
```

Se procede a ejecutar el fichero "cmd\_mig.cmd" para iniciar el proceso de cargue del texto plano al esquema de fuentes.

Nota: En caso de que se genere un archivo "nombre\_de\_archivo.bat" se debe revisar, corregir y volver a cargar los datos.

## 5.9 PROCESO DE CALIDAD

Dispuesta la información en la tabla del esquema Fuentes, se procede a realizar una revisión de datos dispuestos

- **Transformación** Llevar a estándares de migración la data proporcionada. Se procede con la conversión de datos:
  - **Tipos de Documento:** Se conserva la descripción original contenida en la BD, se realiza conversión de minúsculas a mayúsculas
  - **Documento:** el campo debe ser de tipo texto, sin ningún signo de puntuación.
  - **Nombres y Apellidos:** deben estar en mayúscula, separados en primer nombre, segundo nombre, primer apellido y segundo apellido.
  - **Fechas:** deben ser de tipo fecha y seguir el formato "DD/MM/YYYY"
  - **Valores monetarios:** deben ser de tipo numérico.
  - **Etnias:** Se conserva la descripción original contenida en la BD

 <b>El futuro es de todos</b> Unidad para la atención y reparación integral a las víctimas	<b>GUÍA PARA LA ESTANDARIZACIÓN Y MIGRACIÓN DE BASES DE DATOS</b>		Código: 520,06,15-1
	PROCESO: GESTIÓN DE LA INFORMACIÓN		Versión: 3
	PROCEDIMIENTO: INSTRUMENTALIZACIÓN DE LA INFORMACIÓN		Fecha: 15/10/2021 Página <b>7</b> de <b>7</b>

## 5.10 FUENTE PUBLICADA

Cuando el proceso de migración y limpieza culmina, se procede a cambiar el estado en la herramienta de inventarios y se informa al grupo de trabajo de la RNI la disponibilidad de la fuente.

## 6. CONTROL DE CAMBIOS (Cambios realizados al procedimiento)

Versión	Fecha	Descripción de la modificación
1	19/08/2014	Creación del documento.
2	04/09/2015	Se definió el alcance del documento y el glosario del documento.
3	15/10/2021	Actualización a la guía con la descripción y nombre del procedimiento.