**SOFTWARE INTERACTIVO PARA LA INTEGRACIÓN AUTOMÁTICA DE BASES DE DATOS MÉDICAS, CASO DE ESTUDIO: ENCUESTA NACIONAL DE EXAMEN DE SALUD Y NUTRICIÓN (NHANES)**

**Autor**

**Jorge Miguel Anaya León**

****

**Director**

**Msc. José Gerardo Chacón Rangel**

**Ingeniería De Sistemas**

**Departamento De Ingenierías Eléctrica, Electrónica, Sistemas Y Telecomunicaciones**

**Facultad De Ingenierías Y Arquitectura**

**Universidad De Pamplona**

**Villa Del Rosario, 2019**

**Universidad de Pamplona  
Facultad de Ingenierías y Arquitectura  
Departamento de Ingenierías Eléctrica, Electrónica, Sistemas y Telecomunicaciones Ingeniería de Sistemas**

**Trabajo presentado para optar por el título de Ingeniero de Sistemas**

**Título:**

**SOFTWARE INTERACTIVO PARA LA INTEGRACIÓN AUTOMÁTICA DE BASES DE DATOS MÉDICAS, CASO DE ESTUDIO: ENCUESTA NACIONAL DE EXAMEN DE SALUD Y NUTRICIÓN (NHANES)**

**Fecha de inicio del trabajo: Agosto de 2018**

**Fecha de terminación del trabajo: Diciembre de 2018**

**Nombres y firmas de autorización para sustentar:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Jorge Miguel Anaya León**

**Autor**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
Msc. José Gerardo Chacón Rangel**

**Director**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
MSc. Richard Eliseo Mendoza Gafaro**

**Director del programa**

**Jurado calificador:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
 Jurado 1**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
 Jurado 2**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
Jurado 3**

**PROGRAMA:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**MODALIDAD DE TRABAJO DE GRADO**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| x | Investigación |  | Pasantía de Investigación |  | Docencia |  | Práctica Empresarial |
|  | Recital de Grado |  | Diplomado |  | Práctica Integral |  | Articulación Posgrado |

**EL JURADO CALIFICADOR CONFORMADO POR:** (Nombres, apellidos y documento de identidad).

**JURADO 1:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_/ **C.C:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**JURADO 2:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_/ **C.C:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**JURADO 3:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_/ **C.C:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**EN SU SESIÓN EFECTUADA EN: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_A LAS \_\_\_\_\_\_\_\_HORAS, DEL DÍA\_\_\_\_\_\_\_**

**DEL MES \_\_\_\_\_\_\_\_ DEL AÑO \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Terminadas sus deliberaciones, y en cumplimiento de las normas y acuerdos de los órganos de dirección de la Universidad de Pamplona, se ha llegado a la siguiente conclusión:

**Primera Conclusión: Otorgar la Calificación de: \_\_\_. \_\_\_** *(en números)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Meritorio (>=4.51) |  | Excelente (>=4, <=4.49) |  | Aprobado (>=3, <=3.99) |  | Incompleto (<=2.99) |

**AL TRABAJO DE GRADO TITULADO: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**AUTOR(ES):** Número de Autores (\_\_\_\_)

|  |  |  |  |
| --- | --- | --- | --- |
| Nombres: |  | COD. |  |
| Nombres: |  | COD. |  |
| Nombres: |  | COD. |  |

**DIRECTOR Y/O TUTOR:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_/**C.C: \_\_\_\_**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Segunda Conclusión: Emitir los siguientes criterios**

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **DESCRIPCIÓN** | **RECOMENDAR** | |
| **SI** | **NO** |
| 1. | Recomendar para presentar en eventos. |  |  |
| 2. | Recomendar para publicación. |  |  |
| 3. | Recomendar para ser continuado en otros trabajos. |  |  |

Otras: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Tercera Conclusión: Avalar el cumplimiento del Trabajo de Grado, para optar por el Título de**

***Firmas del Jurado Calificador:***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| **JURADO 1** |  | **JURADO 2** |  | **JURADO 3** |

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Director Comité Trabajo de Grado** |  | **Director Unidad Académica** |

**Dedicatoria**

*A mi querida madre Luz Nelly León, que de no ser por ella nada de lo que he sido, soy y seré tendría sentido…*

**Pensamiento**

Arrójale ganas a tus objetivos

con firmeza y mucha potencia

para que perduren mientras

fluyen sobre el lago “Tú

podrás”, *el que llegue más*

*lejos le aconseja la técnica a*

*los demás.*

*Jorge Miguel Anaya León*

**Agradecimiento**

*A mi Mamá, a Papa, a mis hermanas, a mi familia, a los docentes de la Universidad de Pamplona especialmente a los del programa Ingeniería de Sistemas, a los compañeros y a las demás personas que de una u otra forma aportaron en mi formación profesional y personal.*

**Resumen**

La Encuesta Nacional de Examen de Salud y Nutrición (NHANES) es un programa de estudios diseñado para evaluar el estado de salud y nutrición de los adultos y de los niños en Estados Unidos. Esta encuesta es un importante estudio del Centro Nacional de Estadísticas de Salud. Realizar investigaciones sobre la salud de los adolescentes y la salud de personas de mayor edad que permita producir estadísticas confiables para estos grupos es un tema relevante para Colombia y el mundo. La muestra para la encuesta se selecciona para representar a la población de Estados Unidos de todas las edades. Aunque es así, tomarlo como referencia y replicarlo después en la población de Colombia se considera relevante como área de investigación. El objetivo de esta investigación es desarrollar un Software Interactivo para el proceso de Integración Automática de Bases de Datos Médicas mostradas por NHANES, de manera que permita de forma sencilla y rápida la descarga de datas y la realización de fusiones automáticas entre ellas para procesos particulares de investigación en el área de salud. La validación del software se realiza con base a la fusión de 2 datas correspondientes los años 2016 y 2017 de forma manual comparándola con la realizada por el software. La metodología para el desarrollo es una adaptación de la metodología Programación Extrema (XP) por ser esta ágil y flexible, en ella destacan las fases de interacción constante entre el usuario y el equipo de desarrollo, planificación flexible abierta y rápida respuesta a cambios. El lenguaje de desarrollo es PHYTON por ser software libre y haberse convertido en uno de los movimientos tecnológicos de mayor auge en el siglo XXI.

Palabras clave: NHANES, Programación Extrema, Python, Integración Base de Datos, Software Interactivo.

**Abstract**

The National Survey of Health and Nutrition Examination (NHANES) is a program of study designed to assess the health and nutrition of adults and children in the United States. This survey is an important study made by the National Center for Health Statistics. Conduct research on adolescent health and the health of older people that can help produce reliable statistics for these groups is an important issue for Colombia and the world. The survey sample is selected to represent the US population of all ages. Even so, take it as a reference and then replicate the population of Colombia is considered relevant as a research area. The objective of this research is to develop an interactive software for the process of automatic integration of medical databases shown by NHANES, so as to enable simple and fast downloading data and performing automatic mergers between them for particular processes research in the health area. Software validation is performed based on the fusion of two data corresponding 2016 and 2017 manually and comparing the results obtained with the software. The methodology for development is an adaptation of Extreme Programming (XP) methodology to be this agile and flexible, it highlights the phases of constant interaction between the user and the development team, open flexible planning and rapid response to change. the development language is PYTHON due to the fact that is a open source software and has become one of the technological movements of the XXI Century.

Keywords**:** NHANES, Extreme Programming, Python, Database Integration, Interactive Software.

**ÍNDICE GENERAL**

[INTRODUCCIÓN 16](#_Toc26388305)

[CAPÍTULO 1. EL PROBLEMA 18](#_Toc26388306)

[1.1 Planteamiento del problema 18](#_Toc26388307)

[1.2 Formulación del problema 19](#_Toc26388308)

[1.3 Objetivos de la investigación 19](#_Toc26388309)

[1.3.1 Objetivo general 19](#_Toc26388310)

[1.3.2 Objetivos específicos 19](#_Toc26388311)

[1.4 Justificación de la investigación 20](#_Toc26388312)

[CAPÍTULO 2. MARCO TEÓRICO 21](#_Toc26388313)

[2.1 Antecedentes de la investigación 21](#_Toc26388315)

[2.2 Bases teóricas 23](#_Toc26388316)

[2.2.1 Metodología de desarrollo 23](#_Toc26388317)

[2.2.2 NHANES 24](#_Toc26388318)

[2.2.3 Programación Extrema (XP) 24](#_Toc26388319)

[2.2.4 NetBeans 39](#_Toc26388320)

[2.2.5 Lenguaje Python 40](#_Toc26388321)

[2.3 Marco legal 43](#_Toc26388322)

[2.3.1 Datos abiertos 43](#_Toc26388323)

[2.3.2 Ley 37/2007 sobre reutilización de la información del sector público 43](#_Toc26388324)

[2.3.3 RD 1495/2011 sobre reutilización de la información del sector público 44](#_Toc26388325)

[CAPÍTULO 3. MARCO METODOLÓGICO 46](#_Toc26388326)

[3.1 Enfoque epistemológico 46](#_Toc26388328)

[3.2 Diseño de investigación 46](#_Toc26388329)

[3.3 Nivel de investigación 47](#_Toc26388330)

[3.4 Población y muestra 47](#_Toc26388331)

[CAPÍTULO 4. DESARROLLO DE LA INVESTIGACIÓN 48](#_Toc26388332)

[4.1 Descripción de la base de datos en la plataforma NHANES en línea 48](#_Toc26388334)

[4.2 Definición de la metodología de desarrollo y datas a fusionar para la validación del software 49](#_Toc26388335)

[4.2.1 Comparación XP, SCRUM e ICONIX 49](#_Toc26388336)

[4.3 Diseño del software interactivo 50](#_Toc26388337)

[4.3.1 Diseño simple 51](#_Toc26388338)

[4.3.2 Historias de usuario 51](#_Toc26388339)

[4.3.3 Acotaciones del software 56](#_Toc26388340)

[4.3.4 Asignación de roles 57](#_Toc26388341)

[4.3.5 Plan de entrega 57](#_Toc26388342)

[4.3.6 Tareas de ingenierías (Task Card) 58](#_Toc26388343)

[4.3.7 Tarjetas CRC (Clase – Responsabilidades – Colaboradores) 62](#_Toc26388344)

[4.4 Desarrollo del software interactivo 65](#_Toc26388345)

[4.4.1 Fase 1 (Preparación del entorno de desarrollo NetBeans con Python) 65](#_Toc26388346)

[4.4.2 Fase 2 (Desarrollo de SoftDataFusion, GUI) 66](#_Toc26388347)

[4.5 Pruebas de aceptación 68](#_Toc26388348)

[4.6 Análisis e interpretación de resultados 72](#_Toc26388349)

[4.6.1 Archivo con extensión .CSV 73](#_Toc26388350)

[4.6.2 Archivo con extensión .XPT 74](#_Toc26388351)

[4.6.3 Archivo con extensión .XLSX 75](#_Toc26388352)

[4.7 Validación del software de investigación 76](#_Toc26388353)

[CAPÍTULO 5. CONCLUSIONES 78](#_Toc26388354)

[CAPÍTULO 6. RECOMENDACIONES 80](#_Toc26388355)

[CAPÍTULO 7. BIBLIOGRAFÍA 81](#_Toc26388356)

[CAPÍTULO 8. ANEXOS 88](#_Toc26388357)

[8.1 Manual de Usuario 88](#_Toc26388362)

[8.1.1 SoftDataFusion 88](#_Toc26388363)

[8.2 Certificado de registro del Software ante la DNDA en Colombia 98](#_Toc26388364)

[8.3 SoftDataFusion en GitHub 101](#_Toc26388365)

[8.4 Conectando a los investigadores con la investigación - ORCID 101](#_Toc26388366)

[8.5 Sincronización de GitHub con Zenodo (Generación del DOI) 102](#_Toc26388367)

[8.6 Publicación del Artículo de Herramienta de Software en F1000 104](#_Toc26388368)

[8.7 Estadísticos de monitoreo de SoftDataFusion en línea 107](#_Toc26388369)

**LISTA DE TABLAS**

[Tabla 1 Plantilla para Historias de usuario (Sintya Milena…, 2016). 27](#_Toc26388370)

[Tabla 2 Plantilla para Tareas de ingeniería (Sintya Milena..., 2016). 28](#_Toc26388371)

[Tabla 3 Plantilla para Pruebas de aceptación (Sintya Milena..., 2016). 29](#_Toc26388372)

[Tabla 4 Plantilla para Tarjetas CRC (Sintya Milena..., 2016). 29](#_Toc26388373)

[Tabla 5 Resumen cubrimiento porcentual por área de proceso de cada metodología (Jaime Andrés Britto Montoya, 2016). 50](#_Toc26388374)

[Tabla 6 HU: Selección del NHANES (Fuente propia, 2019). 52](#_Toc26388375)

[Tabla 7 HU: Selección de la DATA (Fuente propia, 2019). 53](#_Toc26388376)

[Tabla 8 HU: Selección del FILE (Fuente propia, 2019). 53](#_Toc26388377)

[Tabla 9 HU: Enlistar los FILES a fusionar (Fuente propia, 2019). 53](#_Toc26388378)

[Tabla 10 HU: Des enlistar los FILES a fusionar (Fuente propia, 2019). 54](#_Toc26388379)

[Tabla 11 HU: Cargar las VARIABLES de los FILES (Fuente propia, 2019). 54](#_Toc26388380)

[Tabla 12 HU: Enlistar todas las VARIABLES a fusionar (Fuente propia, 2019). 54](#_Toc26388381)

[Tabla 13 HU: Des enlistar todas las VARIABLES a fusionar (Fuente propia, 2019). 55](#_Toc26388382)

[Tabla 14 HU: Enlistar las VARIABLES a fusionar (Fuente propia, 2019). 55](#_Toc26388383)

[Tabla 15 HU: Des enlistar las VARIABLES a fusionar (Fuente propia, 2019). 55](#_Toc26388384)

[Tabla 16 HU: Valor opcional para las celdas sin coincidencia de fusión (Fuente propia, 2019). 55](#_Toc26388385)

[Tabla 17 HU: Tipo de archivo de exportación (Fuente propia, 2019). 56](#_Toc26388386)

[Tabla 18 Asignación de Roles (Fuente propia, 2019). 57](#_Toc26388387)

[Tabla 19 Plan de Entrega (Fuente propia, 2019). 57](#_Toc26388388)

[Tabla 20 Tareas de Ingeniería (Fuente propia, 2019). 58](#_Toc26388389)

[Tabla 21 TI: Seleccionar el NHANES (Fuente propia, 2019). 58](#_Toc26388390)

[Tabla 22 TI: Seleccionar la DATA (Fuente propia, 2019). 59](#_Toc26388391)

[Tabla 23 TI: Seleccionar el FILE (Fuente propia, 2019). 59](#_Toc26388392)

[Tabla 24 TI: Enlistar los FILES a fusionar (Fuente propia, 2019). 59](#_Toc26388393)

[Tabla 25 TI: Des enlistar los FILES a fusionar (Fuente propia, 2019). 59](#_Toc26388394)

[Tabla 26 TI: Cargar las VARIABLES de los FILES (Fuente propia, 2019). 60](#_Toc26388395)

[Tabla 27 TI: Enlistar todas las VARIABLES a fusionar (Fuente propia, 2019). 60](#_Toc26388396)

[Tabla 28 TI: Des enlistar todas las VARIABLES a fusionar (Fuente propia, 2019). 60](#_Toc26388397)

[Tabla 29 TI: Enlistar las VARIABLES a fusionar (Fuente propia, 2019). 61](#_Toc26388398)

[Tabla 30 TI: Des enlistar las VARIABLES a fusionar (Fuente propia, 2019). 61](#_Toc26388399)

[Tabla 31 TI: Valor opcional para las celdas sin coincidencia de fusión (Fuente propia, 2019). 61](#_Toc26388400)

[Tabla 32 TI: Tipo de archivo de exportación (Fuente propia, 2019). 62](#_Toc26388401)

[Tabla 33 Tarjeta CRC: SoftDataFusion (Fuente propia, 2019). 62](#_Toc26388402)

[Tabla 34 Tarjeta CRC: CreateToolTip (Fuente propia, 2019). 63](#_Toc26388403)

[Tabla 35 Tarjeta CRC: Linkbutton (Fuente propia, 2019). 63](#_Toc26388404)

[Tabla 36 Tarjeta CRC: Application (Fuente propia, 2019). 65](#_Toc26388405)

[Tabla 37 Pruebas de Aceptación (Fuente propia, 2019). 69](#_Toc26388406)

[Tabla 38 Prueba de Aceptación: PA01 (Fuente propia, 2019). 69](#_Toc26388407)

[Tabla 39 Prueba de Aceptación: PA02 (Fuente propia, 2019). 69](#_Toc26388408)

[Tabla 40 Prueba de Aceptación: PA03 (Fuente propia, 2019). 69](#_Toc26388409)

[Tabla 41 Prueba de Aceptación: PA04 (Fuente propia, 2019). 70](#_Toc26388410)

[Tabla 42 Prueba de Aceptación: PA05 (Fuente propia, 2019). 70](#_Toc26388411)

[Tabla 43 Prueba de Aceptación: PA06 (Fuente propia, 2019). 70](#_Toc26388412)

[Tabla 44 Prueba de Aceptación: PA07 (Fuente propia, 2019). 71](#_Toc26388413)

[Tabla 45 Prueba de Aceptación: PA08 (Fuente propia, 2019). 71](#_Toc26388414)

[Tabla 46 Prueba de Aceptación: PA09 (Fuente propia, 2019). 71](#_Toc26388415)

[Tabla 47 Prueba de Aceptación: PA10 (Fuente propia, 2019). 71](#_Toc26388416)

[Tabla 48 Prueba de Aceptación: PA011 (Fuente propia, 2019). 72](#_Toc26388417)

[Tabla 49 Prueba de Aceptación: PA012 (Fuente propia, 2019). 72](#_Toc26388418)

[Tabla 50 Análisis de Data en fusión Manuel y la generada (Fuente propia, 2019). 76](#_Toc26388419)

[Tabla 51 Cantidad de visitas a SoftDataFusión. 108](#_Toc26388420)

[Tabla 52 Cantidad de descargas de SoftDataFusion. 108](#_Toc26388421)

[Tabla 53 Cantidad de fusiones realizadas con SoftDataFusion. 109](#_Toc26388422)

**LISTA DE ILUSTRACIONES**

[Ilustración 1 NHANES (NHANES, 2019). 24](#_Toc26388980)

[Ilustración 2 Programación Extrema (Diego Calvo, 2018). 25](#_Toc26388981)

[Ilustración 3 Entorno de desarrollo NetBeans (NetBeans, 2019). 39](#_Toc26388982)

[Ilustración 4 Lenguaje Python (Guido van Rossum, 2009). 40](#_Toc26388983)

[Ilustración 5 Librería de código abierto, pandas (PandasPyData, 2019). 41](#_Toc26388984)

[Ilustración 10 Estructura de Datos Abiertos NHANES en línea (Fuente propia, 2019). 48](#_Toc26388985)

[Ilustración 11 Proceso de fusión (Fuente propia, 2019). 51](#_Toc26388986)

[Ilustración 12 Diseño simple del modelo de datos (Fuente propia, 2019). 51](#_Toc26388987)

[Ilustración 13 Ventana multipropósito y ágil uso del software (Fuente propia, 2019). 66](#_Toc26388988)

[Ilustración 14 Obtención de la data en línea de NHANES (Fuente propia, 2019). 67](#_Toc26388989)

[Ilustración 15 Lectura de archivos .XPT (Fuente propia, 2019). 67](#_Toc26388990)

[Ilustración 16 Fusión de las datas y la escritura de archivos (Fuente propia, 2019). 68](#_Toc26388991)

[Ilustración 17 Extensiones de exportación de archivos (Fuente propia, 2019). 73](#_Toc26388992)

[Ilustración 18 Data generada en extensión .CSV (Fuente propia, 2019). 74](#_Toc26388993)

[Ilustración 19 Data generada en extensión .XPT (Fuente propia, 2019). 74](#_Toc26388994)

[Ilustración 20 Data generada en extensión .XLSX (Fuente propia, 2019). 75](#_Toc26388995)

[Ilustración 21 MU: Bienvenida e inicio del tutorial (Fuente propia, 2019). 88](#_Toc26388996)

[Ilustración 22 MU: Selección del año del NHANES (Fuente propia, 2019). 89](#_Toc26388997)

[Ilustración 23 MU: Selección de la DATA (Fuente propia, 2019). 89](#_Toc26388998)

[Ilustración 24 MU: Selección de la ARCHIVO/FILE (Fuente propia, 2019). 90](#_Toc26388999)

[Ilustración 25 MU: Lista de ARCHIVOS/FILES a fusionar (Fuente propia, 2019). 90](#_Toc26389000)

[Ilustración 26 MU: Este botón permite Ilustración 24 a Ilustración 25 (Fuente propia, 2019). 90](#_Toc26389001)

[Ilustración 27 MU: Este botón permite desagregar los ARCHIVOS/FILES de Ilustración 25 (Fuente propia, 2019). 91](#_Toc26389002)

[Ilustración 28 MU: Cargar las VARIABLES de los ARCHIVOS/FILES (Fuente propia, 2019). 91](#_Toc26389003)

[Ilustración 29 MU: Lista de VARIABLES A SELECCIONAR (Fuente propia, 2019). 92](#_Toc26389004)

[Ilustración 30 MU: Lista de VARIABLES SELECCIONADAS (Fuente propia, 2019). 92](#_Toc26389005)

[Ilustración 31 MU: Mover todas las VARIABLES a VARIABLES SELECCIONADAS (Fuente propia, 2019). 93](#_Toc26389006)

[Ilustración 32 MU: Mover de a una VARIABLE a VARIABLES SELECCIONADAS (Fuente propia, 2019). 93](#_Toc26389007)

[Ilustración 33 MU: Remover de a una VARIABLE de VARIABLES SELECCIONADAS (Fuente propia, 2019). 94](#_Toc26389008)

[Ilustración 34 MU: Remover todas las VARIABLES de VARIABLES SELECCIONADAS (Fuente propia, 2019). 94](#_Toc26389009)

[Ilustración 35 MU: VALOR para los CAMPOS de REGISTROS SIN RELACIÓN (Fuente propia, 2019). 95](#_Toc26389010)

[Ilustración 36 MU: Formatear para la NUEVA FUSION/NEW MERGE (Fuente propia, 2019). 95](#_Toc26389011)

[Ilustración 37 MU: Seleccionar TIPO DE ARCHIVO para EXPORTAR (Fuente propia, 2019). 96](#_Toc26389012)

[Ilustración 38 MU: Realiza la FUSIÓN/MERGE (Fuente propia, 2019). 96](#_Toc26389013)

[Ilustración 39 MU: Final del Tutorial (Fuente propia, 2019). 97](#_Toc26389014)

[Ilustración 40 Logo de SoftDataFusion (Fuente propia, 2019). 98](#_Toc26389015)

[Ilustración 41 Certificado de registro de soporte lógico - Software ante la DNDA en Colombia (Fuente propia, 2019) 100](#_Toc26389016)

[Ilustración 42 SoftDataFusion en GitHub (Fuente propia, 2019). 101](#_Toc26389017)

[Ilustración 43 7.4 Conectando a los investigadores con la investigación – ORCID. 102](#_Toc26389018)

[Ilustración 44 7.4 Sincronización de GitHub con Zenodo (Generación del DOI). 104](#_Toc26389019)

[Ilustración 45 Envío de el Artículo Herramienta de Software a F1000. 105](#_Toc26389020)

[Ilustración 46 Correcciones u observaciones de F1000 a tener en cuenta para la publicación. 106](#_Toc26389021)

[Ilustración 47 Estadística de eventos por fechas de SoftDataFusion. 109](#_Toc26389022)

[Ilustración 48 Estadística de eventos por horas de SoftDataFusion. 110](#_Toc26389023)

# **INTRODUCCIÓN**

NHANES es un programa de estudios diseñado para evaluar el estado de salud y nutrición de adultos y niños en los Estados Unidos, así mismo, es un programa importante del Centro Nacional de Estadísticas de Salud (NCHS). NCHS es parte de los Centros para el Control y Prevención de Enfermedades (CDC) y tiene la responsabilidad de producir estadísticas vitales y de salud para la Nación. La encuesta examina una muestra nacionalmente representativa de cerca de 5.000 personas cada año. Estas personas están situadas en condados de todo el país, 15 de los cuales son visitados cada año (NHANES, 2003). Para futuras investigaciones en las áreas de la salud de los adolescentes, la salud de personas de mayor edad y en otras que pudieran surgir que permitan producir estadísticas confiables, actualmente no se cuenta con un software que facilite obtener, interpretar y manipular de forma rápida y fácil las datas necesarias que suministra NHANES. Realizar estos procesos actualmente consumen excesivo tiempo, así como también el seleccionar las datas relevantes ya que habría que gestionar estas de forma manual paso a paso y solo para visualización de los datos, lo que hace a estos procesos complejos, engorrosos y complicados para casos de fusión de datos de diferentes años y diferentes estudios que posee NHANES.

Un software interactivo por definición es aquél que necesita la realimentación continúa del usuario para poder ejecutarse. No es común que un software se diseñe para apoyar a los investigadores y que además sea adecuado para utilizarlo en las investigaciones y en la enseñanza (Goodhew Peter, 2002). Como antecedente podemos mencionar los datos para crear las tablas de crecimiento usadas nacionalmente por los pediatras para evaluar el crecimiento de los niños Marrodán Maria. (2013). Estas tablas han sido adaptadas y adoptadas mundialmente como patrones de referencia y han sido recientemente actualizadas usando las cifras más recientes de NHANES. Sin embargo, una debilidad en esta investigación fue la selección de variables para hacer el estudio. Así mismo en la investigación sobre el plomo en la sangre (AGUILAR VALDES, 2003) en la cual fue útil para desarrollar políticas para eliminar el plomo en la gasolina y en la soldadura de las latas de alimentos y refrescos. La información de las encuestas realizadas indica que la política ha sido aún más efectiva de lo que originalmente se imaginaba, con un descenso en los niveles elevados de plomo en la sangre de más de 70% desde los años 70. En esta investigación los datos se descargaron manualmente consumiendo tiempo valioso.

También podemos mencionar que actualmente los programas nacionales para reducir los niveles de hipertensión y de colesterol continúan dependiendo de la información de NHANES para dirigir los programas de educación y prevención hacia aquellos en riesgos y para medir el éxito de la reducción de los factores de riesgo asociados con las enfermedades del corazón, la causa número uno de muerte en la nación (Benjamin EJ, 2017). Para estas futuras investigaciones en esta área y otras que pudieran surgir, actualmente no se cuenta con una aplicación amigable que permita tener en forma rápida y expedita las datas necesarias mostrando de esta forma el problema a abordar y la ocurrencia de este se refleja en : Excesivo tiempo en identificar las datas relevantes, complejidad en la selección de variables a considerar para las investigaciones, engorroso y complicado el proceso de fusión de datos de diferentes años y diferentes estudios ocasionando extemporaneidad de la investigación así como retraso en tomas de decisiones trascendentales en el área de salud.

# **CAPÍTULO 1. EL PROBLEMA**

## **Planteamiento del problema**

Realizar investigaciones sobre la salud de los adolescentes y la salud de personas de mayor edad que permita producir estadísticas confiables para estos grupos es un tema relevante para Colombia y el mundo. NHANES ha proporcionado datos a diversos investigadores para sus estudios, ocasión oportuna como antecedente, mencionar los datos para crear las tablas de crecimiento usadas nacionalmente por los pediatras para evaluar el crecimiento de los niños. En (Marrodán Maria, 2013) las tablas han sido adaptadas y adoptadas mundialmente como patrones de referencia y han sido recientemente actualizadas usando las cifras más recientes de NHANES. Sin embargo, una debilidad en esta investigación fue la selección de variables para hacer el estudio. Así mismo en la investigación sobre el plomo en la sangre (AGUILAR VALDES, 2003) en la cual fue útil para desarrollar políticas para eliminar el plomo en la gasolina y en la soldadura de las latas de alimentos y refrescos. La información de las encuestas realizadas indica que la política ha sido aún más efectiva de lo que originalmente se imaginaba, con un descenso en los niveles elevados de plomo en la sangre de más de 70% desde los años 70. En esta investigación los datos se descargaron manualmente consumiendo tiempo valioso. Los programas nacionales para reducir los niveles de hipertensión y de colesterol continúan dependiendo de la información suministrada por NHANES para dirigir los programas de educación y prevención hacia aquellos en riesgos y para medir el éxito de la reducción de los factores de riesgo asociados con las enfermedades del corazón, la causa número uno de muerte en la nación (Benjamin EJ, 2017). Para futuras investigaciones en las áreas anteriormente mencionadas y en otras que pudieran surgir, actualmente no se cuenta con un software que permita obtener, interpretar y manipular de forma rápida y fácil las datas necesarias que suministra NHANES, causando considerablemente de este modo excesivo tiempo en seleccionar las datas relevantes (ya que habría que gestionar una cuenta en línea para la obtención de un software especializado para solo la visualización de los datos), así mismo disminuya la complejidad en la selección de variables de investigación, y agilice el engorroso y complicado proceso de fusión de datos de diferentes años y diferentes estudios que posee, considerando. Cabe preguntar:

## **Formulación del problema**

*¿Cómo contribuir a mejorar y facilitar a los investigadores en la obtención de forma rápida datos para las investigaciones y evite la extemporaneidad y el retraso en tomas de decisiones trascendentales en el área de salud*

## **Objetivos de la investigación**

### **Objetivo general**

Desarrollar un software interactivo para la Integración Automática de Bases de Datos Médicas, caso de estudio: Encuesta Nacional de Examen de Salud y Nutrición (NHANES).

### **Objetivos específicos**

* Describir la base de datos de la plataforma NHANES utilizando métodos gráficos para definir la manera en la que el software accederá a ellas.
* Definir la metodología de desarrollo y datas a fusionar para la obtención de un producto que valide el software.
* Diseñar el software interactivo usando la metodología de desarrollo definida para generar datas de investigación.
* Validar el software con la implementación de uno de sus productos en grupos de investigación.

## **Justificación de la investigación**

Es importante y necesario desarrollar un Software Interactivo para el proceso de Integración Automática de Bases de Datos Médicas mostradas por NHANES, de manera que permita de forma sencilla y rápida, el proceso, la descarga de datas y la realización de fusiones automáticas entre ellas para procesos particulares de investigación en el área de salud. Evitando de esta manera la extemporaneidad y tiempos adicionales para la culminación de una investigación abordad por temas trascendental en el área de la salud y por ser el tercero de los objetivos de desarrollo sostenible de los Estados Unidos, Salud y Bienestar.

# **CAPÍTULO 2. MARCO TEÓRICO**



## **Antecedentes de la investigación**

En el trabajo presentado en (Sofía Victoria, 2012) se visualiza la utilización de bases de datos y refleja la importancia en el desempeño de los médicos, debido a las facilidades que brindan para el tratamiento de información médica. El objetivo general de ese trabajo fue identificar necesidades de aprendizaje sobre el trabajo con bases de datos que poseen médicos de la atención primaria. Como metodología y datos se encuestaron 105 médicos que trabajaban en los clínicos universitarios del Municipio Playa, durante los años 2009-2010 y que practicaban funciones de docencia e investigación; se entrevistaron conjuntamente 9 profesores de Informática Médica de diferentes facultades de Medicina de ciudad de La Habana en Cuba. Se distinguieron necesidades de enseñanza relativas al estudio de bases de datos. Los médicos reconocieron la importancia que para su ejercicio tiene poder utilizar un gestor de bases de datos. Como conclusión de esa investigación se detectaron necesidades de aprendizaje relacionadas con el diseño y uso de bases de datos para la manipulación de la información de salud. Unos de los indicadores que se usaron para medir la variable necesidad de aprendizaje fueron: Utilización del Gestor de Base de Datos, importancia para el desempeño de la manipulación de datos con un gestor y valoración hacia el uso de un sistema gestión de base de datos. Como gestor de bases de datos se utilizó el software propietario ACCESS 2007 y para análisis estadístico Excel 2007 y STATISTIC versión 8. Igualmente, todos softwares propietarios. Como resultados importantes podemos resaltar: El 73 % de los médicos encuestados han recibido cursos de postgrado sin contenidos programáticos gestores de bases de datos. El 77 % de los profesionales no han recibido curso alguno de actualización en esta temática de gestores de bases de datos y finalmente. La investigación se realizó sobre software propietario y no sobre software libre.

En el trabajo presentado en (Mayenny Linares, 2014) cuyo objetivo general fue “Diseñar un software sobre programación y Gestores de Bases de datos, para mejorar el proceso de preparación de estudiantes, profesores y tutores de la carrera Sistemas de Información en Salud". Su enfoque fue de una investigación aplicada y prospectiva, utilizando métodos histórico-lógico y metodología de ingeniería de software. Como resultado de esta investigación se obtuvo el criterio de los estudiantes en cuanto a diseño de gestores de bases de datos y su importancia para la creación del software que se describe y que será utilizado en los cursos de la carrera proporcionando un medio de motivación y aprendizaje, mejor gestión de la información y una forma de facilitar la preparación de profesores y estudiantes en cualquier momento.

En (María del Carmen, 2014) se propone demostrar las ventajas del uso de las bases de datos distribuidas como parte del desarrollo de esta solución informática para el sector de la salud de Cuba. En este trabajo se desarrolla una plataforma constituida por diferentes componentes con un nivel de conexión y articulación que admite integrar las aplicaciones informáticas para la salud, interactuar entre ellas y reutilizar la información eficientemente; esta información es almacenada en bases de datos independientes, soportadas sobre un servidor MySQL y comunicándose entre sí mediante Servicios Web, basados en XML con arquitectura de 3 capas a través de una interfaz amigable. Esto se realiza con el fin de incorporar nuevas tecnologías, para mejor gestión de la información médica en todas las entidades. Finalmente, para lograr la integración de todas los componentes y bases de datos se crearon diferentes grupos de trabajo para cada componente en cada entidad y se creó un consejo de este grupo donde se analizaba cada proyecto en cada entidad de Cuba y sus resultados, realizándose el estudio de la integración entre los mismos.

La base de datos abierta La Cancer Genome Atlas presentada en la (NHANES, 2019), pone a disposición la recopilación de datos de Cancer Chromophobe (TCGA-KICH, por sus siglas en inglés). Esto es un esfuerzo mayor para construir una comunidad de investigación centrada en conectar fenotipos de cáncer a genotipos proporcionando imágenes clínicas adaptadas a sujetos de The Cancer Genome Atlas (TCGA). Los datos clínicos, genéticos y patológicos residen en el portal de datos de Genomic Data Commons (GDC), mientras que los datos radiológicos se almacenan en The Cancer Imaging Archive (TCIA). Las personas e instituciones que han proporcionado datos para esta recopilación son: National Cancer Institute, Bethesda, MD - Special thanks to Marston Linehan, MD and Rabindra Gautam from the Urologic Oncology Branch. Y Brigham & Women's Hospital Boston, MA Cheryl A. Sadow, MD and Seth Levine. Esta base de datos no cuenta con un software de gestor de selección ni manipulación de datas para descargar, todo el proceso de descarga no es automático lo que hace tediosa y lenta por tener que descargar toda la data se necesite o no y luego seleccionar en forma manual las imágenes para la investigación que se quiera realizar.

## **Bases teóricas**

### **Metodología de desarrollo**

Una Metodología de desarrollo de software se refiere al entorno que se usa para estructurar, planificar y controlar el proceso de desarrollo de un sistema de información, Estas metodologías consisten principalmente en hacer uso de diversas herramientas, técnicas, métodos y modelos para el desarrollo. Regularmente este tipo de metodología, tienen la necesidad de venir documentadas, para que los programadores que estarán dentro de la planeación del proyecto, comprendan perfectamente la metodología y en algunos casos el ciclo de vida del software que se pretende seguir (Orjuela Duarte & Mauricio, 2008).

### **NHANES**

NHANES es un programa de estudios diseñado para evaluar el estado de salud y nutrición de adultos y niños en los Estados Unidos, así mismo, es un programa importante del Centro Nacional de Estadísticas de Salud (NCHS). NCHS es parte de los Centros para el Control y Prevención de Enfermedades (CDC) y tiene la responsabilidad de producir estadísticas vitales y de salud para la Nación. La encuesta examina una muestra nacionalmente representativa de cerca de 5.000 personas cada año. Estas personas están situadas en condados de todo el país, 15 de los cuales son visitados cada año (NHANES, 2003).



Ilustración NHANES (NHANES, 2019).

### **Programación Extrema (XP)**

Es la metodología ágil más conocida, Fue desarrollada por Kent Beck buscando guiar equipos de desarrollo de software pequeños o medianos, entre dos y diez desarrolladores, en ambientes de requerimientos imprecisos o cambiantes.

Las prácticas de esta metodología se derivan de sus valores y principios y están enfocadas en darle solución a las actividades básicas de un proceso de desarrollo, esto es: escribir código, realizar pruebas, escuchar (planear) y diseñar.

Las prácticas de XP incluyen: planning game, pequeñas entregas, diseño simple, programación en pareja, pruebas, refactoring, integración continua, propiedad común del 56 código, pasó sostenible, cliente en sitio, metáfora y estándares de código. (Cadavid et al., 2013)

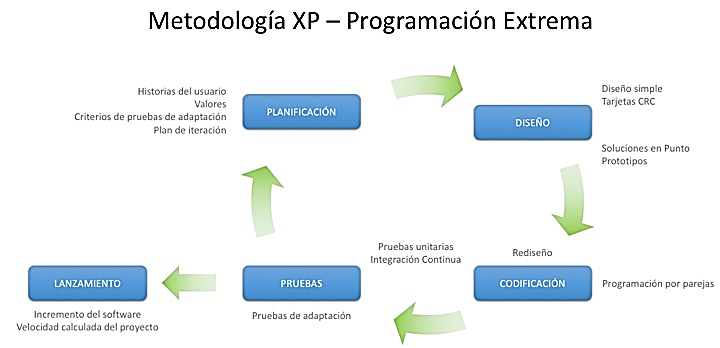


Ilustración Programación Extrema (Diego Calvo, 2018).

#### **Objetivos**

* La Satisfacción del cliente.
* Potenciar el trabajo en grupo.
* Minimizar el riesgo actuando sobre las variables del proyecto: costo, tiempo, calidad, alcance.

#### **Características**

* Metodología basada en prueba y error para obtener un software que funcione realmente.
* Fundamentada en principios.
* Está orientada hacia quien produce y usa software (el cliente participa muy activamente)
* Reduce el coste del cambio en todas las etapas del ciclo de vida del sistema.
* Combina las que han demostrado ser las mejores prácticas para desarrollar software, y las lleva al extremo.
* Cliente bien definido.
* Los requisitos pueden cambiar.
* Grupo pequeño y muy integrado (2-12 personas).
* Equipo con formación elevada y capacidad de aprender.

#### **Herramientas**

##### **Historias de usuario**

Las Historias de Usuario representan una breve descripción del comportamiento del sistema, se realizan por cada característica principal del sistema y son utilizadas para cumplir estimaciones de tiempo y el plan de lanzamientos, así mismo reemplazan un gran documento de requisitos y presiden la creación de las pruebas de aceptación.

Cada historia de usuario debe ser lo suficientemente comprensible y delimitada para que los programadores puedan implementarlas en unas semanas.

La plantilla a utilizarse para la elaboración de las historias de usuario se muestra en la Tabla 1 y cada uno de sus componentes se explica a continuación (Letelier & Penades, 2006).

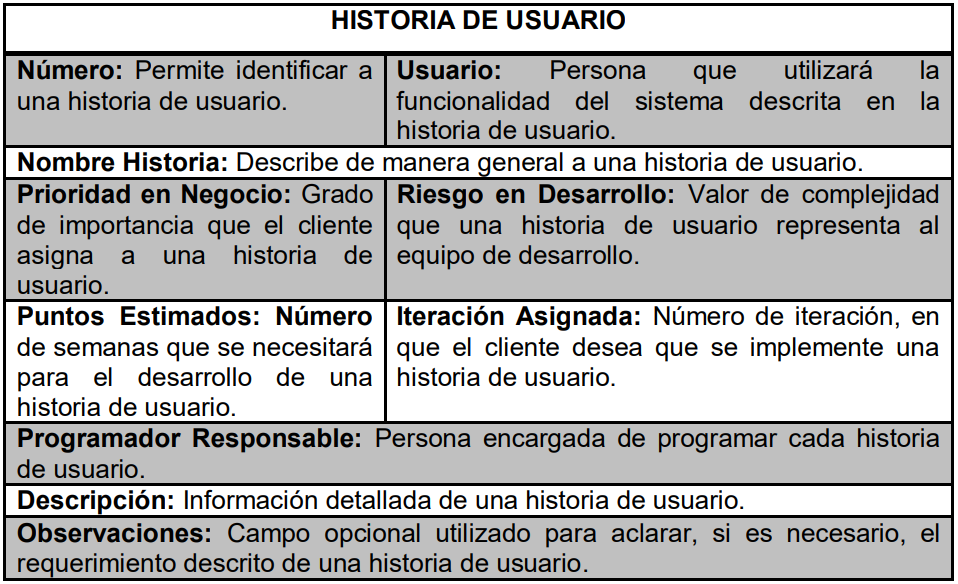


Tabla Plantilla para Historias de usuario (Sintya Milena…, 2016).

##### **Tareas de ingenierías (Task Card)**

Una Historias de Usuario se descompone en varias tareas de ingeniería, las cuales describen las actividades que se realizarán en cada historia de usuario, así mismo las tareas de ingeniería se vinculan más al desarrollador, ya que permite tener un acercamiento con el código. (Ferreira Escutia, 2013)

La plantilla a utilizarse para la elaboración de las tareas de ingeniería se muestra en la Tabla 2 y cada uno de sus componentes.

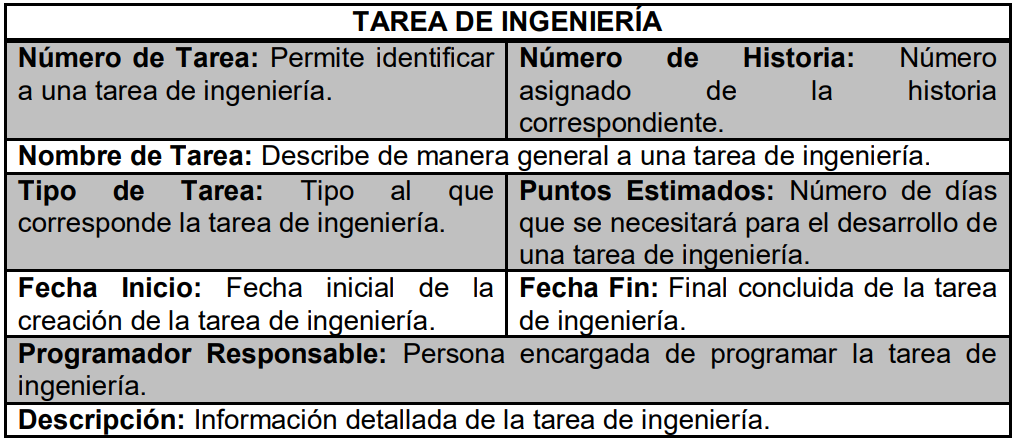


Tabla Plantilla para Tareas de ingeniería (Sintya Milena..., 2016).

##### **Pruebas de aceptación**

Según (Chiluisa Pallo & Loarte Cajamarca, 2014) Las Pruebas de aceptación son de vital importancia para el éxito de una iteración y el comienzo de la siguiente, con lo cual el cliente puede conocer el avance en el desarrollo del sistema y a los programadores lo que les resta por hacer. Además, permite una retroalimentación para el desarrollo de las próximas historias de usuarios a ser entregadas. Estas son comúnmente llamadas pruebas del cliente, por lo que son realizadas por el encargado de verificar si las historias de usuarios de cada iteración cumplen con la funcionalidad esperada.

La plantilla a utilizarse para la elaboración de las pruebas de aceptación se muestra en la Tabla 3 y a continuación se definen cada uno de los componentes.

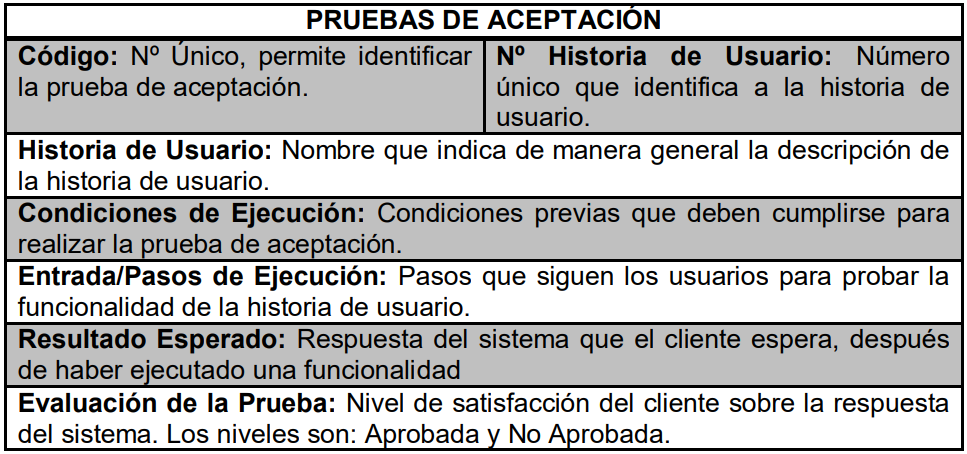


Tabla Plantilla para Pruebas de aceptación (Sintya Milena..., 2016).

##### **Tarjetas CRC (Clase – Responsabilidades – Colaboradores)**

Las Tarjetas CRC (Clase-Responsabilidades-Colaboradores), permiten conocer que clases componen el sistema y cuales interactúan entre sí. Se dividen en tres secciones: Nombre de la Clase, Responsabilidades y Colaboradores (Chiluisa Pallo & Loarte Cajamarca, 2014).

La Plantilla a utilizarse para la elaboración de las Tarjetas CRC se muestra en la Tabla 4 y a continuación se describen cada uno de los componentes.

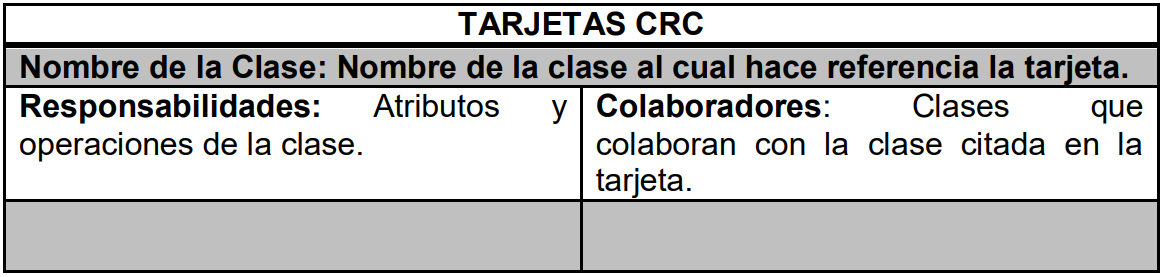


Tabla Plantilla para Tarjetas CRC (Sintya Milena..., 2016).

#### **Roles**

En (Erlijman Piwen & Goyen Fros, 2001) La Propuesta original de Beck incluye

los siguientes roles:

##### **Programador**

Es el responsable de implementar las historias de usuario por el cliente. Además, estima el tiempo de desarrollo de cada historia de usuario para que el cliente pueda asignarle prioridad dentro de la iteración. Cada iteración incorpora nueva funcionalidad de acuerdo a las prioridades establecidas por el cliente. El Programador también es responsable de diseñar y ejecutar los test de unidad del código que ha implementado o modificado.

##### **Cliente**

Determina la funcionalidad que se pretende en cada iteración y define las prioridades de implementación según el valor de negocio que aporta cada historia. El Cliente también es responsable de diseñar y ejecutar los test de aceptación.

##### **Encargado de pruebas (Tester)**

Es el Encargado de ejecutar las pruebas regularmente, difunde los resultados dentro del equipo y es también el responsable de las herramientas de soporte para pruebas.

##### **Encargado de seguimiento (Tracker)**

Una de las tareas más importante del Tracker, consiste en seguir la evolución de las estimaciones realizadas por los programadores y compararlas con el tiempo real de desarrollo. De esta forma, puede brindar información estadística en lo que refiere a la calidad de las estimaciones para que puedan ser mejoradas.

##### **Entrenador (Coach)**

Es responsable del proceso en general. Se encarga de iniciar y de guiar a las personas del equipo en poner en marcha cada una de las prácticas de la metodología XP.

##### **Consultor**

Es un Miembro externo del equipo con un conocimiento específico en algún tema necesario para el proyecto. Guía al equipo para resolver un problema específico.

##### **Gestor (Big Boss)**

Es el vínculo entre el cliente y programadores. Experto en tecnología y labores de gestión. Construye el plantel del equipo, obtiene los recursos necesarios y maneja los problemas que se generan. Administra a su vez las reuniones (planes de iteración, agenda de compromisos, etc). Su labor fundamental es de coordinación.

#### **Fases**

La Programación Extrema consta de 4 fases, las cuales son:

##### **Planeación**

La Metodología XP plantea la planificación como un diálogo continuo entre las partes involucradas en el proyecto, incluyendo al cliente, a los programadores y a los coordinadores. El proyecto comienza recopilando las historias de usuarios, las que constituyen a los tradicionales casos de uso. Una vez obtenidas estas historias de usuarios, los programadores evalúan rápidamente el tiempo de desarrollo de cada una.

Los Conceptos básicos de la planificación son:

**Las Historias de Usuarios**, las cuales son descritas por el cliente, en su propio lenguaje, como descripciones cortas de lo que el sistema debe realizar.

**El Plan de Entregas (Release Plan)**, establece que las historias de usuarios serán agrupadas para conformar una entrega y el orden de las mismas. Este cronograma será el resultado de una reunión entre todos los actores del proyecto.

**Plan de Iteraciones (Iteration Plan)**, las historias de usuarios seleccionadas para cada entrega son desarrolladas y probadas en un ciclo de iteración, de acuerdo al orden preestablecido.

**Reuniones Diarias de Seguimiento (Stand – Up Meeting)**, el objetivo es mantener la comunicación entre el equipo y compartir problemas y soluciones.

##### **Diseño**

La Metodología XP hace especial énfasis en los diseños simples y claros. Los conceptos más importantes de diseño en esta metodología son los siguientes:

**Simplicidad**, Un diseño simple se implementa más rápidamente que uno complejo. Por ello XP propone implementar el diseño más simple posible que funcione.

**Soluciones “Spike”**, Cuando aparecen problemas técnicos, o cuando es difícil de estimar el tiempo para implementar una historia de usuario, pueden utilizarse pequeños programas de prueba (llamados “Spike”), para explorar diferentes soluciones.

**Recodificación (“Refactoring”)**, Consiste en escribir nuevamente parte del código de un programa, sin cambiar su funcionalidad, a los efectos de crearlo más simple, conciso y entendible. Las metodologías de XP sugieren re codificar cada vez que sea necesario.

**Metáforas**, XP sugiere utilizar este concepto como una manera sencilla de explicar el propósito del proyecto, así como guiar la estructura del mismo. Una buena metáfora debe ser fácil de comprender para el cliente y a su vez debe tener suficiente contenido como para que sirva de guía a la arquitectura del proyecto.

##### **Codificación**

**Disponibilidad del Cliente**, Uno de los requerimientos de XP es tener al cliente disponible durante todo el proyecto. No solamente como apoyo a los desarrolladores, sino formando parte del grupo. El Involucramiento del cliente es fundamental para que pueda desarrollarse un proyecto con la metodología XP. Al comienzo del proyecto, el este debe proporcionar las historias de usuarios. Pero, dado que estas historias son expresamente cortas y de “alto nivel”, no contienen los detalles necesarios para realizar el desarrollo del código. Estos detalles deben ser proporcionados por el cliente, y discutidos con los desarrolladores, durante la etapa de desarrollo.

**Uso de Estándares**, XP promueve la programación basada en estándares, de manera que sea fácilmente entendible por todo el equipo, y que facilite la recodificación.

**Programación Dirigida por las Pruebas (“Test-Driven Programming”)**, En las metodologías tradicionales, la fase de pruebas, incluyendo la definición de los test, es usualmente realizada sobre el final del proyecto, o el final del desarrollo de cada módulo. La metodología XP propone un modelo inverso, primero se escribe los test que el sistema debe pasar. Luego, el desarrollo debe ser el mínimo necesario para pasar las pruebas previamente definidas. Las pruebas a los que se refiere esta práctica, son las pruebas unitarias, realizados por los desarrolladores. La definición de estos test al comienzo,

condiciona o “dirige” el desarrollo.

**Programación en Pares**, XP propone que se desarrolle en pares de programadores, ambos trabajando juntos en un mismo ordenador. Si bien parece que ésta práctica duplica el tiempo asignado al proyecto (y por ende, los costos en recursos humanos), al trabajar en pares se minimizan los errores y se logran mejores diseños, compensando la inversión en horas. El producto obtenido es por lo general de mejor calidad que cuando el desarrollo se realiza por programadores individuales.

**Integraciones Permanentes**, Todos los desarrolladores necesitan trabajar siempre con la “última versión”. Realizar cambios o mejoras sobre versiones antiguas causan graves problemas, y retrasan al proyecto. Es por eso que XP promueve publicar lo antes posible las nuevas versiones, aunque no sean las últimas, siempre que estén libres de errores. Idealmente, todos los días deben existir nuevas versiones publicadas. Para evitar errores, solo una pareja de desarrolladores puede integrar su código a la vez.

**Propiedad Colectiva del Código**, En un proyecto XP, todo el equipo puede contribuir con nuevas ideas que apliquen a cualquier parte del proyecto. Asimismo, una pareja de programadores puede cambiar el código que sea necesario para corregir problemas, agregar funciones o re codificar.

**Ritmo Sostenido**, La Metodología XP indica que debe llevarse un ritmo sostenido de trabajo. El concepto que se desea establecer con esta práctica es planificar el trabajo de forma a mantener un ritmo constante y razonable, sin sobrecargar al equipo.

##### **Pruebas**

**Pruebas Unitarias**, Todos los módulos deben de pasar las pruebas unitarias antes de ser liberados o publicados. Por otra parte, como se mencionó anteriormente, las pruebas deben ser definidas antes de realizar el código (“Test-Driven Programmming”). Que todo código liberado pase correctamente las pruebas unitarias, es lo que habilita que funcione la propiedad colectiva del código.

**Detección y Corrección de Errores**, Cuando se encuentra un error (“Bug”), éste debe ser corregido inmediatamente, y se deben tener precauciones para que errores similares no vuelvan a ocurrir. Asimismo, se generan nuevas pruebas para verificar que el error haya sido resuelto.

**Pruebas de Aceptación**, Son creadas en base a las historias de usuarios, en cada ciclo de la iteración del desarrollo. El Cliente debe especificar uno o diversos escenarios para comprobar que una historia de usuario ha sido correctamente implementada. Asimismo, en caso de que fallen varias pruebas, deben indicar el orden de prioridad de resolución. Una historia de usuario no se puede considerar terminada hasta que pase correctamente todas las pruebas de aceptación (Joskowicz, 2008).

##### **Lanzamiento**

Esta es la fase final en la cual el proyecto está terminado de acuerdo al trabajo realizado en conjunto con lo planeado, deseñado, codificado y probado. Listo para quedar en su lugar funcional, desde donde surgió la necesidad.

#### **Prácticas**

En (Echeverry Tobón & Delgado Carmona, 2007) La Metodología Extreme Programmming o XP, está orientada al desarrollo de software cuando los requerimientos son ambiguos o rápidamente cambiantes asumiéndolos como algo natural, por lo que los programadores deben responder a estos cambios cuando el cliente lo solicite.

XP es para pequeños y medianos equipos basándose en la comunicación continúa entre todos los participantes, la simplicidad en las soluciones implementadas y coraje para enfrentar los cambios.

Esta Metodología recomienda a seguir las siguientes prácticas:

##### **Comunicación**

Conversación continúa entre el equipo de desarrollo y el cliente, para implementar cambios lo antes posible.

##### **Entregas pequeñas**

Entrega en versiones operativas.

##### **Diseño simple**

Diseñar lo más posible, pero con la funcionalidad requerida.

##### **Pruebas**

Se realizan pruebas unitarias por parte de los programadores y pruebas de aceptación por parte del cliente.

##### **Refactorización (Refactoring)**

Remover código duplicado para facilitar los posteriores cambios.

##### **Programación en parejas**

Se realiza para contar con menor tasa de errores, mejor diseño y mayor satisfacción de los programadores.

##### **Integración continúa**

Cuando un fragmento de código esté listo, puede ser integrado al sistema.

##### **Cliente IN-SITU**

El Cliente debe estar presente y disponible para el equipo de desarrollo.

##### **Estándares de programación**

Normas definidas por los desarrolladores para tener un código legible.

##### **Juego de la planificación**

Desde el comienzo del desarrollo se requiere que el grupo y el cliente tengan una visión general del proyecto. En el transcurso del mismo se realizan diferentes reuniones, con el fin de organizar las tareas e ideas que surgen tanto por parte del cliente como del equipo.

##### **Propiedad colectiva del código**

El Código no es conocido por una sola persona del grupo del trabajo, esto facilita implementar cambios al programa por parte de otros integrantes del grupo.

##### **Utilización de metáforas del sistema**

Para mejorar el entendimiento de los elementos del sistema por parte del equipo de desarrollo se acude a la utilización de metáforas, como una forma de universalizar el lenguaje del sistema.

##### **Test del cliente**

El Cliente con la ayuda de los desarrolladores, propone sus propias pruebas para validar las Mini-Versiones.

##### **40 horas por semana**

Se debe de trabajar un máximo de 40 horas por semana.

### **NetBeans**

NetBeans es un proyecto exitoso de código abierto con una gran base de usuarios, una comunidad en constante crecimiento, y con cerca de 100 socios (¡y creciendo!) en todo el mundo. Sun MicroSystems fundó el proyecto de código abierto NetBeans en junio 2000 y continúa siendo el patrocinador principal de los proyectos.



Ilustración Entorno de desarrollo NetBeans (NetBeans, 2019).

Al día de hoy hay disponibles dos productos: el NetBeans IDE y NetBeans Platform.

NetBeans IDE es un entorno de desarrollo - una herramienta para que los programadores puedan escribir, compilar, depurar y ejecutar programas. Está escrito en Java - pero puede servir para cualquier otro lenguaje de programación. Existe además un número importante de módulos para extender el NetBeans IDE. NetBeans IDE es un producto libre y gratuito sin restricciones de uso.

También está disponible NetBeans Platform; una base modular y extensible usada como estructura de integración para crear grandes aplicaciones de escritorio. Empresas independientes asociadas, especializadas en desarrollo de software, proporcionan extensiones adicionales que se integran fácilmente en la plataforma y que pueden también utilizarse para desarrollar sus propias herramientas y soluciones (NetBeans, 2019).

### **Lenguaje Python**

Python es un lenguaje de programación poderoso y fácil de aprender. Cuenta con estructuras de datos eficientes y de alto nivel y un enfoque simple pero efectivo a la programación orientada a objetos. La elegante sintaxis de Python y su tipado dinámico, junto con su naturaleza interpretada, hacen de éste un lenguaje ideal para scripting y desarrollo rápido de aplicaciones en diversas áreas y sobre la mayoría de las plataformas (Guido van Rossum, 2009).



Ilustración Lenguaje Python (Guido van Rossum, 2009).

#### **Tkinter**

El módulo Tkinter (“Interfaz Tk”) es la interfaz estándar de Python para el kit de herramientas Tk GUI. Tanto Tk como Tkinter están disponibles en la mayoría de las plataformas Unix, así como en los sistemas Windows. (Tk en sí no es parte de Python; se mantiene en ActiveState).

Ejecutar *python -m Tkinter* desde la línea de comandos debería abrir una ventana que muestre una interfaz Tk simple, que le permita saber que Tkinter está instalado correctamente en su sistema, y ​​que también muestre qué versión de Tcl / Tk está instalada, para que pueda leer la documentación específica de Tcl / Tk versión (Python, 2019).

#### **pandas**

pandas es una biblioteca de código abierto con licencia BSD que proporciona estructuras de datos y herramientas de análisis de datos de alto rendimiento y fáciles de usar para el lenguaje de programación Python.

pandas es un proyecto patrocinado por NumFOCUS. Esto ayudará a asegurar el éxito del desarrollo de los pandas como un proyecto de código abierto de clase mundial, y hace posible donar al proyecto (PandasPyData, 2019).

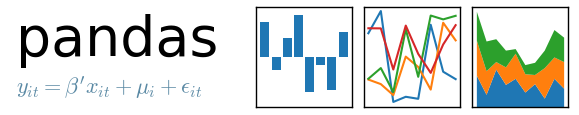


Ilustración Librería de código abierto, pandas (PandasPyData, 2019).

#### **xport**

XPORT es el formato de archivo binario utilizado por un grupo de agencias gubernamentales de los Estados Unidos para publicar conjuntos de datos. Tenía mucho sentido si intentaba leer archivos de datos en su mainframe IBM en 1988.

La especificación oficial de SAS para XPORT es relativamente sencilla. La parte más difícil es convertir el punto flotante en formato IBM al formato IEEE, que la especificación explica en detalle.

Hubo una actualización de la especificación XPT para SAS v8 y superior. Este módulo aún no se ha actualizado para funcionar con la nueva versión. Sin embargo, si está utilizando SAS v8 +, probablemente no esté utilizando el formato XPT. Los cambios en el formato parecen ser cambios triviales en los metadatos, pero la comprobación de errores actual de este módulo generará un ParseError. Si desea una actualización para v8, avíseme enviando un problema (pypi, 2019).

#### **PyInstaller |PyQT5**

PyInstaller agrupa una aplicación Python y todas sus dependencias en un solo paquete. El usuario puede ejecutar la aplicación empaquetada sin instalar un intérprete de Python o ningún módulo. Este lee un script de Python escrito por usted. Analiza su código para descubrir cualquier otro módulo y biblioteca que su script necesita para ejecutarse. Luego, recopila copias de todos esos archivos, ¡incluido el intérprete de Python activo! - y los coloca con su script en una sola carpeta, u opcionalmente en un solo archivo ejecutable.

PyInstaller se prueba con Windows, Mac OS X y Linux. Sin embargo, no es un compilador cruzado: para hacer una aplicación de Windows ejecutas PyInstaller en Windows; para crear una aplicación de Linux, se ejecuta en Linux, etc. PyInstaller se ha utilizado con éxito con AIX, Solaris y FreeBSD, pero no se ha probado con ellos (pypi, 2019).

## **Marco legal**

### **Datos abiertos**

Los datos abiertos son información pública dispuesta en formatos que permiten su uso y reutilización bajo licencia abierta y sin restricciones legales para su aprovechamiento. En Colombia, la Ley 1712 de 2014 sobre Transparencia y Acceso a la Información Pública Nacional, define los datos abiertos en el numeral sexto como “todos aquellos datos primarios o sin procesar, que se encuentran en formatos estándar e interoperables que facilitan su acceso y reutilización, los cuales están bajo la custodia de las entidades públicas o privadas que cumplen con funciones públicas y que son puestos a disposición de cualquier ciudadano, de forma libre y sin restricciones, con el fin de que terceros puedan reutilizarlos y crear servicios derivados de los mismos”. De este modo, la Ley establece la obligatoriedad de las entidades públicas de “divulgar datos abiertos”, teniendo en cuenta las excepciones de acceso a la información, asociadas a información clasificada y reservada establecidas en su título tercero, artículos 18 y 191 (Sylvia Constaín Rengifo, 2019).

### **Ley 37/2007 sobre reutilización de la información del sector público**

En España, las leyes sobre reutilización comienzan con la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público, como una transposición de la Directiva Europea 2003/98/CE. La ley alcanza a todas las Administraciones y organismos del sector público y se aplica a la reutilización de los documentos elaborados o custodiados por ellas. Podemos destacar algunos aspectos:

* Se posibilita la reutilización de documentos puestos a disposición del público sin licencia o con sujeción a condiciones establecidas bajo licencias-tipo.
* En cuanto a los formatos, se hace hincapié en la publicación basada en medios electrónicos y accesibles para personas con discapacidad.
* Se establecen unas condiciones generales de reutilización que, entre otras cosas, prohíben la alteración del contenido o su desnaturalización y obligan a citar la fuente y a indicar la fecha de última actualización. La norma incluye un régimen sancionador en el que se detalla un catálogo de multas (Jefatura del Estado, 2015).

### **RD 1495/2011 sobre reutilización de la información del sector público**

El 24 de octubre de 2011, a propuesta del Ministerio de Industria, Turismo y Comercio y del Ministerio de Política Territorial y Administración Pública, se publica el Real Decreto 1495/2011 con el objetivo de precisar el contenido de la Ley 37/2007 en el ámbito del sector público estatal. Aun cuando el Real decreto solo era de aplicación en la Administración General del Estado, la norma sirvió en varios casos como modelo procedimental para concretar algunas estrategias de carácter autonómico y local. Tomando como base el marco legal establecido por la Ley 37/2007, el Real Decreto regula los siguientes aspectos:

* Se autoriza la reutilización de los documentos del sector público estatal siempre y cuando no exista exclusión alguna y se establece que se realice de una «manera estructurada y usable» y «preferentemente en formatos procesables y accesibles de modo automatizado correspondientes a estándares abiertos».
* Se establece un modelo organizativo con la definición de funciones y obligaciones para la gestión de la Reutilización.
* La obligatoriedad de publicar información sobre los documentos reutilizables preferentemente en la sede electrónica de los organismos, además de ser incluida en el catálogo nacional datos.gob.es.
* Se incluyen una serie de condiciones legales aplicables en la puesta a disposición de los documentos reutilizables que deberán ser recogidas en un aviso legal. Dichas condiciones generales son las que se impondrá por defecto para los conjuntos de datos, existiendo la posibilidad de aplicación de condiciones específicas adicionales.
* Se regula lo relativo al uso de documentos reutilizables que estén sujetos a derechos de propiedad intelectual y lo referido a los documentos que contengan datos de carácter personal (Ministerio de la Presidencia, 2011).

# **CAPÍTULO 3. MARCO METODOLÓGICO**

El fin esencial del marco metodológico es el de asentar en el lenguaje de investigación los procesos e instrumentos que se emplearán en el trabajo, desde la concepción epistemológica, nivel de investigación, diseño de investigación, datos de investigación, población y muestra, desarrollo de la investigación y análisis e interpretación de resultados (Balestrini M, 2009).



## **Enfoque epistemológico**

El paradigma de la investigación con el cual se desarrolla el presente trabajo se considera positivista, ya que, un elevado porcentaje de la investigación que se pretende desarrollar responde a una estructura algorítmica que va a manejar y arrojar datos cuantitativos. Lo cual justifica que la metodología de la investigación sea cuantitativa y no experimental (Hernández R, … 2010).

## **Diseño de investigación**

Se presenta como un proyecto factible consistente en la investigación, elaboración y desarrollo de una propuesta de un modelo operativo viable para solucionar problemas, requerimientos o necesidades de organizaciones o grupo sociales; puede referirse a la formulación de políticas, programas, tecnológicas, métodos o procesos (UPEL, 2003).

## **Nivel de investigación**

Debido al grado de profundidad con el que se aborda este fenómeno de estudio hace que la Investigación sea Exploratoria. Teniendo claro su objetivo de examinar la realización de un sistema como SoftDataFusion (Arias F, 2012).

## **Población y muestra**

Los estudios realizados a las personas seleccionadas por NHANES en los Estados Unidos y compartidos en una Base de Datos Abierta en línea.

# **CAPÍTULO 4. DESARROLLO DE LA INVESTIGACIÓN**



## **Descripción de la base de datos en la plataforma NHANES en línea**

La Base de Datos Abierta NHANES tiene una estructura de datos con accesos mediante enlaces que permiten el desplazamiento web y descarga de datas organizada por NHANES transversales bienales, los cuales a su vez poseen tipos de estudios con sus respectivos archivos .XPT que SoftDataFusion interpretará y fusionará de acuerdo a las necesidades del investigador.

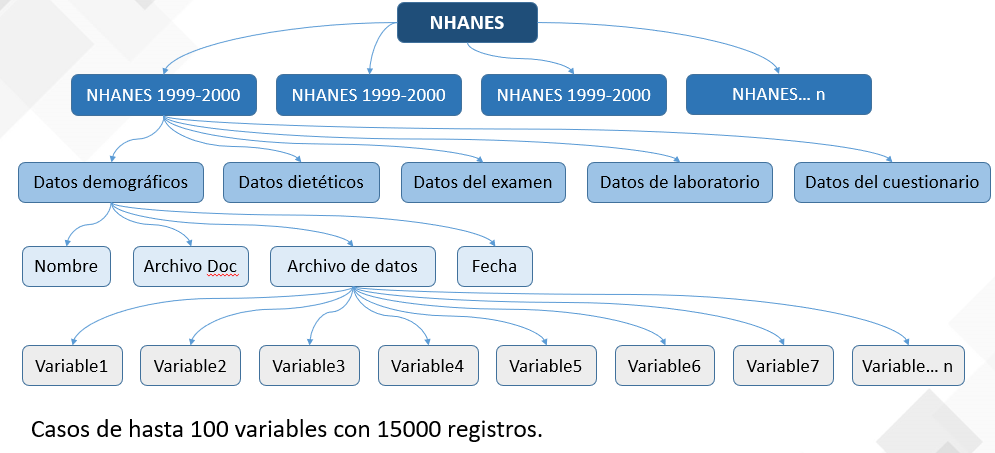


Ilustración Estructura de Datos Abiertos NHANES en línea (Fuente propia, 2019).

## **Definición de la metodología de desarrollo y datas a fusionar para la validación del software**

### **Comparación XP, SCRUM e ICONIX**

Usando el instrumento mencionado en (Jaime Andrés Britto Montoya, 2016) se compararon XP, SCRUM e ICONIX frente a las prácticas específicas de CMMI obteniendo los resultados que se pueden observar en la Tabla 5, en tonos rojos se observan cubrimientos inferiores al 30%, en tonos amarillos entre 30 y 50% y en tonos verdes cubrimientos superiores.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **N°** | **Enfoque** | **Nivel** | **Área de proceso** | **SCRUM** | **XP** | **ICONIX** |
| 1 | Administración básica de proyectos | 2 | Gestión de requerimientos (REQM) | 80% | 80% | 90% |
| 2 | 2 | Planificación de proyectos (PP) | 68% | 43% | 14% |
| 3 | 2 | Proyecto de monitoreo y control (PMC) | 88% | 55% | 0% |
| 4 | 2 | Gestión de acuerdos de proveedores (SAM) | 0% | 0% | 0% |
| 5 | 2 | Medida y análisis (MA) | 38% | 38% | 0% |
| 6 | 2 | Proceso y garantía de calidad del producto (PPQA) | 13% | 50% | 13% |
| 7 | 2 | Gestión de la configuración (CM) | 0% | 100% | 0% |
| 8 | Estandarización de procesos | 3 | Rebautizado como validar los requisitos (REQD) | 93% | 93% | 83% |
| 9 | 3 | Solución técnica (TS) | 0% | 31% | 44% |
| 10 | 3 | Integración de producto (PI) | 0% | 61% | 53% |
| 11 | 3 | Verificación (VER) | 0% | 72% | 69% |
| 12 | 3 | Validación (VAL) | 80% | 90% | 55% |
| 13 | 3 | Enfoque del proceso organizacional (OPF) | 0% | 0% | 0% |
| 14 | 3 | Definición del proceso organizacional (OPD) | 0% | 0% | 0% |
| 15 | 3 | Entrenamiento organizacional (OT) | 0% | 14% | 0% |
| 16 | 3 | Gestión integrada de proyectos (IPM) | 58% | 33% | 0% |
| 17 | 3 | Gestión de riesgos (RSKM) | 7% | 36% | 0% |
| 18 | 3 | Análisis de decisiones y resolución (DAR) | 0% | 0% | 0% |
| 19 | Administración cuantitativa | 4 | Rendimiento del proceso organizacional (OPP) | 0% | 0% | 0% |
| 20 | 4 | Gestión cuantitativa de proyectos (QPM) | 0% | 0% | 0% |
| 21 | Mejoramiento continuo de los procesos | 5 | Gestión del desempeño organizacional (OPM) | 0% | 0% | 0% |
| 22 | 5 | Análisis causal y resolución (CAR) | 0% | 0% | 0% |

Tabla Resumen cubrimiento porcentual por área de proceso de cada metodología (Jaime Andrés Britto Montoya, 2016).

Debido al buen porcentaje que tiene la metodología XP en cada enfoque, superando de manera promedio a las otras dos metodologías. La Programación Extrema se convierte en la metodología ideal para el desarrollo del sistema SoftDataFusion de manera ágil y productiva.

## **Diseño del software interactivo**

Los datos con los cuales el software trabajará será la Base de Datos Abierta NHANES, siendo manipulada por SoftDataFusion de manera local y generando nuevos datos o productos. Este proceso se puede interpretar como se muestra en la siguiente figura.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **NHANES en Línea**   * *Demographics Data* * *Dietary Data* * *Examination Data* * *Laboratory Data* * *Questionnaire Data* |  | **Proceso de Fusión** | | |  | **Exportación**  *.CSV*  *.XPT*  *.XLSX* |
|  | *NHANES*  *Datas*  *Archivos*  *Variables* |  |

Ilustración Proceso de fusión (Fuente propia, 2019).

### **Diseño simple**

A continuación, se representa gráficamente el modelo de datos que utilizará SoftDataFusion para la realización de fusiones automática según la idea del Cliente-Investigador.

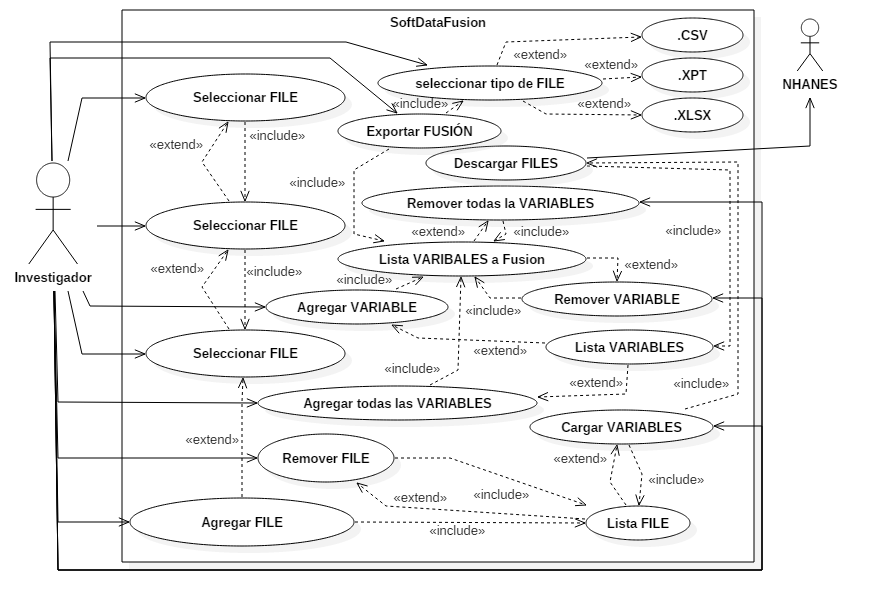


Ilustración Diseño simple del modelo de datos (Fuente propia, 2019).

### **Historias de usuario**

Las Historias de Usuario para SoftDataFusion son las siguientes:

* Selección del NHANES
* Selección de la DATA
* Selección del FILE
* Enlistar los FILES a fusionar
* Des enlistar los FILES a fusionar
* Cargar las VARIABLES de los FILES
* Enlistar todas las VARIABLES a fusionar
* Des enlistar todas las VARIABLES a fusionar
* Enlistar las VARIABLES a fusionar
* Des enlistar las VARIABLES a fusionar
* Valor opcional para las celdas sin coincidencia de fusión
* Tipo de archivo de exportación

A continuación, se tiene la tabla con detalle de cada Historia de Usuario.

|  |  |
| --- | --- |
| **Historia de Usuario** | |
| **Número:** HU01 | **Usuario:** Investigador |
| **Nombre de Historia:** Selección del NHANES | |
| **Prioridad en Negocio:** Alta | **Riesgo en Desarrollo:** Medio |
| **Puntos Estimados:** 0.83 | **Iteración Asignada:** 1 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder seleccionar un NHANES entre los disponibles de los años establecidos en las acotaciones. | |
| **Observaciones:** Ninguna | |

Tabla HU: Selección del NHANES (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Historia de Usuario** | |
| **Número:** HU02 | **Usuario:** Investigador |
| **Nombre de Historia:** Selección de la DATA | |
| **Prioridad en Negocio:** Alta | **Riesgo en Desarrollo:** Medio |
| **Puntos Estimados:** 0.83 | **Iteración Asignada:** 1 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder seleccionar una DATA entre los disponibles para el NHANES seleccionado. | |
| **Observaciones:** Ninguna | |

Tabla HU: Selección de la DATA (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Historia de Usuario** | |
| **Número:** HU03 | **Usuario:** Investigador |
| **Nombre de Historia:** Selección del FILE | |
| **Prioridad en Negocio:** Alta | **Riesgo en Desarrollo:** Medio |
| **Puntos Estimados:** 0.83 | **Iteración Asignada:** 1 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder seleccionar una FILE entre los disponibles para la DATA seleccionada. | |
| **Observaciones:** Ninguna | |

Tabla HU: Selección del FILE (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Historia de Usuario** | |
| **Número:** HU04 | **Usuario:** Investigador |
| **Nombre de Historia:** Enlistar los FILES a fusionar | |
| **Prioridad en Negocio:** Alta | **Riesgo en Desarrollo:** Medio |
| **Puntos Estimados:** 0.83 | **Iteración Asignada:** 1 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder agregar a una lista los FILES que se quieren fusionar. | |
| **Observaciones:** Ninguna | |

Tabla HU: Enlistar los FILES a fusionar (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Historia de Usuario** | |
| **Número:** HU05 | **Usuario:** Investigador |
| **Nombre de Historia:** Des enlistar los FILES a fusionar | |
| **Prioridad en Negocio:** Alta | **Riesgo en Desarrollo:** Medio |
| **Puntos Estimados:** 0.83 | **Iteración Asignada:** 1 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder remover de la lista los FILES que no se quieren fusionar. | |
| **Observaciones:** Ninguna | |

Tabla HU: Des enlistar los FILES a fusionar (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Historia de Usuario** | |
| **Número:** HU06 | **Usuario:** Investigador |
| **Nombre de Historia:** Cargar las VARIABLES de los FILES | |
| **Prioridad en Negocio:** Alta | **Riesgo en Desarrollo:** Alto |
| **Puntos Estimados:** 0.83 | **Iteración Asignada:** 1 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder visualizar en una lista las VARIABLES de loas FILES a fusionar. | |
| **Observaciones:** Ninguna | |

Tabla HU: Cargar las VARIABLES de los FILES (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Historia de Usuario** | |
| **Número:** HU07 | **Usuario:** Investigador |
| **Nombre de Historia:** Enlistar todas las VARIABLES a fusionar | |
| **Prioridad en Negocio:** Alta | **Riesgo en Desarrollo:** Medio |
| **Puntos Estimados:** 0.83 | **Iteración Asignada:** 1 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder agregar a una lista todas las VARIABLES de los FILES a fusionar. | |
| **Observaciones:** Ninguna | |

Tabla HU: Enlistar todas las VARIABLES a fusionar (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Historia de Usuario** | |
| **Número:** HU08 | **Usuario:** Investigador |
| **Nombre de Historia:** Des enlistar todas las VARIABLES a fusionar | |
| **Prioridad en Negocio:** Alta | **Riesgo en Desarrollo:** Medio |
| **Puntos Estimados:** 0.83 | **Iteración Asignada:** 1 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder remover de la lista todas las VARIABLES de los FILES a fusionar. | |
| **Observaciones:** Ninguna | |

Tabla HU: Des enlistar todas las VARIABLES a fusionar (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Historia de Usuario** | |
| **Número:** HU09 | **Usuario:** Investigador |
| **Nombre de Historia:** Enlistar las VARIABLES a fusionar | |
| **Prioridad en Negocio:** Alta | **Riesgo en Desarrollo:** Medio |
| **Puntos Estimados:** 0.83 | **Iteración Asignada:** 1 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder agregar de la lista todas las VARIABLES de los FILES a fusionar. | |
| **Observaciones:** Ninguna | |

Tabla HU: Enlistar las VARIABLES a fusionar (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Historia de Usuario** | |
| **Número:** HU10 | **Usuario:** Investigador |
| **Nombre de Historia:** Des enlistar las VARIABLES a fusionar | |
| **Prioridad en Negocio:** Alta | **Riesgo en Desarrollo:** Medio |
| **Puntos Estimados:** 0.83 | **Iteración Asignada:** 1 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder remover de la lista todas las VARIABLES de los FILES a fusionar. | |
| **Observaciones:** Ninguna | |

Tabla HU: Des enlistar las VARIABLES a fusionar (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Historia de Usuario** | |
| **Número:** HU11 | **Usuario:** Investigador |
| **Nombre de Historia:** Valor opcional para las celdas sin coincidencia de fusión | |
| **Prioridad en Negocio:** Alta | **Riesgo en Desarrollo:** Alto |
| **Puntos Estimados:** 0.83 | **Iteración Asignada:** 1 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder digitar un valor el cual tomen las casillas que no tengan coincidencia al fisionar distintos NHANES, DATAS y FILES. | |
| **Observaciones:** Ninguna | |

Tabla HU: Valor opcional para las celdas sin coincidencia de fusión (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Historia de Usuario** | |
| **Número:** HU12 | **Usuario:** Investigador |
| **Nombre de Historia:** Tipo de archivo de exportación | |
| **Prioridad en Negocio:** Alta | **Riesgo en Desarrollo:** Alto |
| **Puntos Estimados:** 0.83 | **Iteración Asignada:** 1 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder exportar la nueva fusión a cualquier tipo de archivo que se haya especificado en las acotaciones. | |
| **Observaciones:** Ninguna | |

Tabla HU: Tipo de archivo de exportación (Fuente propia, 2019).

### **Acotaciones del software**

* Las datas que estarán disponible en el software serán las que son NHANES continuo y han sido casos de estudios transversales, desde el NHANES 1999-2000 al NHANES 2015-2016.
* El software será ejecutable en computadores de mesa o portátiles, en los cuales corran sistemas operativos de Windows o Linux.
* La cantidad de datas fisionables en simultaneidad dependerá de la cantidad de variables seleccionadas por data y los recursos disponibles del sistema, memoria RAM y el Procesador.
* Es necesaria la conexión a Internet si previamente no se ponen a disposición del software las datas requeridas.
* El tipo de archivo de lectura admitido por el software será .XPT, que es el mismo con el cual NHANES comparte sus datas en su sitio web oficial.
* Los tipos de archivo de exportación que tendrá el software serán .XPT, .CSV Y .XLSX

### **Asignación de roles**

A continuación, se tabula de manera ordenada las personas involucradas con sus respectivos roles.

|  |  |
| --- | --- |
| **Roles** | **Asignado a:** |
| ***Programador:*** | *Jorge Miguel Anaya León* |
| ***Cliente:*** | *PhD Valmore Bermúdez - Investigadores* |
| ***Encargado de Pruebas (Tester):*** | *Jorge Miguel Anaya León* |
| ***Encargado de Seguimiento (Tracker):*** | *Jorge Miguel Anaya León* |
| ***Entrenador (Coach):*** | *Jorge Miguel Anaya León* |
| ***Consultor:*** | *PhD Valmore Bermúdez* |
| ***Gestor (Big Boss):*** | *Msc Gerardo Chacón* |

Tabla Asignación de Roles (Fuente propia, 2019).

### **Plan de entrega**

Con base a las Historias de Usuarios se elabora el siguiente Plan de Entrega, teniendo en cuenta las fechas, prioridad y esfuerzo de cada HU.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Historias de Usuario** | **Iteración** | **Prioridad** | **Esfuerzo** | **Fecha Inicio** | **Fecha Final** |
| HU1 | 1 | Alta | 0.83 | 15/09/2019 | 19/09/2019 |
| HU2 | 1 | Alta | 0.83 | 19/09/2019 | 27/09/2019 |
| HU3 | 1 | Alta | 0.83 | 27/09/2019 | 01/10/2019 |
| HU4 | 1 | Alta | 0.83 | 01/10/2019 | 05/10/2019 |
| HU5 | 1 | Alta | 0.83 | 05/10/2019 | 09/10/2019 |
| HU6 | 1 | Alta | 0.83 | 09/10/2019 | 13/10/2019 |
| HU7 | 1 | Alta | 0.83 | 13/10/2019 | 17/10/2019 |
| HU8 | 1 | Alta | 0.83 | 17/10/2019 | 21/10/2019 |
| HU9 | 1 | Alta | 0.83 | 21/10/2019 | 25/10/2019 |
| HU10 | 1 | Alta | 0.83 | 25/10/2019 | 29/10/2019 |
| HU11 | 1 | Alta | 0.83 | 29/10/2019 | 03/11/2019 |
| HU12 | 1 | Alta | 0.83 | 03/11/2019 | 07/11/2019 |

Tabla Plan de Entrega (Fuente propia, 2019).

### **Tareas de ingenierías (Task Card)**

|  |  |  |
| --- | --- | --- |
| **N° TI** | **N° HU** | **Nombre de la Tarea de Ingeniería** |
| TI01 | HU01 | Posicionamiento en la NHANES |
| TI02 | HU02 | Posicionamiento en la GUI y código funcional para la DATA |
| TI03 | HU03 | Posicionamiento en la FILE |
| TI04 | HU04 | Posicionamiento en la Enlistar los FILES a fusionar |
| TI05 | HU05 | Posicionamiento en la Des enlistar los FILES a fusionar |
| TI06 | HU06 | Posicionamiento en la Cargar las VARIABLES de los FILES |
| TI07 | HU07 | Posicionamiento en la Enlistar todas las VARIABLES a fusionar |
| TI08 | HU08 | Posicionamiento en la Des enlistar todas las VARIABLES a fusionar |
| TI09 | HU09 | Posicionamiento en la Enlistar las VARIABLES a fusionar |
| TI10 | HU10 | Posicionamiento en la Des enlistar las VARIABLES a fusionar |
| TI11 | HU11 | Posicionamiento en la Valor opcional para las celdas sin coincidencia de fusión |
| TI12 | HU12 | Posicionamiento en la Tipo de archivo de exportación |

Tabla Tareas de Ingeniería (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Tarea de Ingeniería** | |
| **Número de Tarea:** TI01 | **Numero de Historia:** HU01 |
| **Nombre de Tarea:** Posicionamiento en la NHANES | |
| **Tipo de Tarea:** Desarrollo | **Puntos Estimados:** 0.83 |
| **Fecha Inicio:** 15/09/2019 | **Fecha Fin:** 19/09/2019 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder visualizar de forma sencilla en dónde y cómo seleccionar el NHANES a utilizar. | |

Tabla TI: Seleccionar el NHANES (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Tarea de Ingeniería** | |
| **Número de Tarea:** TI02 | **Numero de Historia:** HU02 |
| **Nombre de Tarea:** Posicionamiento en la GUI y código funcional para la DATA | |
| **Tipo de Tarea:** Desarrollo | **Puntos Estimados:** 0.83 |
| **Fecha Inicio:** 19/09/2019 | **Fecha Fin:** 27/09/2019 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder visualizar de forma sencilla en dónde y cómo seleccionar la DATA a utilizar. | |

Tabla TI: Seleccionar la DATA (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Tarea de Ingeniería** | |
| **Número de Tarea:** TI03 | **Numero de Historia:** HU03 |
| **Nombre de Tarea:** Posicionamiento en la FILE | |
| **Tipo de Tarea:** Desarrollo | **Puntos Estimados:** 0.83 |
| **Fecha Inicio:** 27/09/2019 | **Fecha Fin:** 01/10/2019 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder visualizar de forma sencilla en dónde y cómo seleccionar el FILE a utilizar. | |

Tabla TI: Seleccionar el FILE (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Tarea de Ingeniería** | |
| **Número de Tarea:** TI04 | **Numero de Historia:** HU04 |
| **Nombre de Tarea:** Posicionamiento en la Enlistar los FILES a fusionar | |
| **Tipo de Tarea:** Desarrollo | **Puntos Estimados:** 0.83 |
| **Fecha Inicio:** 01/10/2019 | **Fecha Fin:** 05/10/2019 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder visualizar de forma sencilla en dónde y cómo Enlistar los FILES a fusionar. | |

Tabla TI: Enlistar los FILES a fusionar (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Tarea de Ingeniería** | |
| **Número de Tarea:** TI05 | **Numero de Historia:** HU05 |
| **Nombre de Tarea:** Posicionamiento en la Des enlistar los FILES a fusionar | |
| **Tipo de Tarea:** Desarrollo | **Puntos Estimados:** 0.83 |
| **Fecha Inicio:** 05/10/2019 | **Fecha Fin:** 09/10/2019 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder visualizar de forma sencilla en dónde y cómo Des enlistar los FILES a fusionar. | |

Tabla TI: Des enlistar los FILES a fusionar (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Tarea de Ingeniería** | |
| **Número de Tarea:** TI06 | **Numero de Historia:** HU06 |
| **Nombre de Tarea:** Posicionamiento en la Cargar las VARIABLES de los FILES | |
| **Tipo de Tarea:** Desarrollo | **Puntos Estimados:** 0.83 |
| **Fecha Inicio:** 09/10/2019 | **Fecha Fin:** 13/10/2019 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder visualizar de forma sencilla en dónde y cómo Cargar las VARIABLES de los FILES. | |

Tabla TI: Cargar las VARIABLES de los FILES (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Tarea de Ingeniería** | |
| **Número de Tarea:** TI07 | **Numero de Historia:** HU07 |
| **Nombre de Tarea:** Posicionamiento en la Enlistar todas las VARIABLES a fusionar | |
| **Tipo de Tarea:** Desarrollo | **Puntos Estimados:** 0.83 |
| **Fecha Inicio:** 13/10/2019 | **Fecha Fin:** 17/10/2019 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder visualizar de forma sencilla en dónde y cómo Enlistar todas las VARIABLES a fusionar. | |

Tabla TI: Enlistar todas las VARIABLES a fusionar (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Tarea de Ingeniería** | |
| **Número de Tarea:** TI08 | **Numero de Historia:** HU08 |
| **Nombre de Tarea:** Posicionamiento en la Des enlistar todas las VARIABLES a fusionar | |
| **Tipo de Tarea:** Desarrollo | **Puntos Estimados:** 0.83 |
| **Fecha Inicio:** 17/10/2019 | **Fecha Fin:** 21/10/2019 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder visualizar de forma sencilla en dónde y cómo Des enlistar todas las VARIABLES a fusionar. | |

Tabla TI: Des enlistar todas las VARIABLES a fusionar (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Tarea de Ingeniería** | |
| **Número de Tarea:** TI09 | **Numero de Historia:** HU09 |
| **Nombre de Tarea:** Posicionamiento en la Enlistar las VARIABLES a fusionar | |
| **Tipo de Tarea:** Desarrollo | **Puntos Estimados:** 0.83 |
| **Fecha Inicio:** 21/10/2019 | **Fecha Fin:** 25/10/2019 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder visualizar de forma sencilla en dónde y cómo Enlistar las VARIABLES a fusionar. | |

Tabla TI: Enlistar las VARIABLES a fusionar (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Tarea de Ingeniería** | |
| **Número de Tarea:** TI10 | **Numero de Historia:** HU10 |
| **Nombre de Tarea:** Posicionamiento en la Des enlistar las VARIABLES a fusionar | |
| **Tipo de Tarea:** Desarrollo | **Puntos Estimados:** 0.83 |
| **Fecha Inicio:** 25/10/2019 | **Fecha Fin:** 29/10/2019 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder visualizar de forma sencilla en dónde y cómo Des enlistar las VARIABLES a fusionar. | |

Tabla TI: Des enlistar las VARIABLES a fusionar (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Tarea de Ingeniería** | |
| **Número de Tarea:** TI11 | **Numero de Historia:** HU11 |
| **Nombre de Tarea:** Posicionamiento en la Valor opcional para las celdas sin coincidencia de fusión | |
| **Tipo de Tarea:** Desarrollo | **Puntos Estimados:** 0.83 |
| **Fecha Inicio:** 29/10/2019 | **Fecha Fin:** 03/11/2019 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder visualizar de forma sencilla en dónde y cómo Valor opcional para las celdas sin coincidencia de fusión. | |

Tabla TI: Valor opcional para las celdas sin coincidencia de fusión (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Tarea de Ingeniería** | |
| **Número de Tarea:** TI11 | **Numero de Historia:** HU11 |
| **Nombre de Tarea:** Posicionamiento en la Tipo de archivo de exportación | |
| **Tipo de Tarea:** Desarrollo | **Puntos Estimados:** 0.83 |
| **Fecha Inicio:** 03/11/2019 | **Fecha Fin:** 07/11/2019 |
| **Programador Responsable:** Jorge Miguel Anaya León | |
| **Descripción:** Se debería poder visualizar de forma sencilla en dónde y cómo Tipo de archivo de exportación. | |

Tabla TI: Tipo de archivo de exportación (Fuente propia, 2019).

### **Tarjetas CRC (Clase – Responsabilidades – Colaboradores)**

|  |  |
| --- | --- |
| **Tarea de Ingeniería** | |
| **Nombre de la Clase:** SoftDataFusion | |
| **Responsabilidades:**   * **Atributos** * import os.path * import pandas * import datetime * import time * import tkinter as tk * from tkinter import \* * from tkinter import font * from tkinter import messagebox * from tkinter import ttk * from tkinter.font import Font * from tkinter.font import nametofont * import urllib.request * import xport * NHANES\_DATA * **Operaciones** * def FuseData(self, InfoVariable, default, exportto): return DFS * def DownloadFile(WebAddress, LocalAddress): * def IsInternetUp(): return boolean * def Licencia(): return Boolean * def ItemsColors(self): * def LoadVariables(self): | **Colaboradores:**   * Ninguno |

Tabla Tarjeta CRC: SoftDataFusion (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Tarea de Ingeniería** | |
| **Nombre de la Clase:** CreateToolTip | |
| **Responsabilidades:**   * **Atributos** * self.waittime * self.wraplength * self.widget * self.text * self.id * self.tw * **Operaciones** * def enter(self, event=None): * def leave(self, event=None): * def schedule(self): * def unschedule(self): * def showtip(self, event=None): * def hidetip(self): | **Colaboradores:**   * Ninguno |

Tabla Tarjeta CRC: CreateToolTip (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Tarea de Ingeniería** | |
| **Nombre de la Clase:** Linkbutton | |
| **Responsabilidades:**   * **Atributos** * label\_font * self.font * style * **Operaciones** * def on\_mouse\_enter(self, event): * def on\_mouse\_leave(self, event): | **Colaboradores:**   * Ninguno |

Tabla Tarjeta CRC: Linkbutton (Fuente propia, 2019).

|  |  |  |
| --- | --- | --- |
| **Tarea de Ingeniería** | | |
| **Nombre de la Clase:** Application | | |
| **Responsabilidades:**   * **Atributos** | | **Colaboradores:**   * SoftDataFusion * CreateToolTip * Linkbutton |
| * WidthWindow * HeightWindow * BackgroundWindow * BackgroundEvents * self.MnBar * self.Mn0 * self.Add00 * self.Move00 * self.Load00 * self.Mn1 * self.AddAll01 * self.Add01 * self.Move01 * self.MoveAll01 * self.ValCel01 * self.Mn2 * self.Merg02 * self.NewM02 * self.Mn3 * self.Tuto03 * self.Abou03 * f1 * f2 = 60 * f3 = 90 * f4 = 160 * self.Canv1 * self.Labe1 * self.Comb1 * self.Labe2 * self.Comb2 * self.Labe3 * self.Comb3 | * self.List1 * self.Canv2 * self.Butt1 * self.Butt2 * self.Canv3 * self.Canv4 * self.Canv41 * self.Comb4 * self.Butt3 * self.Butt4 * self.Canv6 * self.Butt8 * self.Canv7 * self.Butt9 * self.Butt10 * self.Butt11 * self.Butt12 * self.Entry1 * self.Labe3 * self.Labe4 * self.Labe5 * self.List2 * self.List3 * self.Canv5 * self.LoImg * self.JButton111 * self.ImgAct * self.Fondo * self.Back * self.Next * self.CloseT |
| * **Operaciones** * def Tutorial(self, \* event): * def CloseTutorial(self, \* event): * def NextTutorial(self, \* event): * def BackTutorial(self, \* event): * def About(self, \* event): * def link\_clicked(self): * def ChangeComb1(self, event): * def ChangeComb2(self, event): * def OppreButt1(self, \* event): * def OppreButt2(self, \* event): * def OppreButt3(self, \* event): * def OppreButt4(self, \* event): * def OppreButt8(self, \* event): * def OppreButt9(self, \* event): * def OppreButt10(self, \* event): * def OppreButt11(self, \* event): * def OppreButt12(self, \* event): * def Copy(self, \* event): | |

Tabla Tarjeta CRC: Application (Fuente propia, 2019).

## **Desarrollo del software interactivo**

El software interactivo SoftDataFusion fue desarrollado en el entorno de desarrollo integrado NetBeans, el cual se adecuo para la interpretación del lenguaje Python mediante librerías adicionales. Las fases de desarrollo fueron:

### **Fase 1 (Preparación del entorno de desarrollo NetBeans con Python)**

En esta fase se prepara el entorno de desarrollo instalando librerías como manejador de paquetes desde la consola de comandos, librerías para la lectura de archivos, variables de entono.

1. Instalar Python
2. Crear su variable de entorno
3. Descargar e instalar get-pip.py
4. *pip install xport*
5. *pip install pandas*
6. *pip install openpyxl*
7. *pip install --upgrade certifi*
8. Agregar el Plugin de Python a NetBeans
9. Configurar plataforma de Python en NetBeans

### **Fase 2 (Desarrollo de SoftDataFusion, GUI)**

El tipo de programación implementada para el desarrollo de SoftDataFusion ha sido funcional, las cuales de manera estratégica se dividieron de acuerdo a su propósito.

Funciones para la modificación de la única interfaz del programa permitiendo a esta ser una ventana multipropósito y funciones para la agilización del uso del software con atajos de teclado, Ilustración 13.

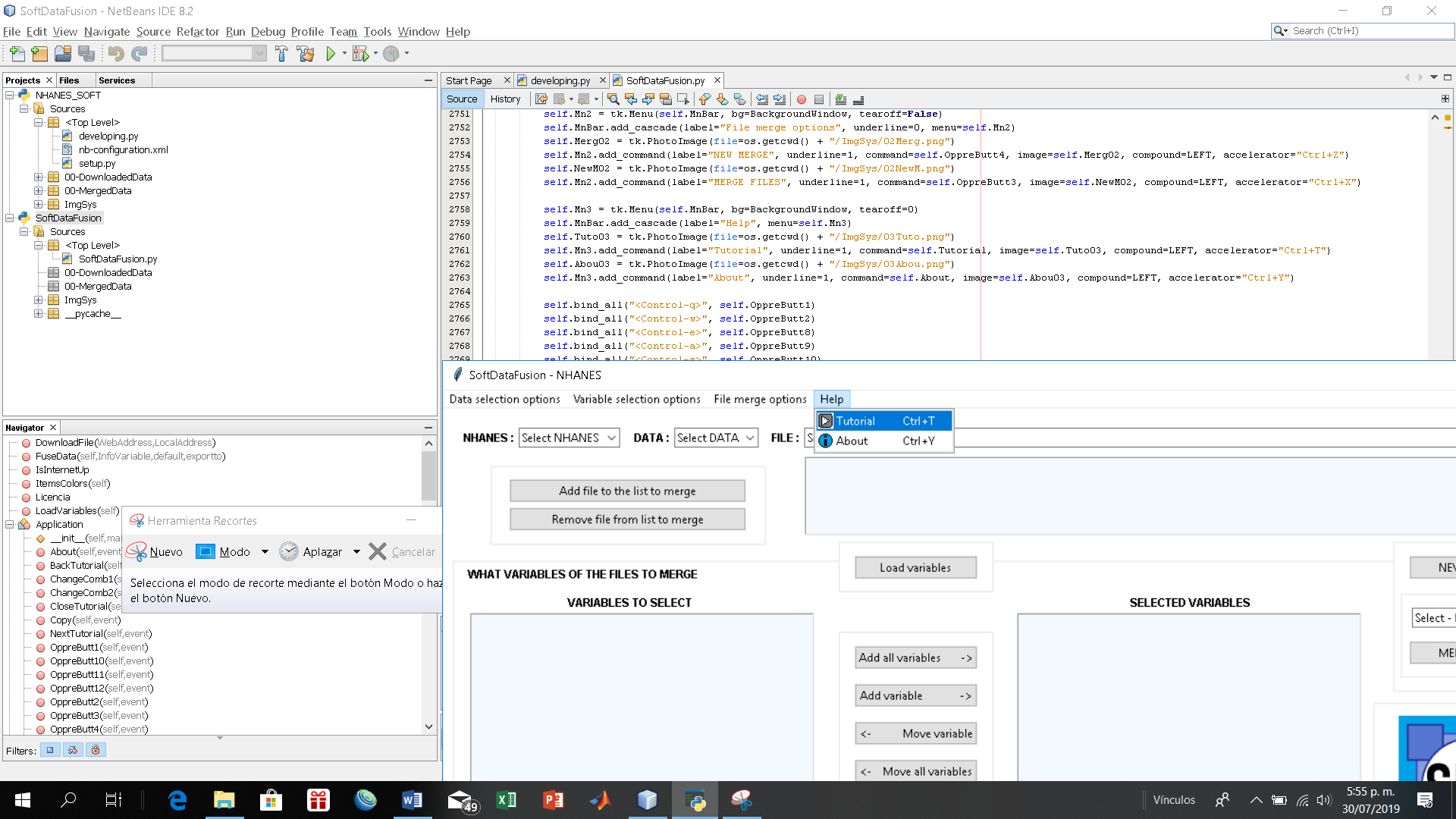


Ilustración Ventana multipropósito y ágil uso del software (Fuente propia, 2019).

Las Funciones para la obtención de la data en línea de NHANES se muestra en la Ilustración 14.

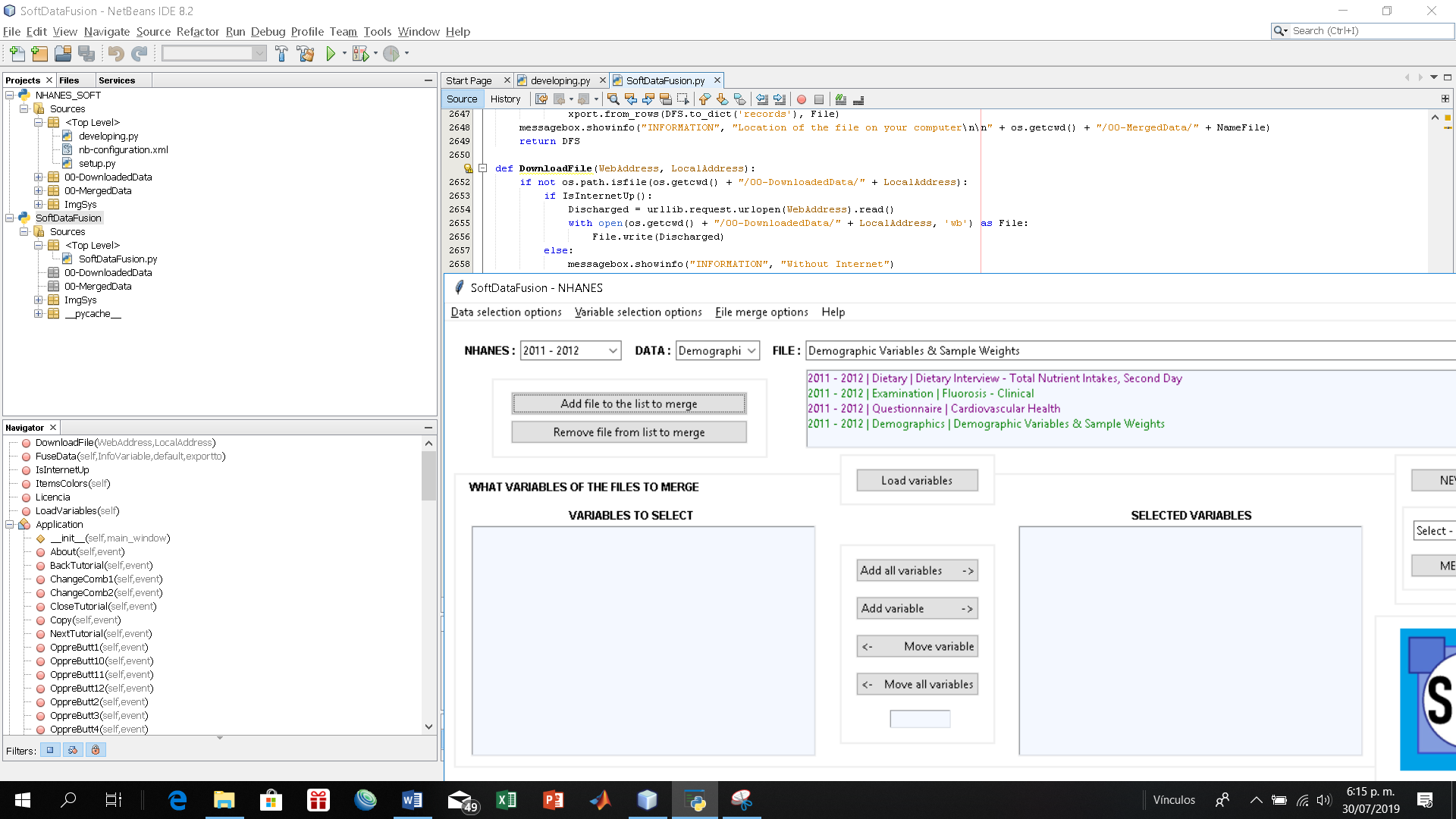


Ilustración Obtención de la data en línea de NHANES (Fuente propia, 2019).

Funciones de validación, funciones para la lectura de archivos .XPT Ilustración 15.

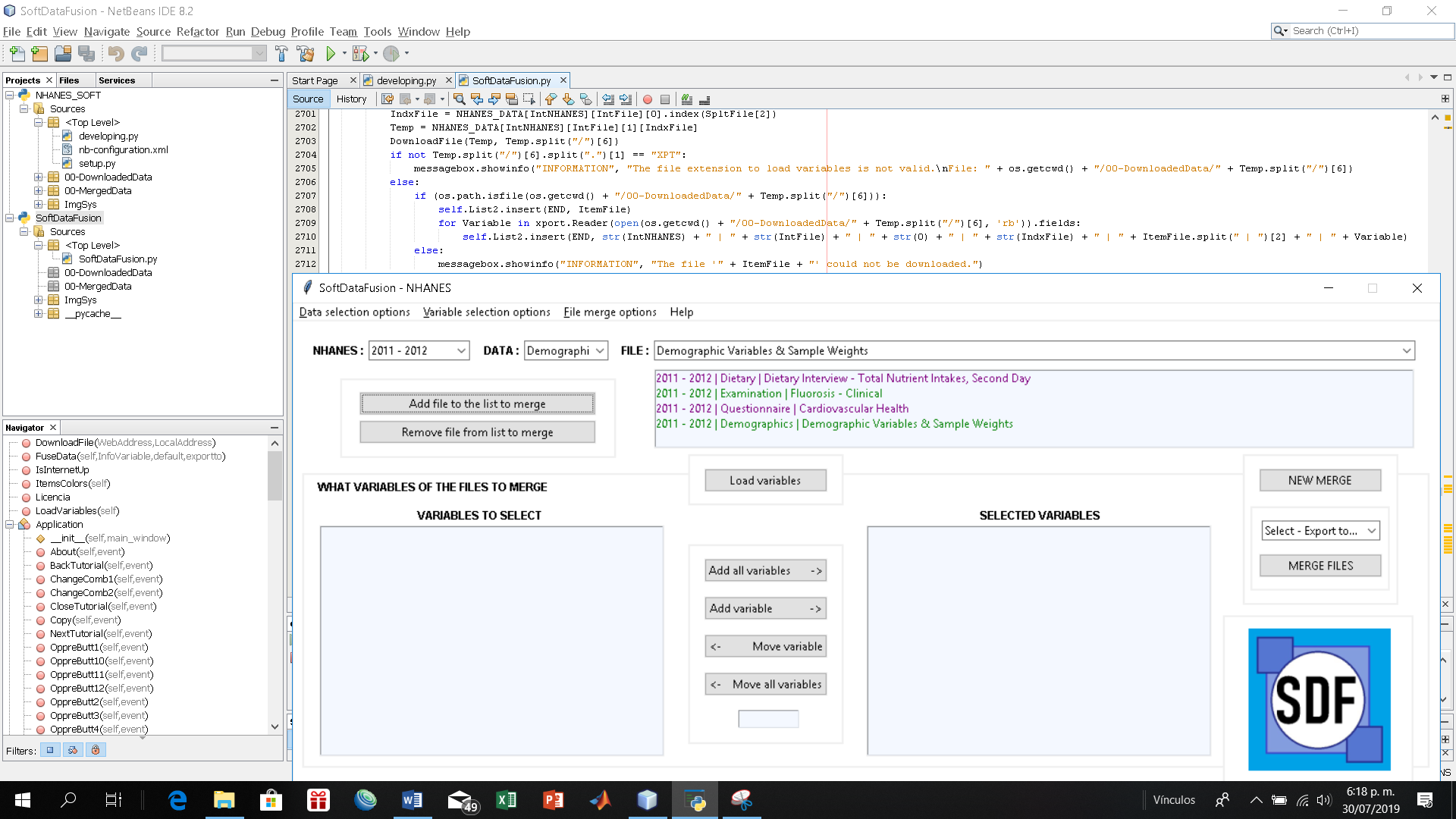


Ilustración Lectura de archivos .XPT (Fuente propia, 2019).

Funciones para la fusión de las datas y la escritura de archivos en los diferentes formatos establecidos en el levantamiento de requerimientos para el caso de las bases de datos demográficos y base de datos de laboratorio se tomaron Variables demográficas, pesos de muestra, Insulina, Glucosa en ayunas de plasma se muestran en la Ilustración 16.

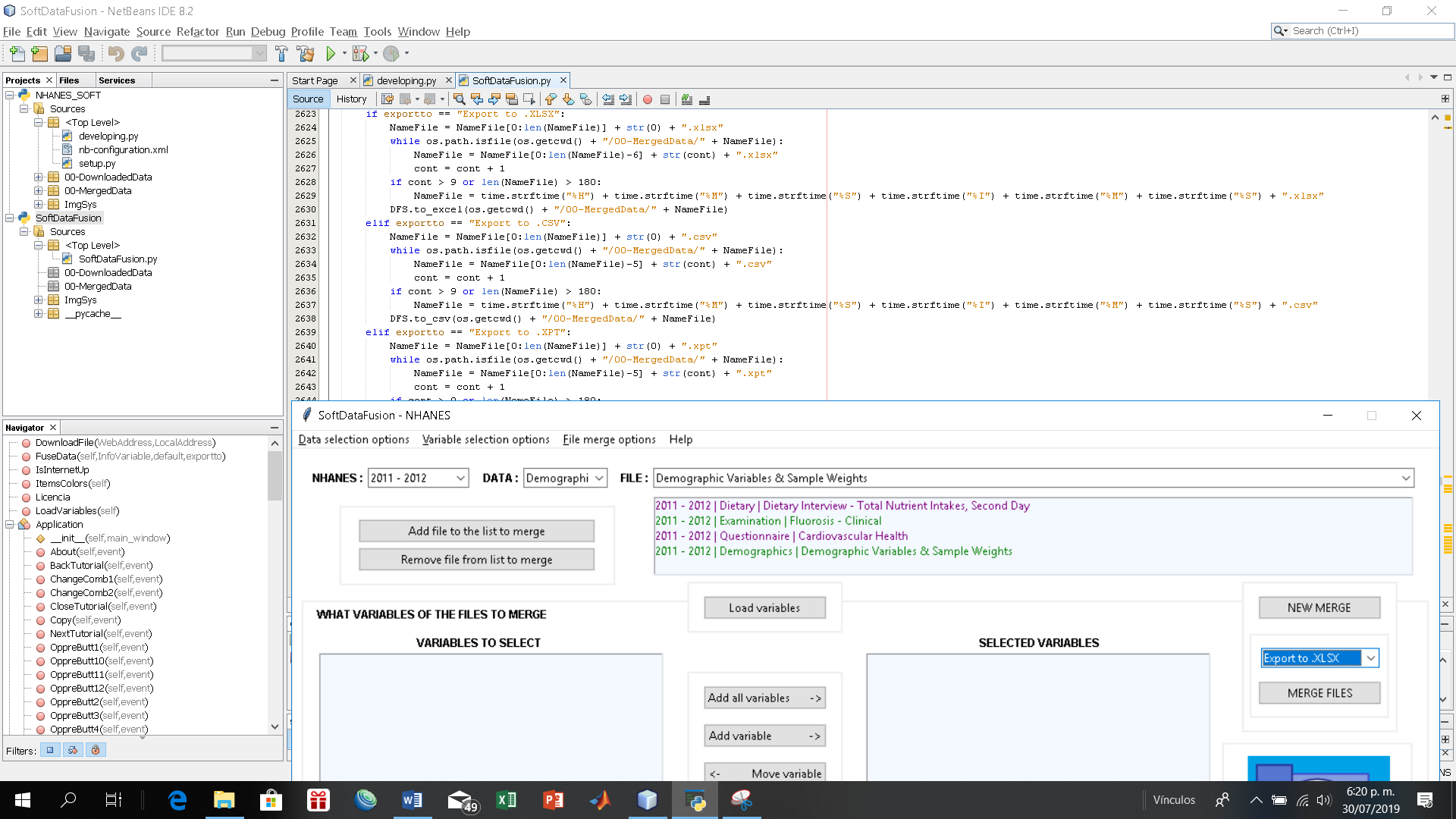


Ilustración Fusión de las datas y la escritura de archivos (Fuente propia, 2019).

## **Pruebas de aceptación**

|  |  |  |
| --- | --- | --- |
| **N° PA** | **N° HU** | **Nombre de la Tarea de Ingeniería** |
| PA01 | HU01 | Selección del NHANES |
| PA02 | HU02 | Selección de la DATA |
| PA03 | HU03 | Selección del FILE |
| PA04 | HU04 | Enlistar los FILES a fusionar |
| PA05 | HU05 | Des enlistar los FILES a fusionar |
| PA06 | HU06 | Cargar las VARIABLES de los FILES |
| PA07 | HU07 | Enlistar todas las VARIABLES a fusionar |
| PA08 | HU08 | Des enlistar todas las VARIABLES a fusionar |
| PA09 | HU09 | Enlistar las VARIABLES a fusionar |
| PA10 | HU10 | Des enlistar las VARIABLES a fusionar |
| PA11 | HU11 | Valor opcional para las celdas sin coincidencia de fusión |
| PA12 | HU12 | Tipo de archivo de exportación |

Tabla Pruebas de Aceptación (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Prueba de Aceptación** | |
| **Número de Prueba:** PA01 | **Numero de Historia:** HU01 |
| **HU:** Selección del NHANES | |
| **Condiciones de Ejecución:** Ejecutar el programa. | |
| **Entrada/Pasos de Ejecución:** Buscar SoftDataFusion en nuestro Sistema Operativo y ejecutarlo. | |
| **Resultado Esperado:** Ver una GUI con el contenido de la Historia de Usuario. | |
| **Evaluación de la Prueba:** Aprobada | |

Tabla Prueba de Aceptación: PA01 (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Prueba de Aceptación** | |
| **Número de Prueba:** PA02 | **Numero de Historia:** HU02 |
| **HU:** Selección de la DATA | |
| **Condiciones de Ejecución:** Seleccionar un NHANES. | |
| **Entrada/Pasos de Ejecución:** Desplegar el combo de NHANES y seleccionar uno. | |
| **Resultado Esperado:** Ver una GUI con el contenido de la Historia de Usuario. | |
| **Evaluación de la Prueba:** Aprobada | |

Tabla Prueba de Aceptación: PA02 (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Prueba de Aceptación** | |
| **Número de Prueba:** PA03 | **Numero de Historia:** HU03 |
| **HU:** Selección del FILE | |
| **Condiciones de Ejecución:** Seleccionar un NHANES y una DATA. | |
| **Entrada/Pasos de Ejecución:** Desplegar el combo de DATAS y seleccionar una. | |
| **Resultado Esperado:** Ver una GUI con el contenido de la Historia de Usuario. | |
| **Evaluación de la Prueba:** Aprobada | |

Tabla Prueba de Aceptación: PA03 (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Prueba de Aceptación** | |
| **Número de Prueba:** PA04 | **Numero de Historia:** HU04 |
| **HU:** Enlistar los FILES a fusionar | |
| **Condiciones de Ejecución:** Seleccionar un NHANES, una DATA y un FILE. | |
| **Entrada/Pasos de Ejecución:** Desplegar el combo de FILES y seleccionar uno. Para luego a agregarlo a la lista. | |
| **Resultado Esperado:** Ver una GUI con el contenido de la Historia de Usuario. | |
| **Evaluación de la Prueba:** Aprobada | |

Tabla Prueba de Aceptación: PA04 (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Prueba de Aceptación** | |
| **Número de Prueba:** PA05 | **Numero de Historia:** HU05 |
| **HU:** Des enlistar los FILES a fusionar | |
| **Condiciones de Ejecución:** Seleccionar un FILE de la lista. | |
| **Entrada/Pasos de Ejecución:** Seleccionar un FILE de la lista y luego dar en el botón de Remover FILE de la lista a fusionar. | |
| **Resultado Esperado:** Ver una GUI con el contenido de la Historia de Usuario. | |
| **Evaluación de la Prueba:** Aprobada | |

Tabla Prueba de Aceptación: PA05 (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Prueba de Aceptación** | |
| **Número de Prueba:** PA06 | **Numero de Historia:** HU06 |
| **HU:** Cargar las VARIABLES de los FILES | |
| **Condiciones de Ejecución:** Se deben hacer seleccionado y agregado FILES a la lista. También | |
| **Entrada/Pasos de Ejecución:** Dar clic en el botón de cargar VARIABLES. | |
| **Resultado Esperado:** Ver una GUI con el contenido de la Historia de Usuario. | |
| **Evaluación de la Prueba:** Aprobada | |

Tabla Prueba de Aceptación: PA06 (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Prueba de Aceptación** | |
| **Número de Prueba:** PA07 | **Numero de Historia:** HU07 |
| **HU:** Enlistar todas las VARIABLES a fusionar | |
| **Condiciones de Ejecución:** Haber VARIABLES cargadas para los FILES enlistados. | |
| **Entrada/Pasos de Ejecución:** Hay que dar clic en el botón Agregar todas las VARIABLES. | |
| **Resultado Esperado:** Ver una GUI con el contenido de la Historia de Usuario. | |
| **Evaluación de la Prueba:** Aprobada | |

Tabla Prueba de Aceptación: PA07 (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Prueba de Aceptación** | |
| **Número de Prueba:** PA08 | **Numero de Historia:** HU08 |
| **HU:** Des enlistar todas las VARIABLES a fusionar | |
| **Condiciones de Ejecución:** Haber VARIABLES agregadas a la lista final de fusión. | |
| **Entrada/Pasos de Ejecución:** Hay que dar clic en el botón Remover todas las VARIABLES. | |
| **Resultado Esperado:** Ver una GUI con el contenido de la Historia de Usuario. | |
| **Evaluación de la Prueba:** Aprobada | |

Tabla Prueba de Aceptación: PA08 (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Prueba de Aceptación** | |
| **Número de Prueba:** PA09 | **Numero de Historia:** HU09 |
| **HU:** Enlistar las VARIABLES a fusionar | |
| **Condiciones de Ejecución:** Haber VARIABLES cargadas para los FILES enlistados. | |
| **Entrada/Pasos de Ejecución:** Hay que dar clic en el botón Agregar la VARIABLE. | |
| **Resultado Esperado:** Ver una GUI con el contenido de la Historia de Usuario. | |
| **Evaluación de la Prueba:** Aprobada | |

Tabla Prueba de Aceptación: PA09 (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Prueba de Aceptación** | |
| **Número de Prueba:** PA10 | **Numero de Historia:** HU10 |
| **HU:** Des enlistar las VARIABLES a fusionar | |
| **Condiciones de Ejecución:** Haber VARIABLES agregadas a la lista final de fusión. | |
| **Entrada/Pasos de Ejecución:** Hay que dar clic en el botón Remover la VARIABLE. | |
| **Resultado Esperado:** Ver una GUI con el contenido de la Historia de Usuario. | |
| **Evaluación de la Prueba:** Aprobada | |

Tabla Prueba de Aceptación: PA10 (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Prueba de Aceptación** | |
| **Número de Prueba:** PA11 | **Numero de Historia:** HU11 |
| **HU:** Valor opcional para las celdas sin coincidencia de fusión | |
| **Condiciones de Ejecución:** Es opcional colocar este valor, por defecto es SDF (SoftDataFusion). | |
| **Entrada/Pasos de Ejecución:** Hay que digitar un valor en la caja detesto que se encuentra entre las listas de VARIABLES cargadas y VARIABLES a fusionar. | |
| **Resultado Esperado:** Ver una GUI con el contenido de la Historia de Usuario. | |
| **Evaluación de la Prueba:** Aprobada | |

Tabla Prueba de Aceptación: PA011 (Fuente propia, 2019).

|  |  |
| --- | --- |
| **Prueba de Aceptación** | |
| **Número de Prueba:** PA12 | **Numero de Historia:** HU12 |
| **HU:** Tipo de archivo de exportación | |
| **Condiciones de Ejecución:** Haber VARIABLES en la lista de VARIABLES a fusionar. | |
| **Entrada/Pasos de Ejecución:** Hay que digitar un valor en la caja detesto que se encuentra entre las listas de VARIABLES cargadas y VARIABLES a fusionar. | |
| **Resultado Esperado:** Ver una GUI con el contenido de la Historia de Usuario. | |
| **Evaluación de la Prueba:** Aprobada | |

Tabla Prueba de Aceptación: PA012 (Fuente propia, 2019).

## **Análisis e interpretación de resultados**

Las datas de fusiones realizadas por SoftDataFusion son archivos con el mismo nombre de los archivos fusionados y exportados a un cierto formato, la representación de estos es en una tabla la cual consta del número de sujetos y las variables que se avaluaron para este.

A continuación, se puede visualizar las datas generadas por SoftDataFusion de acuerdo a su extensión:

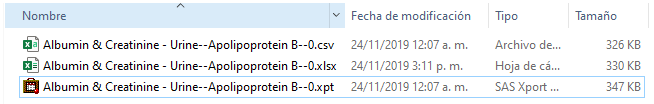


Ilustración Extensiones de exportación de archivos (Fuente propia, 2019).

### **Archivo con extensión .CSV**

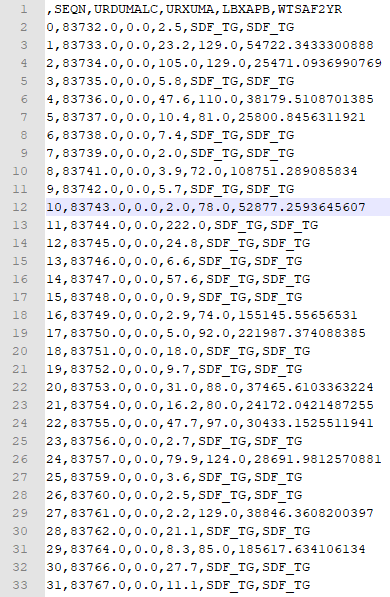


Ilustración Data generada en extensión .CSV (Fuente propia, 2019).

### **Archivo con extensión .XPT**

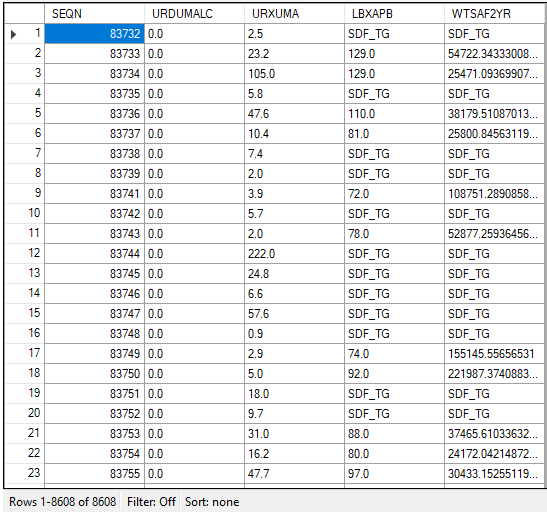


Ilustración Data generada en extensión .XPT (Fuente propia, 2019).

### **Archivo con extensión .XLSX**

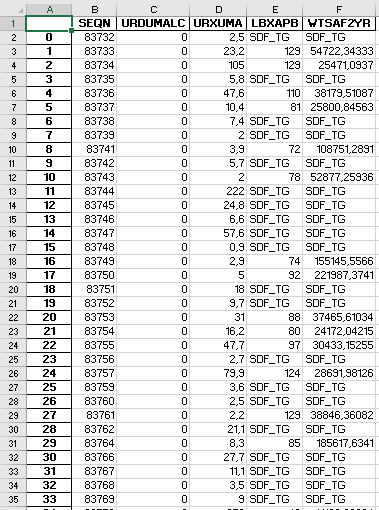


Ilustración Data generada en extensión .XLSX (Fuente propia, 2019).

## **Validación del software de investigación**

La validación del software se realiza con base a la fusión de 2 datas correspondientes los años 2016 y 2017 de forma manual comparándola con la realizada por el software SoftDataFusion, ver la Tabla 50.

|  |  |  |
| --- | --- | --- |
| **DETALLE / TIPO** | **MANUAL (120 minutos)** | **SOFTDATAFUSION (30 segundos)** |
| **NHANES** | 2015-2016 | |
| **DATA** | * Datos demográficos.   Datos de laboratorio | |
| **ARCHIVO** | * Variables demográficas y pesos de muestra. * Insulina. * Glucosa en ayunas de plasma. | |
| **VARIABLES** | Todas las variables de cada archivo, Total 56. | |
| **RELACIONES** | **2919** | |
| **TOTALES** | **9979** | |
| **SENSIBILIDAD** | **100%** | |
| **ESPECIFICIDAD** | **100%** | |

Tabla Análisis de Data en fusión Manuel y la generada (Fuente propia, 2019).

La fusión de estas datas ha sido de gran apoyo en el proceso de preparación de los datos para la investigación en curso titulada **Sistema computacional fundamentado en operadores inteligentes para la estimación funcional del factor resistencia a la insulina y HOMA: Caso práctico NHANES 1999-2000**, la cual se desarrolla desde un proyecto de aula en la Electiva de Ingeniería 3, del pensum académico de Ingeniería de Sistemas en la Universidad de Pamplona.

También SoftDataFusion está siendo utilizado para la preparación de datas en el Trabajo de Grado de un docente en la Universidad Simón Bolívar para optar por su siguiente nivel profesional, Maestría. De igual forma al este software estar disponible en repositorios en línea de código abierto, ha sido recomendado y compartido para todo investigador que quiera dar uso de la herramienta practica en su investigación, como lo es la Universidad de Zulia en Venezuela.

# **CAPÍTULO 5. CONCLUSIONES**

Un entorno de desarrollo adaptable a las necesidades del ejecutor y su amplia configuración con librerías hace que los tiempos y experiencias de desarrollo sean más agiles y agradables, dando una mejor productividad. Siendo un claro ejemplo NetBeans, el cual en la carrera universitaria ha sido una herramienta indispensable y configurable de acuerdo a las necesidades, permitiendo la implementación de Plataformas de Python y codificación en este mismo lenguaje.

Usar una metodología de desarrollo como la Programación Extrema XP, es de vital importancia en la ejecución de proyectos con poco tiempo y que requieran buena interpretación de Requerimientos o Historias de usuario para la mejor implementación GUI’s a la media como desea el usuario final.

El continuo trabajo en equipo, integración con nuevos proyectos e investigación de materiales, herramientas o metodologías en nombre de la ciencia hace que nos formemos a nivel personal y profesional con un mejor complemento de conocimientos; los cueles pueden desembocar en la generación de productos relevantes para la comunidad, el centro de estudios o grupo de investigación al que nos asociemos. Como en este caso la Universidad de Pamplona se vería considerablemente resaltada a nivel investigativo con publicaciones de las tesis de los graduados.

Los procesos para la publicación de un artículo científico como herramienta de software no son realmente complicados o complejos, son de constancia y atención a las descripciones o pautas de la revista a la que se desea aspirar a publicar. Algunos de los pasos importantes son el registro del software para el establecimiento de derechos de autor, el alojamiento del software en un repositorio como GitHub, la sincronización con un repositorio científico como Zenodo para la generación del Identificador de Objeto Digital (DOI) y la unión del documento formal (Artículo) con el DOI o DOI’s de los artefactos involucrados para que la revista pueda replicar el proceso en su respectiva revisión.

# **CAPÍTULO 6. RECOMENDACIONES**

Sugiero que se asigne un espacio investigativo para la recopilación, selección y análisis de Bases de Datos Abiertas con las cuales se pueda hacer investigación, implementación o mejora de herramientas prácticas para el proceso, como lo es SoftDataFusion para la fusión automática de bases de datos de NHANES. Esto con el fin de dar la oportunidad a los Jóvenes Investigadores que enfoque su motivación e ingenio en la implementación de herramientas y no en la recopilación de datos para trabajar.

# **CAPÍTULO 7. BIBLIOGRAFÍA**

NHANES. (2003). *DEPARTAMENTO DE SALUD Y SERVICIOS HUMANOS DE ESTADOS UNIDOS* Centros para el Control y Prevención de Enfermedades Centro Nacional de Estadísticas de Salud Data Disemination Branch National Center for Health Statistics Room 1064, Metro III Bldg. 6525 Belcrest Road Hyattsville, Maryland 20782-2003

Goodhew Peter. (2002). *PROGRAMAS INTERACTIVOS (SOFTWARE) PARA LA ENSEÑANZA DE MATERIALES* Journal of Materials Education, año/vol. 24, número 1-3 University of North Texas Denton, México pp. 39-43

Marrodán Maria. (2013). *ESTÁNDARES PARA LA EVALUACIÓN DEL CRECIMIENTO EVALUACIÓN DEL CRECIMIENTO Y LA CONDICIÓN NUTRICIONAL*. VIII Congreso Internacional SEDCA Antropometría en Cooperación al Desarrollo y Ayuda Humanitaria Grupo de Investigación EPINUT-UCM (eitces)(www.epinut.ucm.es)

AGUILAR VALDES, Juan et al. (2003). *Niveles de plomo en sangre y factores asociados, en niños del municipio de Centro Habana*. Rev Cubana Hig Epidemiol [online]. vol.41, n.1 [citado 2019-04-26]. Disponible en: <http://scielo.sld.cu/scielo.php?script=sci\_arttext&pid=S1561-30032003000100002&lng=es&nrm=iso>. ISSN 1561-3003.

Benjamin EJ, Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R, de Ferranti SD, Floyd J, Fornage M, Gillespie C, Isasi CR, Jimenez MC, Jordan LC, Judd SE, Lackland D, Lichtman JH, Lisabeth L, Liu S, Longenecker CT, Mackey RH, Matsushita K, Mozaffarian D, Mussolino ME, Nasir K, Neumar RW, Palaniappan L, Pandey DK, Thiagarajan RR, Reeves MJ, Ritchey M, Rodriguez CJ, Roth GA, Rosamond WD, Sasson C, Towfighi A, Tsao CW, Turner MB, Virani SS, Voeks JH, Willey JZ, Wilkins JT, Wu JHY, Alger HM, Wong SS, Muntner P. (2017). *En representación del Comité de Estadísticas y del Subcomité de Estadísticas de Ataque Cerebral de la American Heart Association*. Estadísticas de enfermedad cardíaca y de ataque cerebral --información actualizada: un informe de l American Heart Association [se publicó en línea el 25 de enero de 2017] Circulation. doi: 10.1161/CIR.0000000000000485

Sofía Victoria Cruz Varona, María Rosa Más Camacho, Eugenio Santander Pozo. (2012). *Identificación de necesidades de aprendizaje sobre el trabajo con bases de datos para tratamiento de información médica*. Revista de Educación Médica Superior 26(1):61-73

Mayenny Linares Río. (2014). *Diseño de un software para la enseñanza de la asignatura Programación y Gestores de Bases de Datos en la carrera de Tecnología de la Salud*. Revista de Ciencias Médicas de Pinar del Río vol.18 no.5 sep.-oct. 2014 versión On-line ISSN 1561-3194 Rev

María del Carmen Paderni López, Idorys Aguilar León. Mirna Cabrera Hernández. Ariel Delgado Ramos. (2014). *Revista Cubana de Informática Médica*. 6(2)227-235. Disponible en: http://scielo.sld.cu

NHANES. (2019). *Bases de datos imágenes medicas cáncer*. About the National Health and Nutrition Examination Survey. Disponible en: https://wiki.cancerimagingarchive.net/display/Public/TCGA-KICH

Sylvia Constaín Rengifo. (2019). *Guía para el uso y aprovechamiento de Datos Abiertos en Colombia*. Disponible en: http://gobiernodigital.gov.co/623/articles-9407\_guia\_datos.pdf

Jefatura del Estado. (2015). *Ley 18/2015, de 9 de julio, por la que se modifica la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público*. Disponible en: https://www.boe.es/buscar/pdf/2015/BOE-A-2015-7731-consolidado.pdf

Ministerio de la Presidencia. (2011). *Real Decreto 1495/2011, de 24 de octubre, por el que se desarrolla la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público, para el ámbito del sector público estatal*. Disponible en: https://www.boe.es/eli/es/rd/2011/10/24/1495

F1000Research Ltd. (2019). *How it Works: Rapid & Transparent Publishing*. Disponible en: <https://f1000research.com/about>

GitHub, Inc. (2019). *About work in GitHub*. Disponible en: <https://github.blog/category/company/policy/>

CERN. (2019*). Zenodo | About: Passionate about Open Science!*. Disponible en: <https://about.zenodo.org/>

DNDA. (2019). *Acerca de la Dirección Nacional de Derecho de Autor*. Colombia. Disponible en: <http://www.derechodeautor.gov.co/web/guest/definicion1;jsessionid=8B6A79366320734AFAFDFF32A6337C88.worker0>

NetBeans. (2019). *¿Qué es NetBeans?* Disponible en: <https://netbeans.org/index_es.html>

Guido van Rossum. (2009). *El tutorial de Python.* Disponible en: <http://docs.python.org.ar/tutorial/pdfs/TutorialPython2.pdf>

Python. (2019). *kinter — Python interface to Tcl/Tk.* Disponible en: <https://docs.python.org/2/library/tkinter.html>

PandasPyData. (2019). *Python Data Analysis Library.* Disponible en: <https://pandas.pydata.org/>

pypi. (2019). *xport 2.0.2*. Disponible en: <https://pypi.org/project/xport/>

pypi. (2019). *PyInstaller 3.2*. Disponible en: <https://pypi.org/project/PyInstaller/3.2/>

Orjuela Duarte, A., & Mauricio, R. C. (2008). *The Methodologies of Agile Development like an Opportunity for the Engineering of Educative Software*. Revista Avances En Sistemas E Informática, 5(2), 159–171.

Palsetia, N., Deepa, G., Khan, F. A., Thilagam, P. S., & Pais, A. R. (2016). *Securing Native XML Database-Driven Web Applications from XQuery Injection Vulnerabilities*. The Journal of Systems & Software. https://doi.org/10.1016/j.jss.2016.08.094

Diego Calvo. (2018). *Metodología XP Programación Extrema (Metodología ágil).* Disponible en: http://www.diegocalvo.es/metodologia-xp-programacion-extrema-metodologia-agil/

Letelier, P., & Penades, M. C. (2006). *Metodologías Ágiles para el desarrollo del software:Extreme Programming (XP).* Disponible en: http://www.cyta.com.ar/ta0502/v5n2a1.htm

SINTYA MILENA MELÉNDEZ VALLADAREZ, MARIA ELIZABETH GAITAN, NELDIN NOEL PÉREZ REYES. (2016). *METODOLOGIA ÁGIL DE DESARROLLO DE SOFTWARE PROGRAMACION EXTREMA.* Disponible en: http://repositorio.unan.edu.ni/1365/1/62161.pdf

Ferreira Escutia, R. (2013). *XP Extreme Programming*. Disponible en: http://slideplayer.es/slide/84721/

Chiluisa Pallo, A. P., & Loarte Cajamarca, B. G. (2014). *Desarrollo e Implantación del Sistema de Control de Inventarios y Gestión de Laboratorios para la Facultad de Ciencias de la Escuela Politécnica Nacional.* Quito.

Erlijman Piwen, A., & Goyen Fros, A. (2001). *Problemas y Soluciones en la Implementación de Extreme Programming*. Montevideo.

Joskowicz, J. (2008). *Reglas y Prácticas en Extreme Programming*. Disponible en: http://iie.fing.edu.uy/~josej/docs/XP%20-%20Jose%20Joskowicz.pdf

Echeverry Tobón, L. M., & Delgado Carmona, L. E. (2007). *Caso Práctico de la Metodología Ágil XP al Desarrollo del Software*. Pereira.

Balestrini M. (2009). *Como se elabora el proyecto de investigación*. Caracas. Editorial BL. Consultores asociados.

Hernández R, Fernández C, Baptista M (2010 ). Metodología de la investigación, McGRAW-HILL / INTERAMERICANA EDITORES, S.A. Quinta edición ISBN: 978-607-15-0291-9 C.P. 01376, México D.F.

Universidad Pedagógica Experimental Libertador. (2003). *Manual de Trabajos de grado de maestría y tesis doctorales.* Caracas. Vicerrectorado de investigación y postgrado. UPEL.

Arias F (2012). *El proyecto de investigación. Introducción a la metodología científica*. Caracas: Episteme.

Jorge Miguel Anaya Leon, Gerardo Chacón, Johel E. Rodriguez, Valmore Bermúdez, Alexander Bayona, Anderson Smith Flórez Fuentes, & Jose Ricardo Bernal Leon. (2019, October 17). *SoftDataFusion V.01.01 - NHANES (Version V.01.01)*. Zenodo. http://doi.org/10.5281/zenodo.3496224

Jaime Andrés Britto Montoya. (2016). Comparación de metodologías ágiles y procesos de desarrollo de software mediante un instrumento basado en CMMI. Universidad Autónoma de Manizales, Manizales, Colombia. Disponible en: file:///C:/Users/JorgeMAL/Downloads/9249-30101-2-PB.pdf

Andrés Navarro Cadavid, Juan Daniel Fernández Martínez, Jonathan Morales Vélez. (2013). *Revisión de metodologías ágiles para el desarrollo de software*. s, Grupo de investigación i2T, Universidad Icesi. Disponible en: https://www.redalyc.org/pdf/4962/496250736004.pdf

# **CAPÍTULO 8. ANEXOS**



## **Manual de Usuario**

### **SoftDataFusion**

Para iniciar el proceso de fusión de datas (cual sea la cantidad de datas y variables que se desee), es necesario iniciar SoftDataFusion desde su archivo ejecutable​ ***SoftDataFusion.exe*** dentro de la carpeta **SoftDataFusion​**.

*Es importante tener en cuenta que* ***SoftDataFusion*** *cuenta con una guía en tiempo real, la cual puede ser ejecutada en cualquier momento durante un trabajo de fusión y no perder lo que se tiene en la ventana. ​ ​Dar ​****ctrl + t***

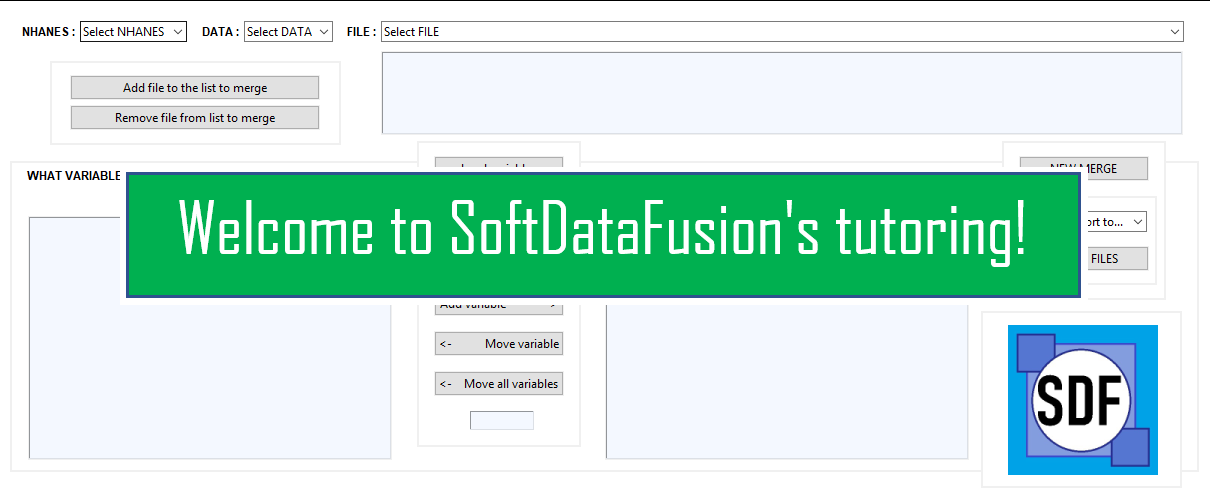


Ilustración MU: Bienvenida e inicio del tutorial (Fuente propia, 2019).

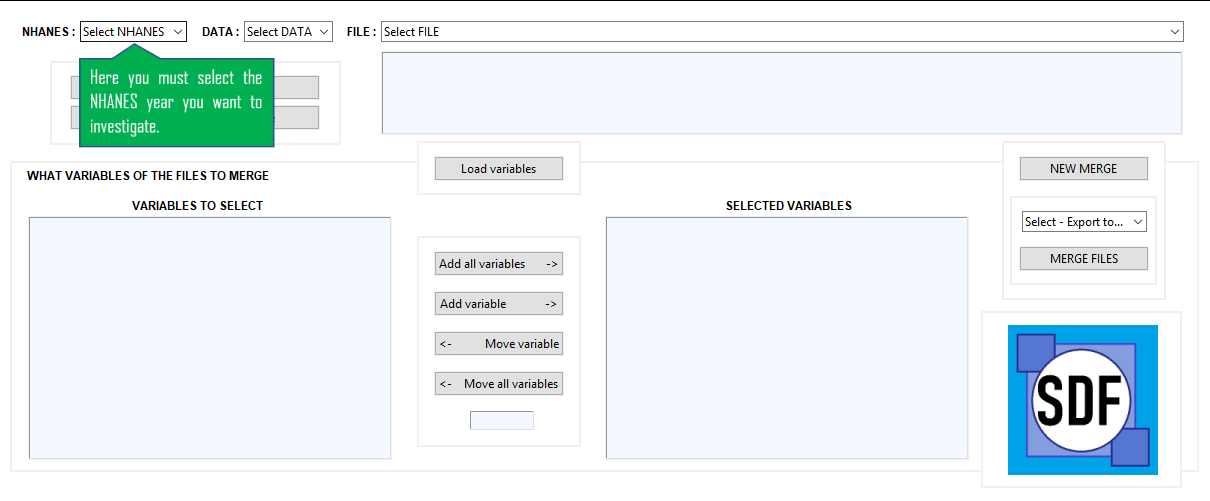


Ilustración MU: Selección del año del NHANES (Fuente propia, 2019).

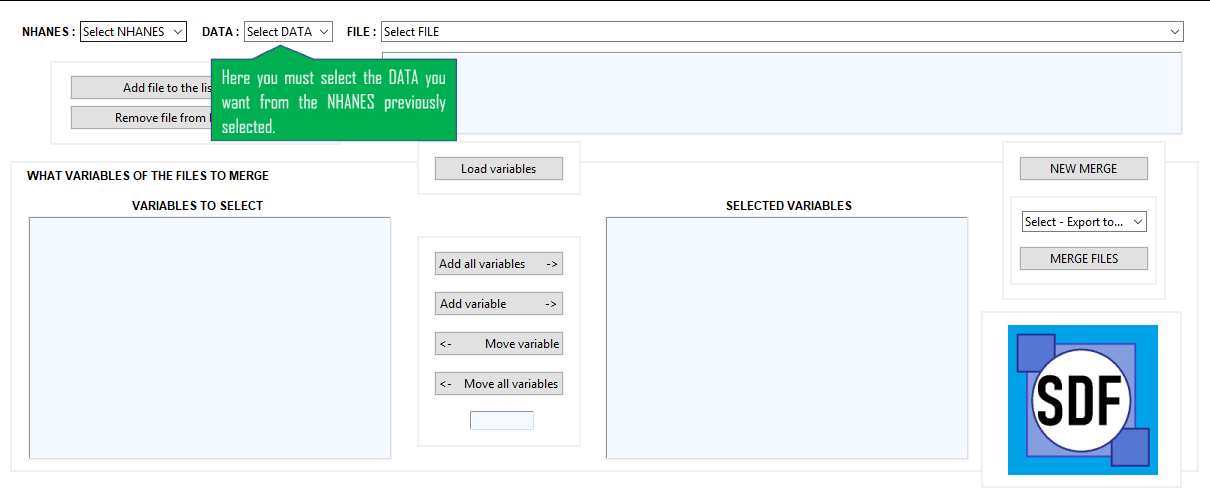


Ilustración MU: Selección de la DATA (Fuente propia, 2019).

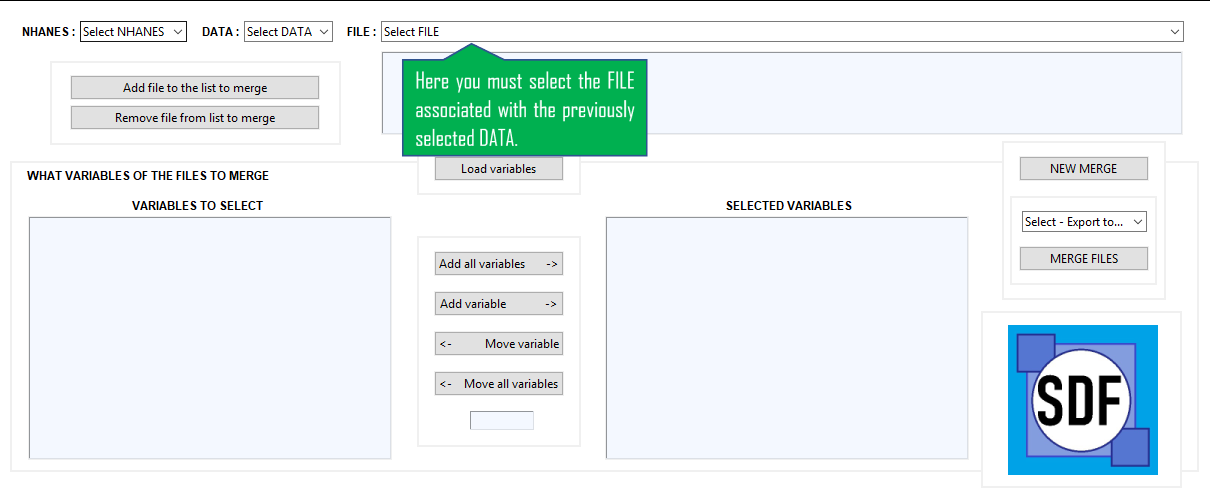


Ilustración MU: Selección de la ARCHIVO/FILE (Fuente propia, 2019).

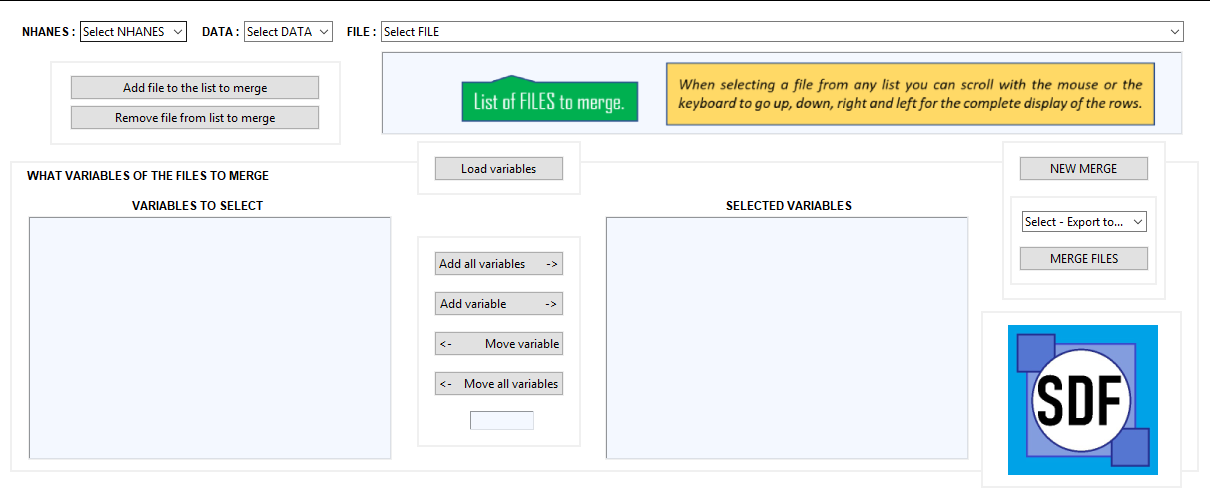


Ilustración MU: Lista de ARCHIVOS/FILES a fusionar (Fuente propia, 2019).

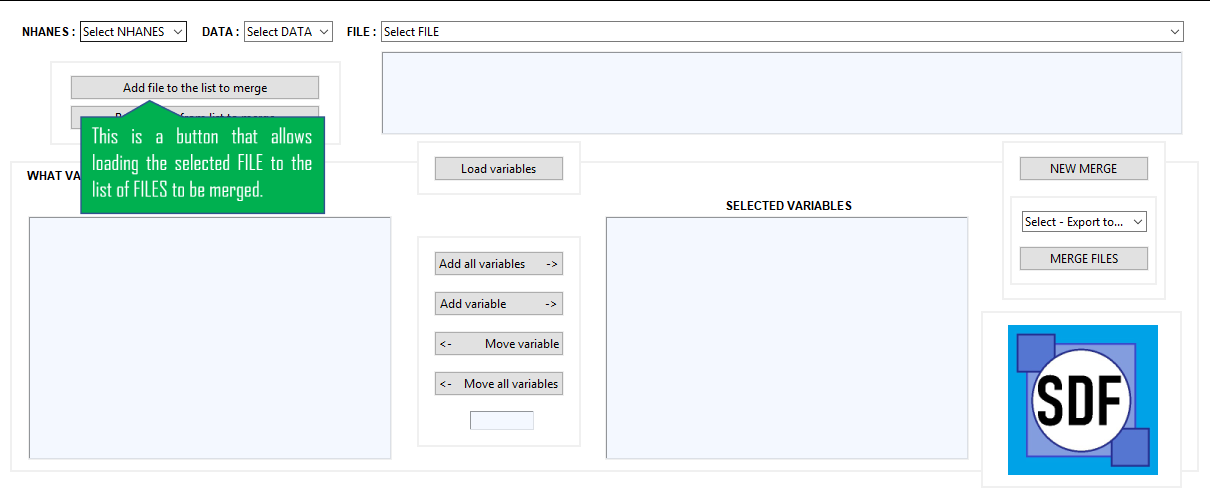


Ilustración MU: Este botón permite Ilustración 24 a Ilustración 25 (Fuente propia, 2019).

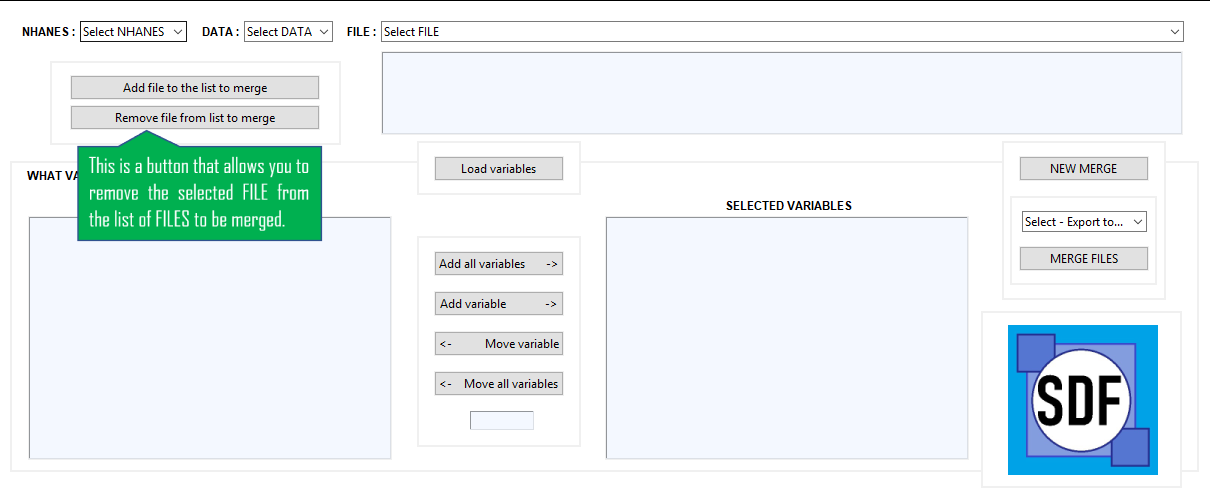


Ilustración MU: Este botón permite desagregar los ARCHIVOS/FILES de Ilustración 25 (Fuente propia, 2019).

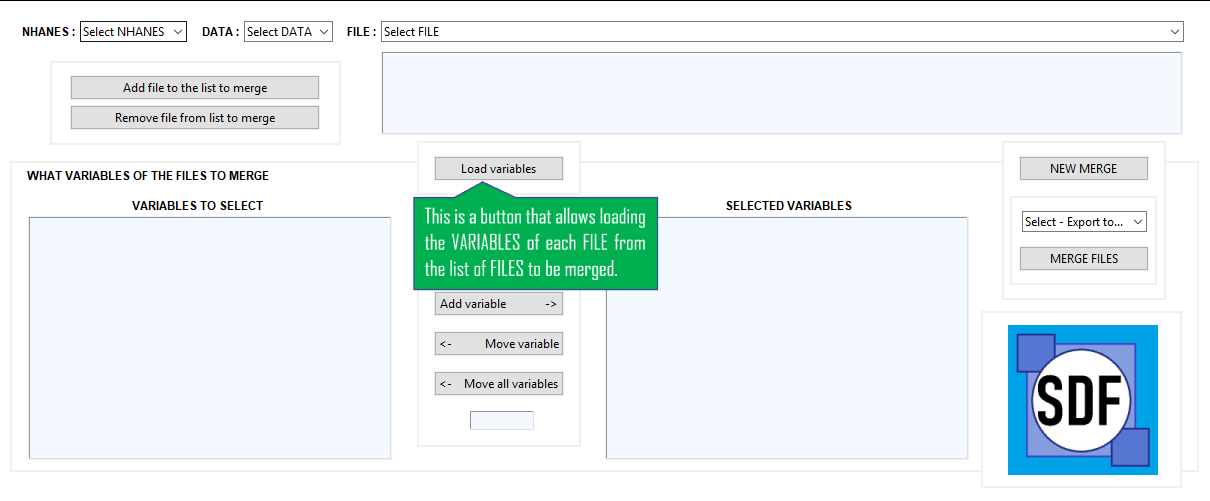


Ilustración MU: Cargar las VARIABLES de los ARCHIVOS/FILES (Fuente propia, 2019).

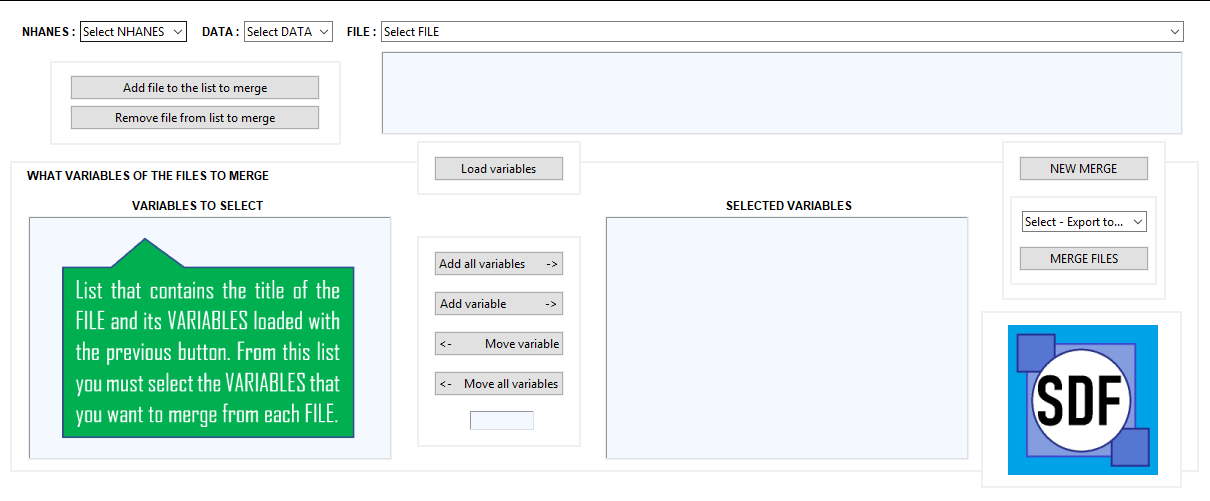


Ilustración MU: Lista de VARIABLES A SELECCIONAR (Fuente propia, 2019).

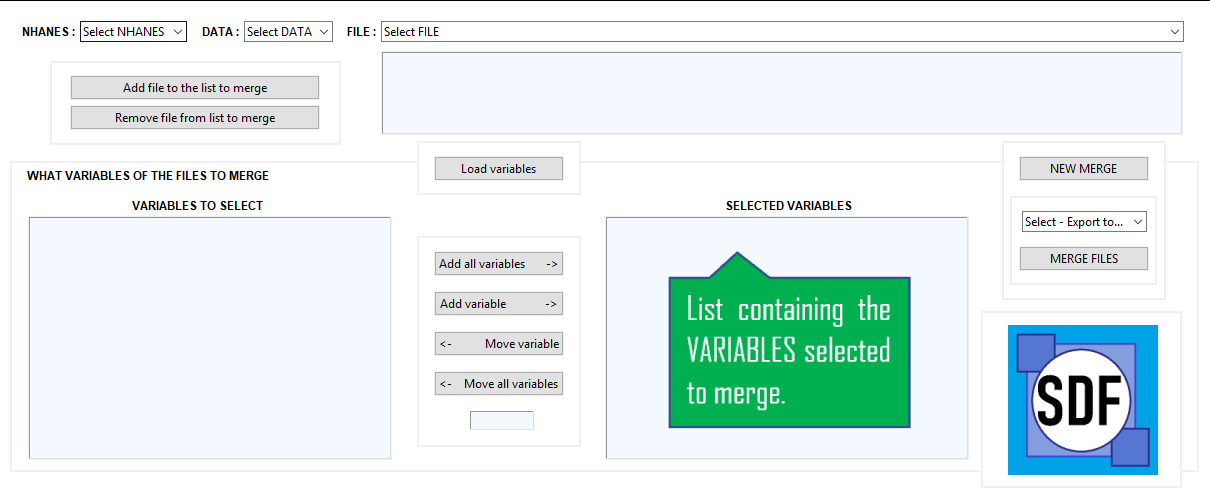


Ilustración MU: Lista de VARIABLES SELECCIONADAS (Fuente propia, 2019).

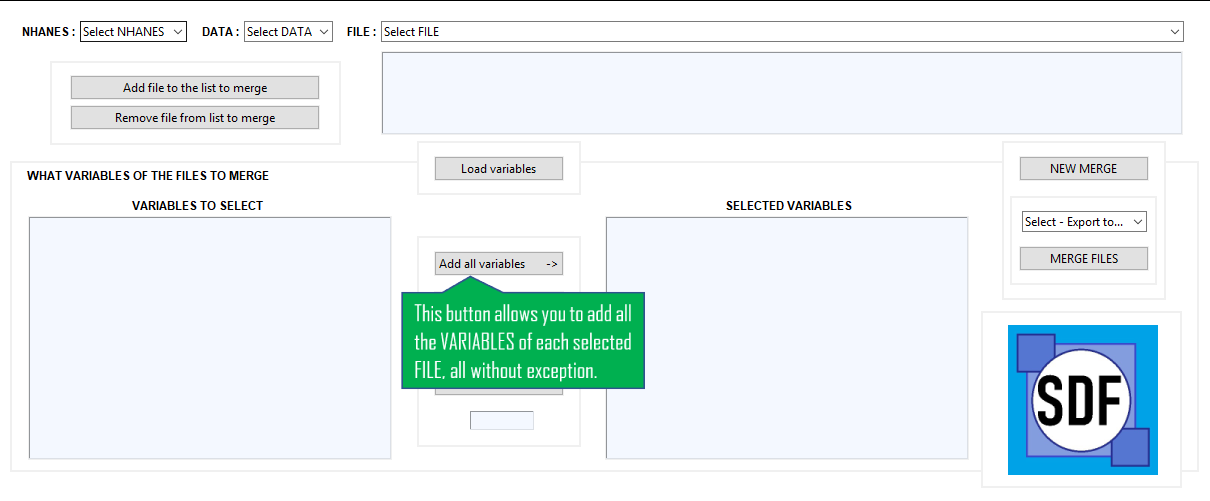


Ilustración MU: Mover todas las VARIABLES a VARIABLES SELECCIONADAS (Fuente propia, 2019).

\*\*\*

**Importante seleccionar el SEQN por cada título de NHANES-DATA-FILE, este es el que le permite a SoftDataFusion relacionar los casos.**

\*\*\*

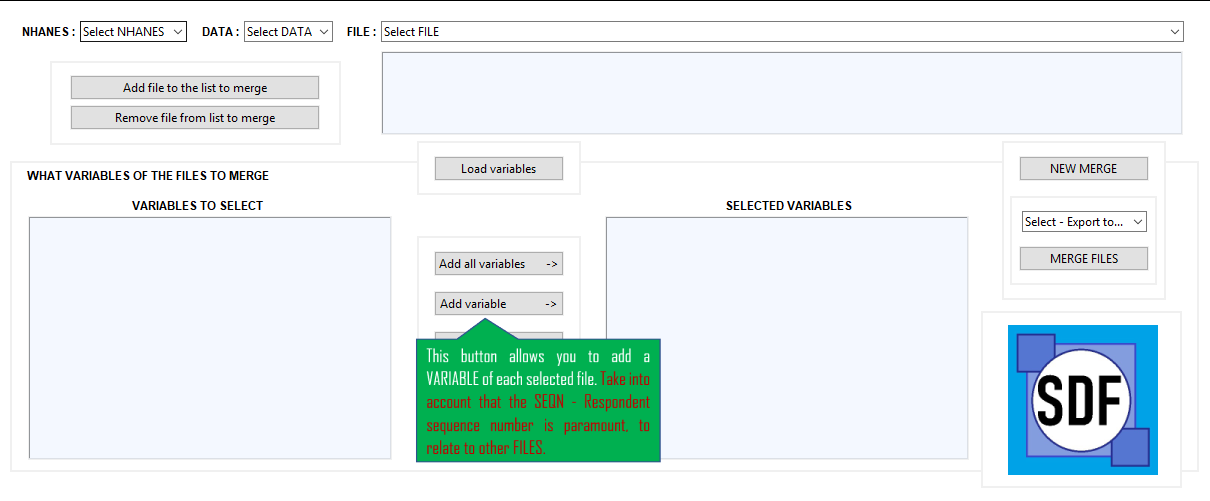


Ilustración MU: Mover de a una VARIABLE a VARIABLES SELECCIONADAS (Fuente propia, 2019).

\*\*\*

**Importante seleccionar el SEQN por cada título de NHANES-DATA-FILE, este es el que le permite a SoftDataFusion relacionar los casos.**

\*\*\*

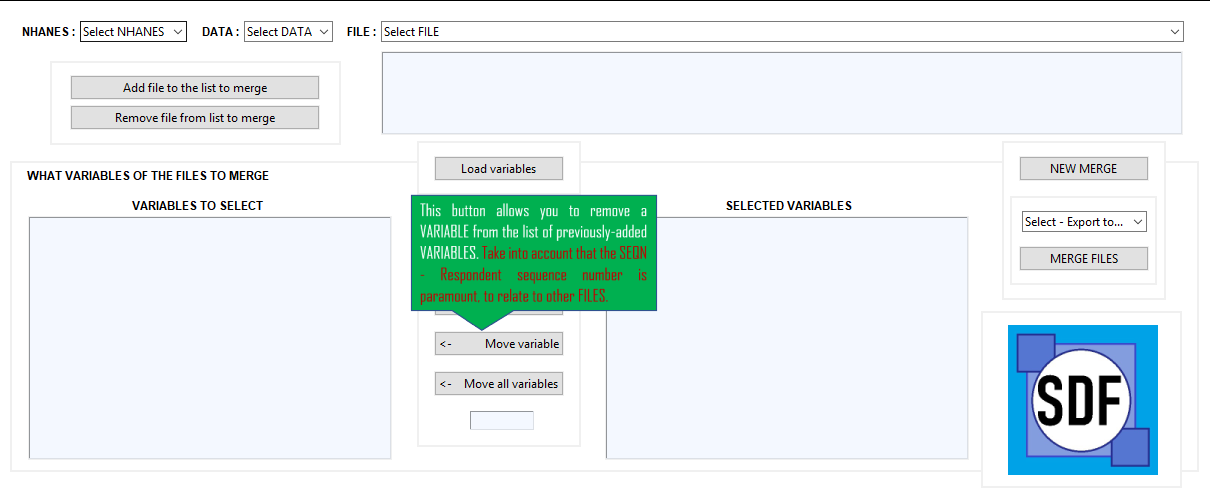


Ilustración MU: Remover de a una VARIABLE de VARIABLES SELECCIONADAS (Fuente propia, 2019).

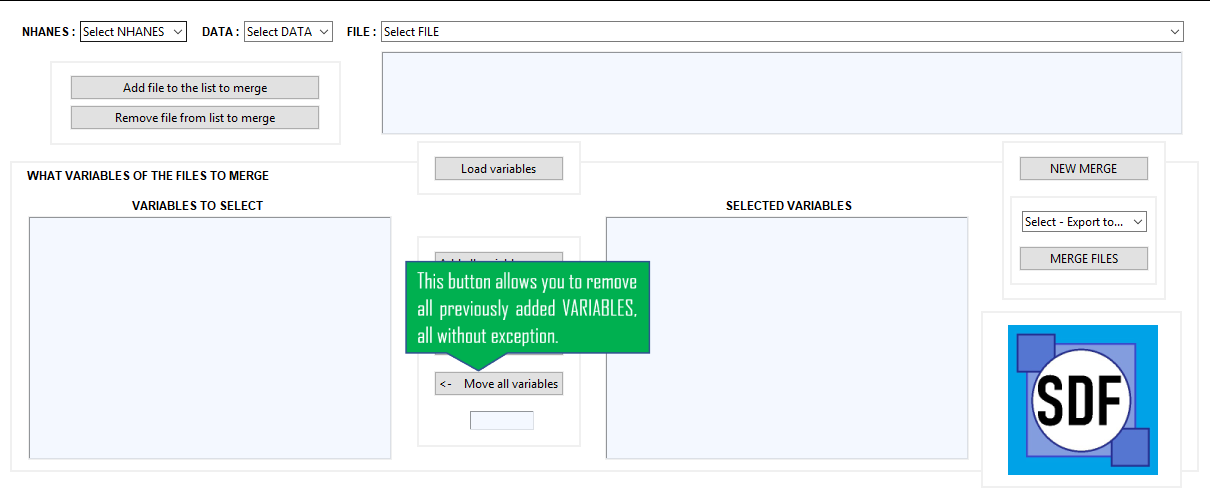


Ilustración MU: Remover todas las VARIABLES de VARIABLES SELECCIONADAS (Fuente propia, 2019).

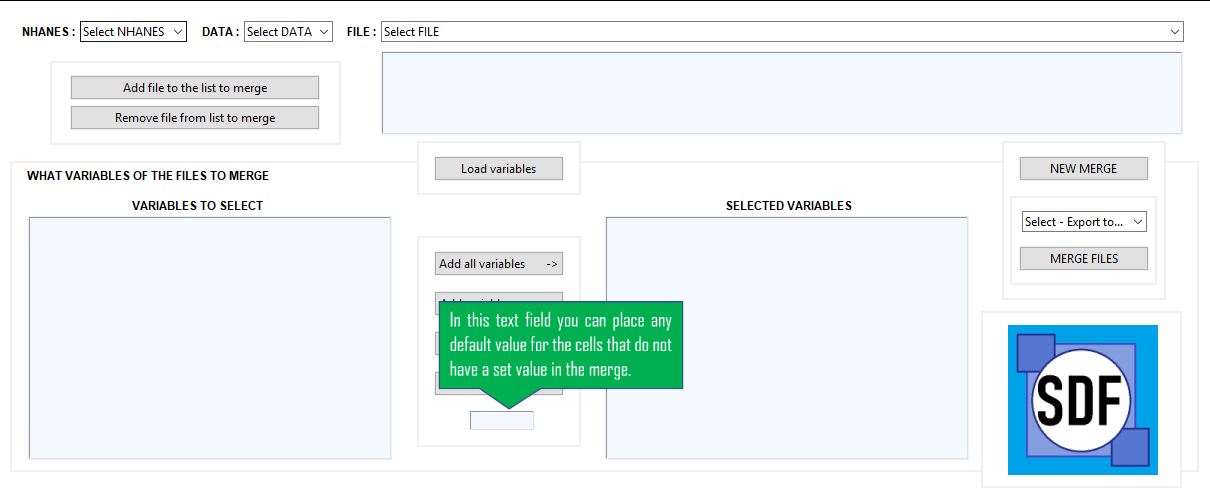


Ilustración MU: VALOR para los CAMPOS de REGISTROS SIN RELACIÓN (Fuente propia, 2019).

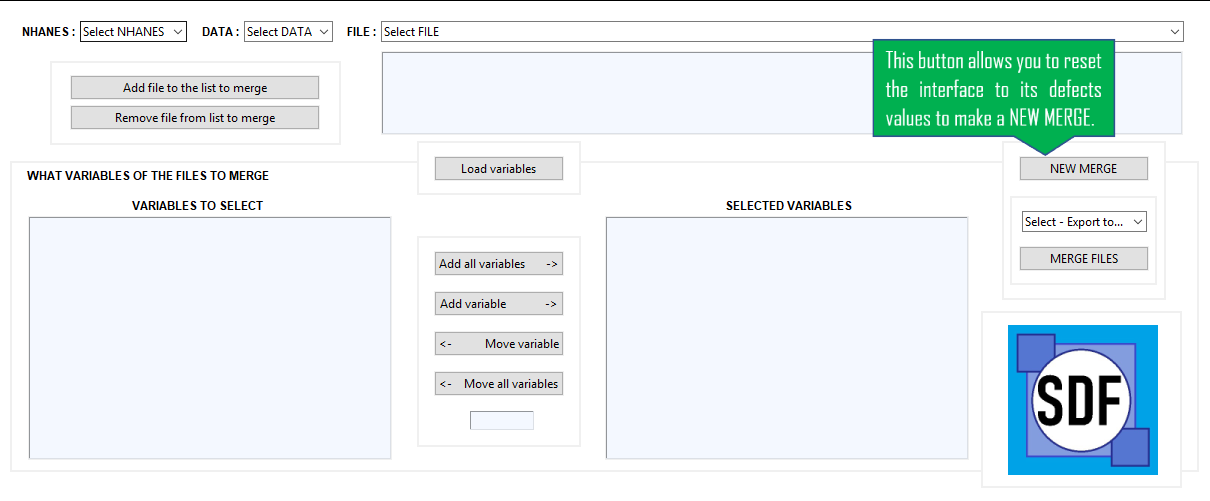


Ilustración MU: Formatear para la NUEVA FUSION/NEW MERGE (Fuente propia, 2019).

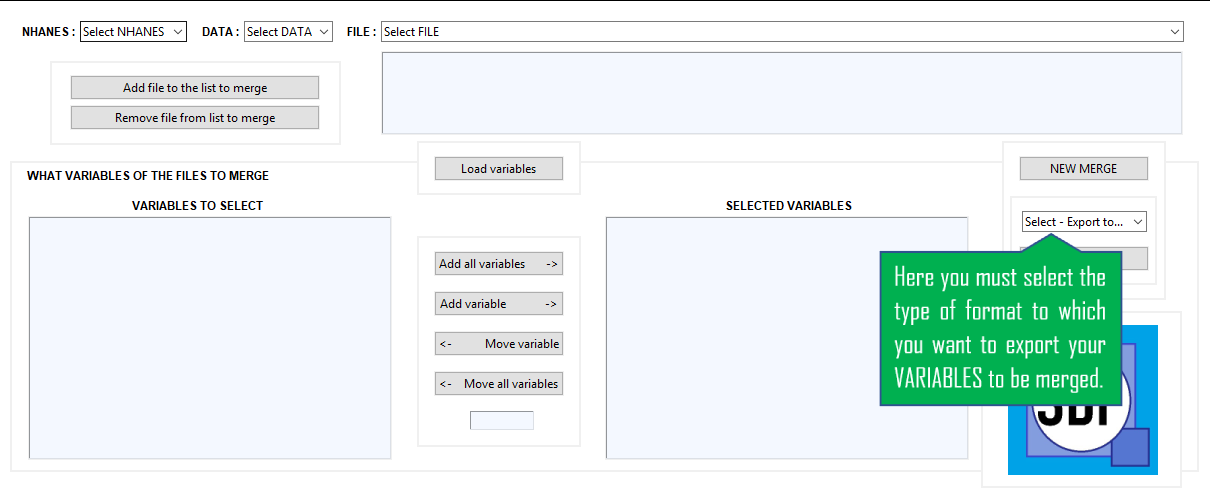


Ilustración MU: Seleccionar TIPO DE ARCHIVO para EXPORTAR (Fuente propia, 2019).

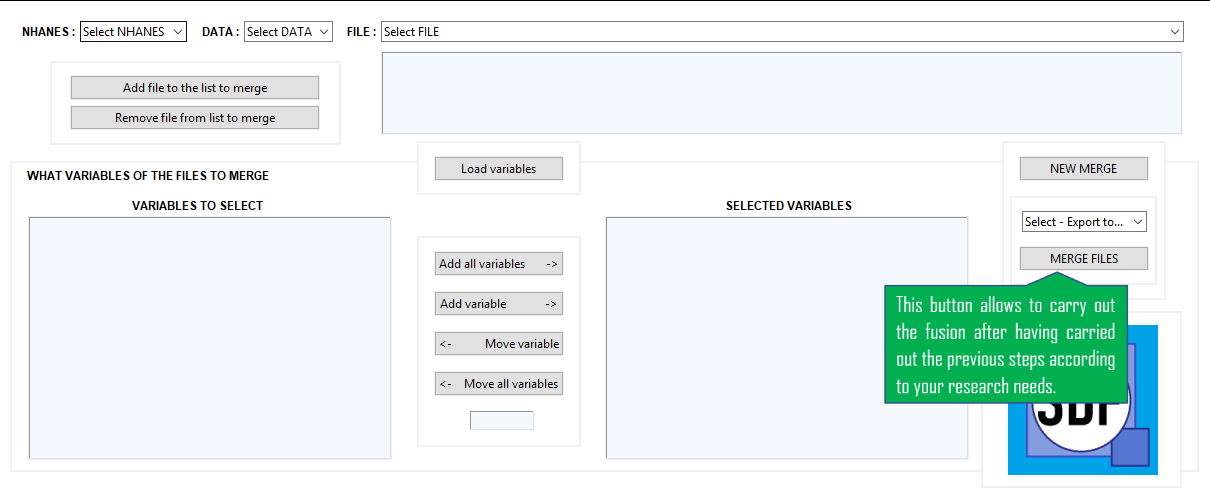


Ilustración MU: Realiza la FUSIÓN/MERGE (Fuente propia, 2019).

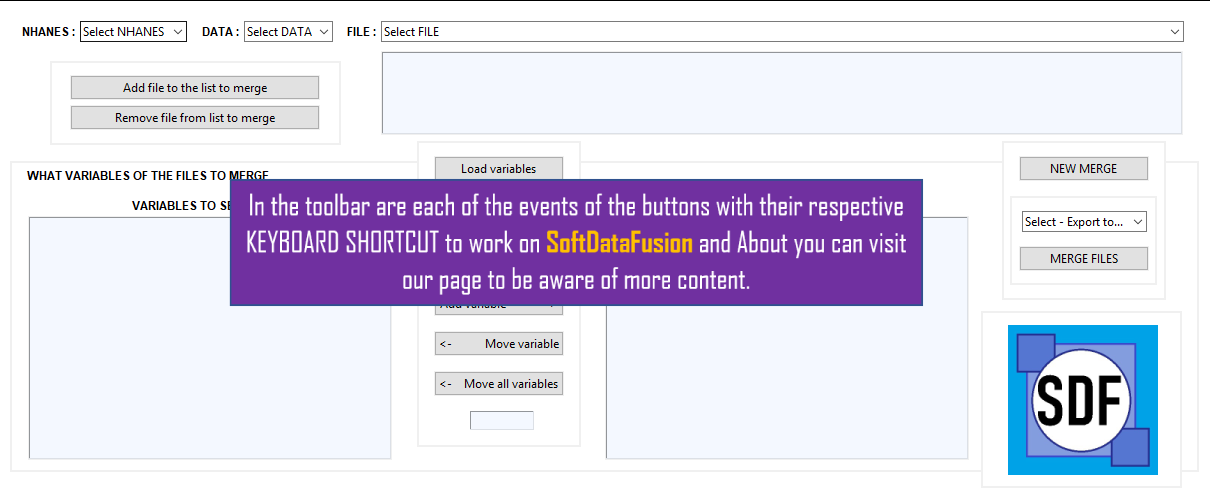


Ilustración MU: Final del Tutorial (Fuente propia, 2019).

\*\*\*\*

***En todo momento SoftDataFusion interactivamente le estará informando al investigador que es lo que está realizando o lo que hizo.***

\*\*\*\*

En la barra de herramientas se encuentran con atajos de teclado todas las acciones de los botones que se muestran en la interfaz, los mismos que fueron anteriormente descritos. Esto con el objetivo de hacer más ágil el proceso de fusión.

Los FILES descargados directamente de NHANES se alojan en la carpeta ***00-DownloadedData,*** que se encuentra al mismo nivel de archivos que el ejecutable de **SoftDataFusion.exe**

Los FILES fusionados se alojan en la carpeta ***00-MergedData,*** que se encuentra al mismo nivel de archivos que el ejecutable de **SoftDataFusion.exe**



Ilustración Logo de SoftDataFusion (Fuente propia, 2019).

## **Certificado de registro del Software ante la DNDA en Colombia**

La Dirección Nacional de Derecho de Autor es un organismo del Estado Colombiano, que posee la estructura jurídica de una Unidad Administrativa Especial adscrita al Ministerio del Interior y es el órgano institucional que se encarga del diseño, dirección, administración y ejecución de las políticas gubernamentales en materia de derecho de autor y derechos conexos. En tal calidad posee el llamado institucional de fortalecer la debida y adecuada protección de los diversos titulares del derecho de autor y los derechos conexos, contribuyendo a la formación, desarrollo y sustentación de una cultura nacional de respeto por los derechos de los diversos autores y titulares de las obras literarias y artísticas.

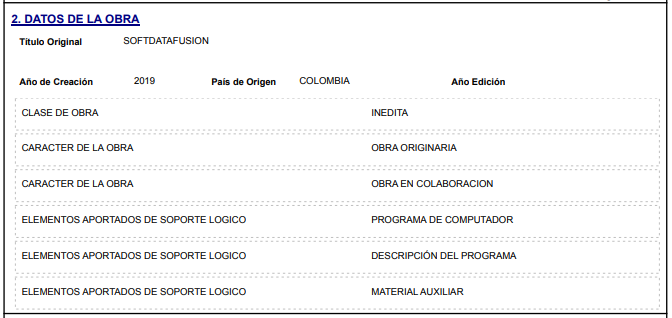
Dentro de este entorno, la acción institucional de la DNDA involucra el estudio y proceso de expedición, de la normatividad autoral de nuestro país, así como la adhesión a los principales convenios internacionales sobre protección del derecho de autor y los derechos conexos.

De igual forma, la Dirección Nacional de Derecho de Autor participa activamente en todos los procesos de negociación comercial que adelanta nuestro país a nivel bilateral y multilateral, y en los cuales se discuten los temas del derecho de autor y los derechos conexos (DNDA, 2019).









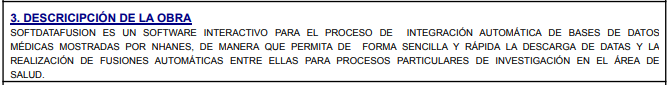


Ilustración Certificado de registro de soporte lógico - Software ante la DNDA en Colombia (Fuente propia, 2019)

## **SoftDataFusion en GitHub**

GitHub es una plataforma de alojamiento de código para el control de versiones y la colaboración. Te permite a ti y a otros trabajar juntos en proyectos desde cualquier lugar (GitHub, Inc, 2019).



Ilustración SoftDataFusion en GitHub (Fuente propia, 2019).

## **Conectando a los investigadores con la investigación - ORCID**

ORCID es una organización sin fines de lucro ayudando a crear un mundo en el que todos los que participan en la investigación, el estudio y la innovación son identificados y conectados de forma única a sus contribuciones y afiliaciones a lo largo de disciplinas, fronteras y el tiempo.

La comunidad ORCID incluye investigadores individuales, universidades, laboratorios nacionales, organizaciones de investigación comercial, financiadores para la investigación, editores, agencias científicas nacionales, repositorios de datos y asociaciones de profesionales internacionales, los cuales se han visto afectados gravemente por la falta de un registro central para los investigadores. ORCID coordina con la comunidad a través de Grupos de Trabajo y Reuniones Informativas semestrales.

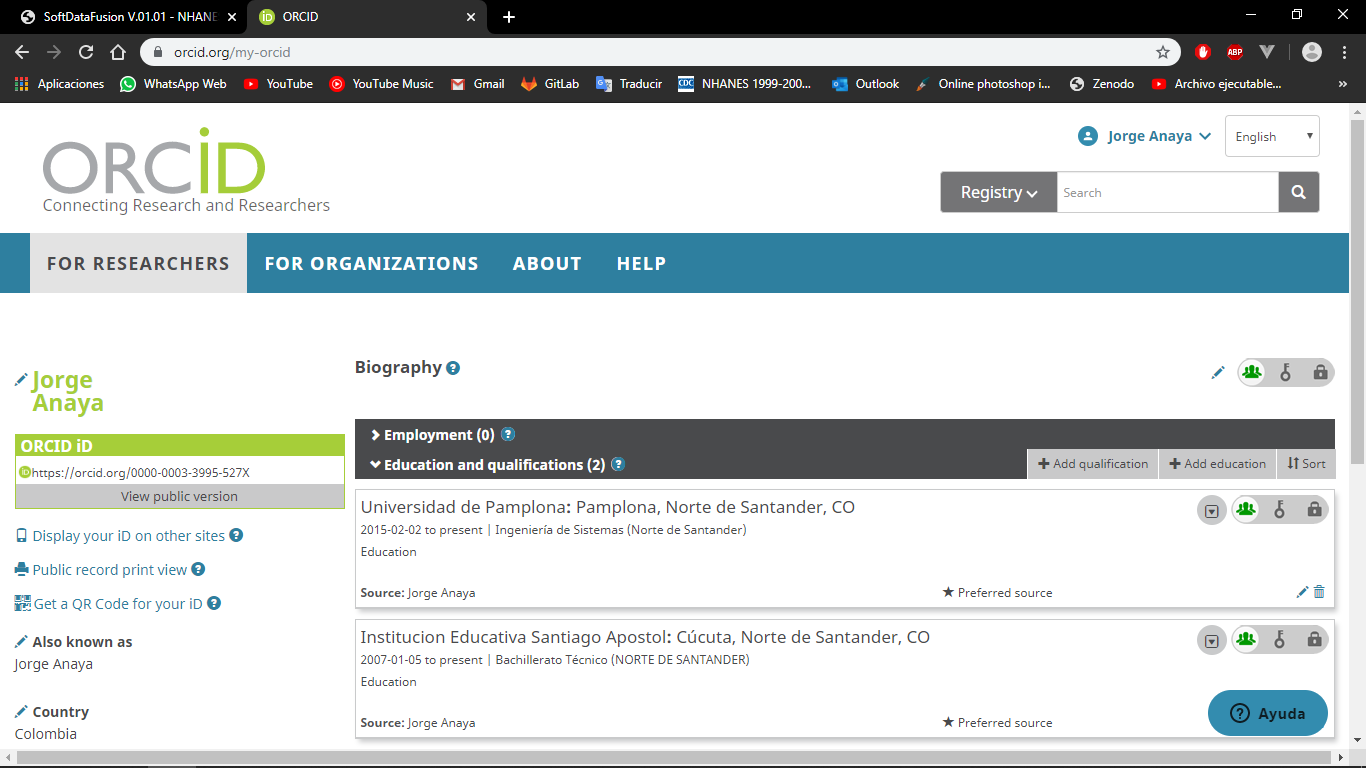


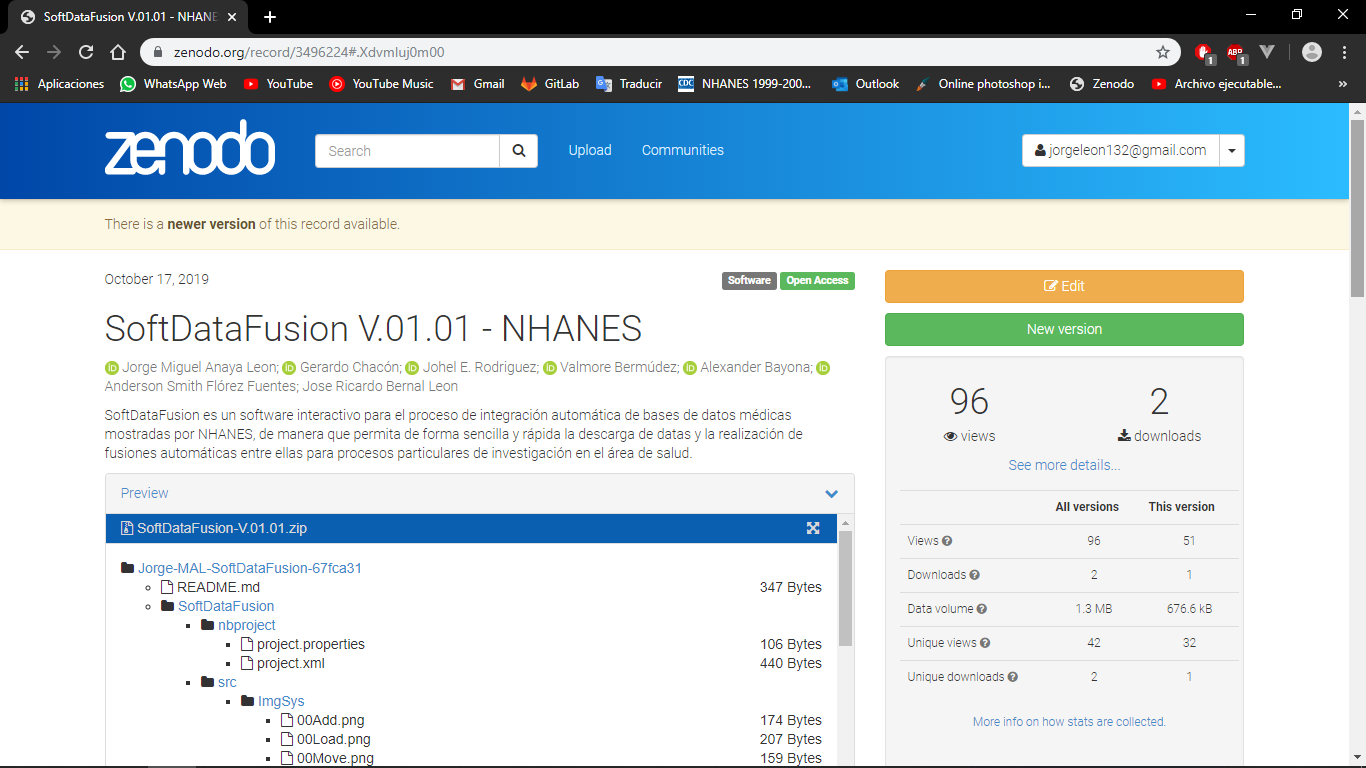
Ilustración 7.4 Conectando a los investigadores con la investigación – ORCID.

.

## **Sincronización de GitHub con Zenodo (Generación del DOI)**

El proyecto OpenAIRE, a la vanguardia del acceso abierto y los movimientos de datos abiertos en Europa, fue comisionado por la CE para apoyar su política naciente de Datos Abiertos al proporcionar un depósito general para la investigación financiada por la CE. CERN, un socio de OpenAIRE y pionero en código abierto, acceso abierto y datos abiertos, proporcionó esta capacidad y Zenodo se lanzó en mayo de 2013.

En apoyo de su programa de investigación, CERN ha desarrollado herramientas para la gestión de Big Data y capacidades ampliadas de la Biblioteca Digital para Open Data. A través de Zenodo, estas herramientas de Big Science podrían compartirse efectivamente con la larga cola de la investigación (CERN, 2019).



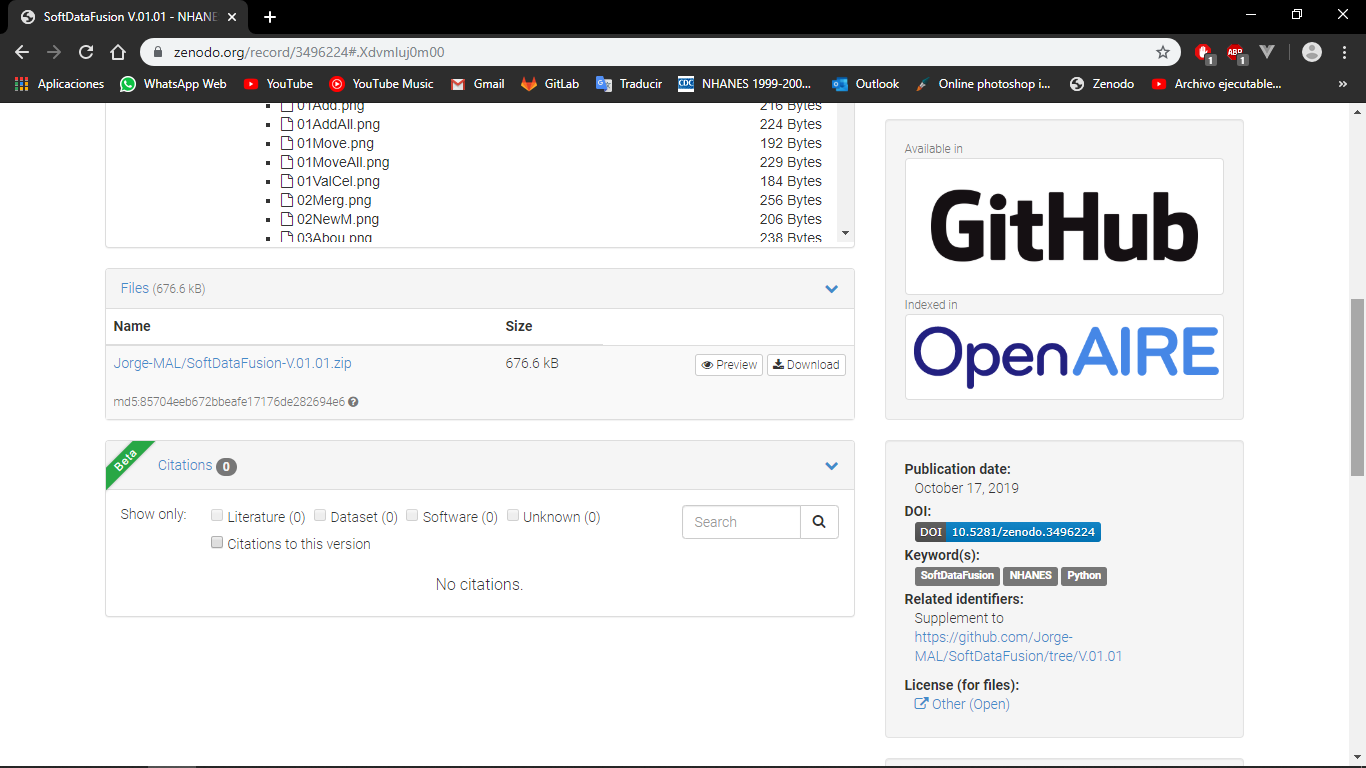


Ilustración 7.4 Sincronización de GitHub con Zenodo (Generación del DOI).

## **Publicación del Artículo de Herramienta de Software en F1000**

F1000Research es una plataforma de Open Science: todos los artículos se publican en acceso abierto; los procesos de publicación y revisión por pares son totalmente transparentes; y se les pide a los autores que incluyan descripciones detalladas de los métodos y que proporcionen acceso completo y fácil a los datos fuente subyacentes a los resultados para mejorar la reproducibilidad.

#### **Envío de artículos**

El envío se realiza a través de un sistema de envío de una página. El equipo editorial interno lleva a cabo un conjunto completo de comprobaciones previas a la publicación para garantizar que se cumplan todas las políticas y pautas éticas.

#### **Publicación y deposición de datos**

Una vez que los autores han finalizado el manuscrito, el artículo se publica dentro de una semana, lo que permite su visualización y cita inmediata.

#### **Revisión por pares abierta y comentarios de usuarios**

Los revisores expertos son seleccionados e invitados, y sus informes y nombres se publican junto con el artículo, junto con las respuestas de los autores y los comentarios de los usuarios registrados.

#### **Revisión del artículo**

Se alienta a los autores a publicar versiones revisadas de su artículo. Todas las versiones de un artículo están vinculadas y son citables de forma independiente. Los artículos que pasan la revisión por pares se indexan en bases de datos externas como PubMed, Scopus y Google Scholar (F1000Research Ltd, 2019).

Durante el proceso de publicación del Articulo de Herramienta de Software se permite llevar un completo monitoreo del estado en que se encuentra nuestro artículo partiendo desde el envío del mismo.

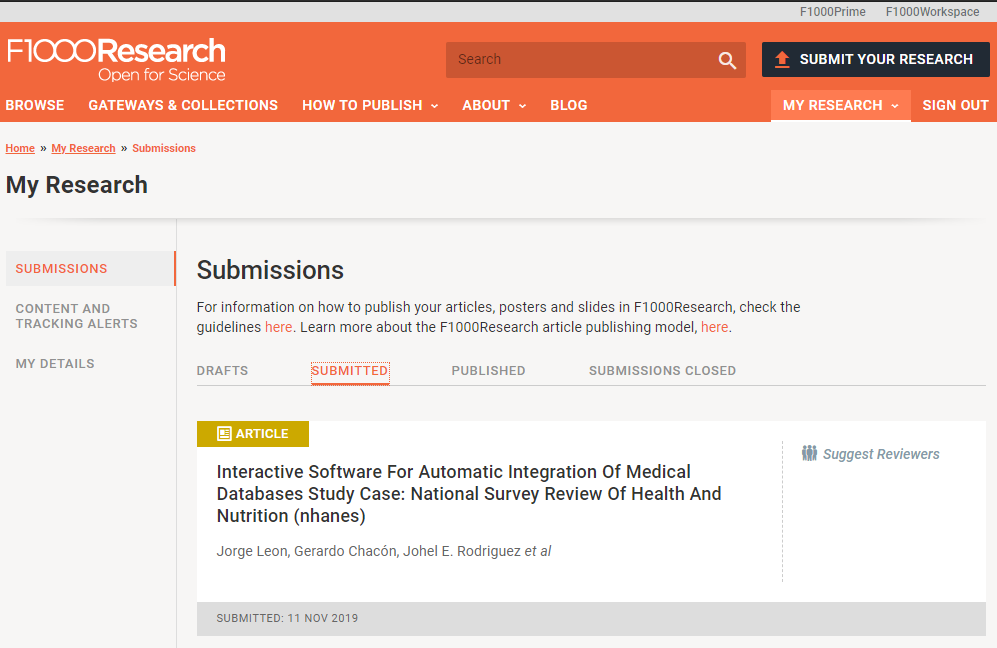
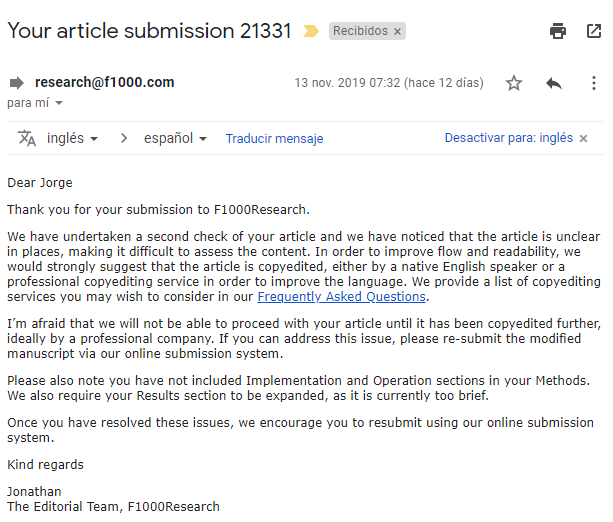


Ilustración Envío de el Artículo Herramienta de Software a F1000.

En donde por medio de correos electrónicos la revista notifica a el autor de correspondencia observaciones o correcciones a tener en cuenta para la publicación.



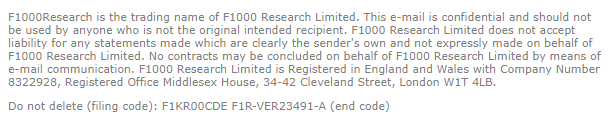


Ilustración Correcciones u observaciones de F1000 a tener en cuenta para la publicación.

## **Estadísticos de monitoreo de SoftDataFusion en línea**

A continuación, se presentan estadísticos sobre la cantidad de visualizaciones, descargas de archivos de NHANES con SoftDataFusion y fusiones que se han realizado por medio de este. Con las tablas de seguimiento en línea.

|  |  |
| --- | --- |
| **VISITAS** | |
| Identificador | Fecha |
| 1 | 19/11/2019 23:11 |
| 2 | 19/11/2019 23:13 |
| 3 | 19/11/2019 23:21 |
| 4 | 20/11/2019 20:00 |
| 5 | 20/11/2019 20:01 |
| 6 | 20/11/2019 20:02 |
| 7 | 20/11/2019 20:09 |
| 8 | 20/11/2019 20:10 |
| 9 | 20/11/2019 20:10 |
| 10 | 20/11/2019 20:11 |
| 11 | 20/11/2019 20:12 |
| 12 | 20/11/2019 20:13 |
| 13 | 20/11/2019 20:14 |
| 14 | 20/11/2019 20:14 |
| 15 | 20/11/2019 20:14 |
| 16 | 20/11/2019 20:15 |
| 17 | 20/11/2019 20:16 |
| 18 | 20/11/2019 20:17 |
| 19 | 20/11/2019 20:19 |
| 20 | 20/11/2019 20:19 |
| 21 | 20/11/2019 20:19 |
| 22 | 20/11/2019 20:19 |
| 23 | 20/11/2019 20:19 |
| 24 | 20/11/2019 20:19 |
| 25 | 20/11/2019 20:20 |
| 26 | 20/11/2019 20:20 |
| 27 | 20/11/2019 20:29 |
| 28 | 20/11/2019 20:29 |
| 29 | 20/11/2019 23:05 |
| 30 | 20/11/2019 23:44 |
| 31 | 20/11/2019 23:49 |
| 32 | 21/11/2019 8:08 |
| 33 | 21/11/2019 8:58 |
| 34 | 22/11/2019 5:06 |
| 35 | 22/11/2019 5:22 |
| 36 | 22/11/2019 5:26 |
| 37 | 22/11/2019 5:26 |
| 38 | 22/11/2019 18:22 |
| 39 | 23/11/2019 7:05 |
| 40 | 23/11/2019 10:44 |
| 41 | 26/11/2019 19:22 |
| 42 | 30/11/2019 17:46 |
| 43 | 1/12/2019 17:33 |
| 44 | 1/12/2019 20:48 |
| 45 | 1/12/2019 21:01 |
| 46 | 1/12/2019 18:42 |

Tabla Cantidad de visitas a SoftDataFusión.

|  |  |
| --- | --- |
| **DESCARGAS** | |
| Identificador | Fecha |
| 1 | 19/11/2019 23:11 |
| 2 | 19/11/2019 23:21 |
| 3 | 20/11/2019 20:16 |
| 4 | 21/11/2019 8:08 |
| 5 | 22/11/2019 5:26 |
| 6 | 22/11/2019 18:25 |
| 7 | 23/11/2019 7:05 |
| 8 | 23/11/2019 10:44 |
| 9 | 29/11/2019 17:33 |

Tabla Cantidad de descargas de SoftDataFusion.

|  |  |
| --- | --- |
| **FUSIONES** | |
| Identificador | Fecha |
| 1 | 19/011/2019 23:11 |
| 2 | 23/011/2019 7:26 |
| 3 | 29/11/2019 20:26 |

Tabla Cantidad de fusiones realizadas con SoftDataFusion.



Ilustración Estadística de eventos por fechas de SoftDataFusion.

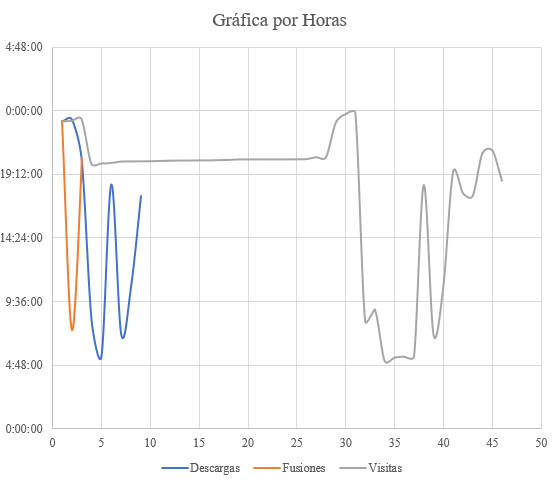


Ilustración Estadística de eventos por horas de SoftDataFusion.