



Tratamiento de datos para Mapa de Servicios de Atención Integral a la Primera Infancia en República Dominicana



Tratamiento de datos para Mapa de Servicios de Atención Integral a la Primera Infancia en República Dominicana

Dirección de Evaluación e Investigación

Julio César Mejía

Autor

Noel Rodríguez

Divulgación Científica

Julián Álvarez Acosta

Francisco Martínez

Corrección de estilo

Luis Emilio Segura

Diseño y Diagramación

Yeimy Rosa Olivier Salcedo

Natasha Mercedes Arias

Centro de Gestión de la Información y Documentación

Dilcia Armesto Núñez

Derechos Reservados

Ministerio de Educación de la República Dominicana

Se permite reproducir parcialmente este documento siempre que se cite la fuente

ISBN DIGITAL: 978-9945-499-58-2

Julio 2020

Santo Domingo, D.N.
República Dominicana



AUTORIDADES

Danilo Medina Sánchez
Presidente de la República

Margarita Cedeño de Fernández
Vicepresidenta de la República

Antonio Peña Mirabal
Ministro de Educación

Denia Burgos
Viceministra de Educación, Encargada de Servicios Técnicos y Pedagógicos

Rafael Darío Rodríguez Tavares
Viceministro de Educación, Encargado de Asuntos Administrativos y Financieros

Manuel Ramón Valerio Cruz
Viceministro de Educación, Encargado de Certificación Docente

Víctor Ricardo Sánchez
Viceministro de Educación, Encargado de Planificación y Desarrollo Educativo

Adarberto Martínez
Viceministro de Educación, Encargado de Supervisión y Evaluación de la Calidad Educativa

Luis de León
Viceministro de Educación, Encargado de Descentralización

Julio Leonardo Valeirón
Director Ejecutivo del Instituto Dominicano de Evaluación e Investigación de la Calidad Educativa

Aclaración

Este documento está supuesto a ser un anexo de la producción académica en la Línea de Primera Infancia, llevada a cabo por el IDEICE. Mas no pretende ser en sí mismo un producto final, sino un reporte guía que transparente los procesos para generar el insumo principal del *Mapa de Servicios de Atención Integral a la Primera Infancia en República Dominicana*, que es la base de datos consolidada del INAIPI y el SIGERD.

Agradecimientos

Se agradece atentamente, la participación en el proceso de revisión, a la División General de Educación Inicial del MINERD, representada en el proceso por Jahaziely Martínez Corporán, así como al Departamento de Desarrollo e Implementación de Sistemas TIC del INAPI, representado en este proceso por Jairo Moreno y Germán Félix. También se agradece la participación del SIGERD, en la persona de Juan Pablo Berroa.

A su vez, se agradece el fiel seguimiento por parte de Rita Cruz y Daniel Morales, del IDEICE, quienes serán los administradores principales del producto *Mapa de Servicios de Atención Integral a la Primera Infancia en República Dominicana*, que estará disponible dentro de la mayor brevedad posible, en la página del propio IDEICE.

Contenido

Introducción	1
Insumos utilizados	1
Insumos primarios	1
Insumos secundarios	1
Procesamiento e integración de las bases de datos	2
Procesamiento: SIGERD.....	2
Procesamiento: INAIPI	3
Procesamiento: ONE.....	4
Integración.....	4
Resultados	4
Softwares, paquetes utilizados y scripts	5
Softwares y paquetes utilizados	5
Scripts de trabajo	5
Índice de cuadros	
Cuadro 1. Variables de interés incluidas en el SIGERD	2
Cuadro 2. Intervenciones de Educación Inicial presentes en el SIGERD	2
Cuadro 3. Variables de interés incluidas en INAIPI	3
Cuadro 4. Intervenciones de Educación Inicial presentes en el INAIPI	4
Cuadro 5. Paquetes utilizados para la integración INAIPI-SIGERD.....	5

Introducción

El siguiente documento tiene como finalidad explicar la rúbrica de procesos adoptados para crear las bases de datos que tributan al producto *Mapa de Servicios de Atención Integral a la Primera Infancia en República Dominicana*, de la Línea de Investigación en Educación Inicial llevada a cabo por el IDEICE. Este en particular, explica la consolidación de los datos provenientes del Sistema de Gestión de Centros de la República Dominicana (SIGERD), plataforma de datos administrativos utilizada por el Ministerio de Educación (MINERD). A su vez, también se dispuso de los registros acumulados en el Instituto Nacional de Atención Integral a la Primera Infancia (INAIFI), el cual cuenta con toda la matrícula de los centros CAIFI y CAFI, desde el año calendario 2015 hasta el 2019. Por último, de manera auxiliar, se usaron los listados de centros del Instituto Dominicano de Seguros Sociales (IDSS) y de los Centros Modelos de Educación Inicial (CMEI) del MINERD.

Se emplearon técnicas computacionales que serán brevemente descritas en este documento, cuyos scripts estarán disponibles a nivel interno en el portal del IDEICE.

Insumos utilizados

Insumos primarios

Las bases de datos que se tomaron como insumos principales son las siguientes:

1. Cortes del SIGERD, desde el año escolar 2013 - 2014 hasta el 2018 - 2019, a nivel de estudiantes;
2. Las bases de datos entregadas por el INAIFI:
Coordenadas de los centros; Matrícula de atendidos; Movimientos de atendidos.
3. Proyección poblacional de la ONE, segmentada por sexo, fecha y provincia.

Insumos secundarios

Las bases de datos que se tomaron como insumos secundarios son las siguientes:

1. Listado de centros del IDSS.
2. Tabla relacional del SIGERD, con datos provinciales.
3. Tabla relacional creada manualmente, para vincular los CAIFI que aparecen en el SIGERD con los datos reportados por el INAIFI.
4. Shapefiles de regionales y distritos educativos.
5. Listado de los CMEI.

Cabe destacar que estos recursos fueron empleados únicamente para identificar centros ya existentes en el SIGERD que pertenecen a esas intervenciones.

Procesamiento e integración de las bases de datos

A continuación, se describe el procesamiento, dividido en las fuentes de información de datos y su posterior integración. A pesar de disponer de más variables en esta sección solo se mencionarán los utilizados en el procesamiento de los datos que se buscan generar.

Procesamiento: SIGERD

Los cortes del SIGERD son homogéneos, por lo que todos tienen los mismos campos y atributos. Solo se diferencian en que cada uno corresponde a las informaciones relacionadas a sus respectivos años escolares.

Cuadro 1. Variables de interés incluidas en el SIGERD

VARIABLE	CÓDIGO DE LA VARIABLE	DESCRIPCIÓN
Código de 5 dígitos	código_centro	Código principal de centros en el SIGERD
Nombre del centro	centro	Nombre del centro escolar
Grado del estudiante	grado	Grado al que pertenece el estudiante
Sector	sector	Sector del centro escolar
Provincia	prov_nom	Provincia a la que pertenece el centro escolar
Sexo	sexo	Sexo del estudiante

Fuente: Elaboración propia en base al SIGERD.

El código de 5 dígitos fue utilizado para agrupar a los estudiantes en cada uno de sus centros y, a su vez, como llave más adelante para unir los datos en un documento agregado final descrito en el acápite de Integración. Las demás variables se utilizaron para agrupar los estudiantes dentro del propio centro y hacer segmentaciones, especialmente la de sector y el nombre del centro escolar, de donde se creó un nuevo atributo: intervención, el cual se divide en las siguientes categorías:

Cuadro 2. Intervenciones de Educación Inicial presentes en el SIGERD

ETIQUETA DE INTERVENCIÓN	DEFINICIÓN
MINERD-Privado	Centros cuyo sector es el privado
MINERD-Público	Centros cuyo sector es el público, sin denominación distintiva
INAIPI-CAIPI	Centros de denominación CAIPI, dentro de la administración del INAIPI
EPES	Centros dentro de la administración de Vicepresidencia de RD
CMEI	Centros del sector público y son parte de la denominación CMEI
IDSS	Centros que son parte del IDSS

Fuente: Elaboración propia en base al SIGERD.

Cabe señalar que el SIGERD solo empieza a reportar informaciones de estudiantes en los 3 primeros cursos de educación inicial (maternal, infantes y párvulos) a partir del año escolar 2017 - 2018. Los centros marcados como «EPES» son aquellos que, en la variable centro mencionan explícitamente ser de tal denominación (por ejemplo, EPES - Escuela A). La tabla resultante del script de procesamiento de datos del SIGERD termina siendo un panel donde se repiten los centros por año escolar y muestran, por sexo y provincia, la cantidad de estudiantes inscritos en los distintos grados de educación inicial.

Procesamiento: INAIPI

Los registros en las bases del INAIPI no poseen el código de 5 dígitos del SIGERD. A su vez, un centro puede ser regido por códigos: de centro y de red. Estas y otras variables utilizadas en el procesamiento son descritas en la siguiente tabla:

Cuadro 3. Variables de interés incluidas en INAIPI

VARIABLE	CÓDIGO DE LA VARIABLE	DESCRIPCIÓN
Código del centro	id_centro	Código principal de centros en el INAIPI
Código de la red	id_red	Código de la red a la que pertenece el centro
Tipo de centro	tipo_centro	Tipo de centro (CAIPI o CAFI)
Coordenadas	coordenadas	Coordenadas de geolocalización del centro escolar
Provincia	provincia	Provincia a la que pertenece el centro escolar
Sexo	sexo	Sexo del estudiante
Inicio en servicio	fecha_inicio_en_inaipi	Fecha de inicio a ser atendido en el INAIPI
Salida del servicio	fecha_salida	Fecha de salida a ser atendido en el INAIPI
Fecha nacimiento	fecha_nacimiento	Fecha en la que el atendido nació
Fecha movimiento	fecha_movimiento	Fecha en la cual el atendido realizó un movimiento

Fuente: Elaboración propia en base al INAIPI.

Al ser bases de datos casi totalmente planas, se procedió a crear divisiones, siguiendo el patrón de cohortes del SIGERD. Para ello se utilizó el calendario escolar del MINERD de los años escolares 2015 - 2016 hasta el 2018 - 2019, para definir cuándo inician y cuándo finalizan los años escolares. La separación de los atendidos en función de sus respectivos años escolares se marcaba cuando el mismo tuviera una fecha de inicio a tomar clases en algún centro del INAIPI menor a la del inicio del año escolar y que su fecha de salida, en caso de existir, por lo menos fuera mayor o igual a la fecha de inicio del año escolar posterior al evaluado de manera específica. Por ejemplo, si el atendido tomó atención integral en el año escolar 2016 - 2017, se verificaba que su fecha de inicio en el INAIPI fuera menor a la del inicio del año escolar y que su fecha de salida sea mayor o igual a la del inicio del año escolar 2017 - 2018.

Esto, a su vez, se corrigió con la base de datos de movimientos, la cual contiene información de si el atendido migró de un centro a otro; esto con el objetivo de indicar de manera precisa el último centro al que el atendido asistió dentro del año escolar a evaluar.

La variable de edad se utilizó para estimar en cuál grado teórico el atendido estaría, dado a que en la base del INAIPI existen registros de estudiantes de centros CAFI, que no necesariamente reciben actividades pedagógicas dentro de una infraestructura escolar. Con ella también se excluyeron algunos registros inconsistentes, como de estudiantes con 6 años o más al inicio del año escolar, los cuales ya están en edad de tomar clases en primaria de manera obligatoria, respondiendo a la Ley General de Educación. La fecha de corte para calcular la edad fue el 31 de diciembre de cada primer año calendario que traslape un año escolar. Es decir, si es la edad para el año escolar 2017-2018, se calculaba la edad al 31 de diciembre del 2017¹.

1 Ese es el criterio asumido por el MINERD para aceptar estudiantes en un grado específico.

Al disponerse de cohortes de atendidos por año escolar, se reportaron agrupados por tipo de intervención, en las siguientes categorías:

Cuadro 4. Intervenciones de Educación Inicial presentes en el INAIPI

ETIQUETA DE INTERVENCIÓN	DEFINICIÓN
INAIPI-CAIPI	Centros de denominación CAIPI, dentro de la administración del INAIPI
INAIPI-CAFI	Centros de denominación CAFI, dentro de la administración del INAIPI

Fuente: Elaboración propia en base al SIGERD.

Al final quedó un panel con una estructura similar a la descrita para el SIGERD. Con las coordenadas de los centros, se extrajeron sus datos de regionales y distritos educativos a partir de la información geoespacial de las capas cartográficas del MINERD (*shapefiles*).

Procesamiento: ONE

Se procedió a trabajar con las estimaciones poblacionales que están disponibles al público en la página de la ONE. La rúbrica de procesamiento fue simple: se conservó la población del país, dividida por provincia y sexo de aquellas personas en edad teórica de recibir educación inicial.

Integración

Se creó una tabla relacional de los centros CAIPI que aparecen tanto en el SIGERD como en el INAIPI, a fines de no duplicar información a la hora de integrar ambas fuentes. Esta tiene tanto el código utilizado por el INAIPI, del centro, como su contraparte en la plataforma del SIGERD. Ya disponiendo de llaves comunes entre bases de datos, se procedió con la consolidación. Se unieron ambos paneles en función de su año escolar, su tipo de intervención y el código del centro. Esto para constatar de un consolidado realizado por centro escolar, pero para los fines primordiales del mapa, también se hizo una integración por provincias. Ya con una base general del SIGERD y del INAIPI, se creó una serie de variables de cantidad de estudiantes/atendidos matriculados en función del grado². La regla es la siguiente: ya como hay casos donde grados se repiten (centros que aparecen tanto en las bases del SIGERD como en las del INAIPI), se indicó que la cantidad de atendidos matriculados final, por grado, iba a ser igual a los que salen reportados en las bases del INAIPI, cuando la intervención es *INAIPI-CAFI* o *INAIPI-CAIPI*; en caso contrario, lo que reporte el SIGERD.

Al final, el insumo que tributa al mapa es un documento en formato *xlsx*, reportando todos los datos transformados que se discutieron previamente.

Resultados

Los resultados del mapa serán de acceso al público general, a través de la página oficial del IDEICE. Los mismos serán discutidos más adelante en el documento académico del cual este reporte es un anexo.

² Se hace la aclaración que las estrategias guiadas por el INAIPI no se basan principalmente en la asistencia a un centro escolar, sino en la atención integral de los niños y niñas. La clasificación por «grado» se hizo para homogeneizar todas las intervenciones, pero estas operan bajo conceptos distintos.

Softwares, paquetes utilizados y scripts

Softwares y paquetes utilizados

Para preparar la base de la ONE, se utilizó el software MS Excel 2016, ya que esta no requería de ningún tratamiento sofisticado. En cuanto a las bases del SIGERD, INAIPI y el consolidado, se prepararon utilizando el lenguaje de programación R, en su versión 3.6.0, auxiliado con RStudio 1.2.1335. Los paquetes utilizados en R para las tareas específicas fueron los siguientes:

Cuadro 5. Paquetes utilizados para la integración INAIPI-SIGERD

PAQUETE	VERSIÓN DEL PAQUETE	RAZÓN DE USO
tidyverse	1.2.1	Gestión y visualización de datos
tmap	2.2	Análisis geoespacial
janitor	1.2.0	Limpieza de nombres de campos
openxlsx	4.1.0	Manipulación de datos en formato de Excel
readxl	1.3.1	Manipulación de datos en formato de Excel
sf	0.7-4	Gestión de datos geoespacial
lubridate	1.7.4	Manipulación de datos en formato de Excel
rgdal	1.4-3	Gestión de datos geoespacial
maptools	0.9-	Gestión de datos geoespacial

Fuente: Elaboración propia.

Las versiones de tales paquetes son las que se encontraban en R 3.6.0 y R 3.6.1.

Scripts de trabajo

Los scripts de la gestión de datos en R fueron almacenados. Estos son cuatro: uno para los datos del INAIPI, otro para los del SIGERD, un tercero que integra todas las fuentes de información y el último realiza los cálculos de indicadores. Estos están disponibles en el IDEICE.



Instituto Dominicano de Evaluación e
Investigación de la Calidad Educativa

www.ideice.gob.do



ISBN 978-9945-499-58-2

