

## **TERCER INFORME ESTADO DE LA EDUCACIÓN**

Megabases de datos georreferenciados para la educación  
primaria y secundaria de los centros educativos de Costa Rica  
(2000-2009)

***Investigador:***  
***Dagoberto Murillo Delgado***

***Con el apoyo de:***  
***Antonella Mazzei***

**Abril, 2011**

# Megabases de datos georreferenciados para la educación primaria y secundaria de los centros educativos de Costa Rica (2000-2009)

---

## Índice

<b>INTRODUCCIÓN</b> .....	<b>4</b>
<b>ASPECTOS METODOLÓGICOS</b> .....	<b>6</b>
ETAPA I. AGRUPACIÓN DE BASES DE DATOS DISPERSAS .....	7
ETAPA II. ELABORACIÓN DE DICCIONARIOS DE VARIABLES .....	8
ETAPA III. CONSTRUCCIÓN DE MATRIZ DE INFORMACIÓN .....	9
ETAPA IV. PROBLEMAS ENCONTRADOS EN LA INFORMACIÓN DE BASE .....	10
<i>Centros con información duplicada</i> .....	11
<i>Centros con información en dos filas, pero una de ellas con variables de seguimiento iguales a cero</i> .....	11
<i>Centros con errores en el código secuencial o en la variable “rama académica”</i> .....	12
<i>Centros de educación secundaria con una sección nocturna</i> .....	12
<i>Centros de educación secundaria con matrícula académica y técnica</i> .....	13
<i>Cambios en centros educativos</i> .....	14
ETAPA V. GEORREFERENCIACIÓN .....	15
ETAPA VI. VERIFICACIÓN DE LA INFORMACIÓN .....	17
<b>ESTRUCTURA DE LAS MEGABASES</b> .....	<b>19</b>
INFORMACIÓN Y CARACTERÍSTICAS GENERALES DEL CENTRO EDUCATIVO .....	20
<i>Código secuencial y código presupuestario</i> .....	20
<i>Nombre del centro educativo</i> .....	20
<i>Año de creación</i> .....	21
<i>Dependencia</i> .....	21
<i>Tipo de jornada (exclusivo de primaria)</i> .....	21
<i>Imparte el servicio de preescolar (exclusivo de primaria)</i> .....	21
<i>Servicios de educación especial (exclusivo de primaria)</i> .....	21
<i>Pertenencia a programas (exclusivo de primaria)</i> .....	21
<i>Rama académica (exclusivo de secundaria)</i> .....	21
<i>Es telesecundaria (exclusivo de secundaria)</i> .....	22
<i>Estructura curricular (exclusivo de secundaria académica pública)</i> .....	22
VARIABLES DE SEGUIMIENTO .....	22
<i>Matrícula inicial total</i> .....	22
<i>Repitentes</i> .....	22
<i>Porcentaje de repitencia</i> .....	22
<i>Deserción intra-anual</i> .....	22
<i>Abandono</i> .....	23
<i>Porcentaje de abandono</i> .....	23
RENDIMIENTO .....	23
<i>Aprobación</i> .....	23

<i>Reprobación</i> .....	23
MATRÍCULA INICIAL DE ESTUDIANTES EXTRANJEROS .....	23
SEGUNDO IDIOMA (EXCLUSIVO PARA PRIMARIA) .....	24
ALUMNAS EMBARAZADAS EN EDUCACIÓN TRADICIONAL.....	24
CASOS DE VIOLENCIA EN LOS CENTROS EDUCATIVOS .....	24
ADECUACIONES CURRICULARES EN EDUCACIÓN TRADICIONAL .....	24
INFRAESTRUCTURA EDUCATIVA .....	25
PERSONAL.....	25
<i>Personal docente-administrativo</i> .....	25
<i>Personal docente</i> .....	25
<i>Personal docente titulado</i> .....	25
<i>Personal docente autorizado</i> .....	26
<i>Personal docente aspirante</i> .....	26
<i>Porcentaje de docentes titulados</i> .....	26
<i>Porcentaje de docentes no titulados</i> .....	26
<i>Personal administrativo y de servicios</i> .....	26
BACHILLERATO (SOLO SECUNDARIA) .....	26

## Introducción

El *Informe Estado de la Educación* es una iniciativa impulsada por el Consejo Nacional de Rectores (Conare) y el Programa Estado de la Nación, que tiene como principal objetivo proveer al país conocimiento actualizado sobre el desempeño de la educación costarricense.

En especial, se pretende consolidar un sistema de seguimiento que ayude a determinar cuánto se acerca o se aleja Costa Rica de la aspiración de ofrecer oportunidades para que su población tenga acceso equitativo a una educación de calidad, esto es, una educación que promueva el desarrollo de capacidades que les permitan a las personas realizarse individualmente y, a la vez, contribuir al bienestar colectivo.

El Informe utiliza distintas fuentes de información. Su fuente primaria son los hallazgos de sus propios estudios, pero también emplea información proveniente de investigaciones realizadas por terceros y los registros administrativos de instituciones públicas (fuentes secundarias). Los registros administrativos no solo son útiles para generar estadísticas oficiales, sino para diseñar políticas de mejora del sistema educativo. Es importante saber cuántos niños, niñas y jóvenes están matriculados en las escuelas y colegios, cuántos de ellos reprueban o desertan cada año, o cuántas aulas se encuentran en buen estado; todos estos son insumos esenciales para diagnosticar, establecer metas, planificar, evaluar resultados y dar seguimiento a los programas.

La información de los registros administrativos se publica anualmente con una desagregación máxima a nivel de direcciones regionales del Ministerio de Educación Pública (MEP). Estas últimas son unidades geográficas creadas para organizar la prestación de los servicios y facilitar la atención de las comunidades educativas (decreto 35513-MEP, 2010). Los datos disponibles incluyen series de tiempo para algunos indicadores principales; sin embargo, es necesario incorporar la dimensión espacial para entender mejor los diversos fenómenos educativos que se dan en el territorio.

La georreferenciación de los centros educativos consiste en la localización espacial de las instituciones de enseñanza primaria y secundaria en un sistema de coordenadas, mediante el registro de su latitud y longitud. Para Costa Rica, este proceso constituye un paso importante en la consolidación un Sistema de Información Geográfica (SIG), entendido como una base de datos computarizada con información espacial. Este es un valioso insumo para la toma de decisiones estratégicas y operativas en el sistema educativo; ayuda en la planificación de políticas públicas como la inversión en infraestructura, la construcción de establecimientos en lugares seguros, la asignación de docentes y la ampliación de la cobertura educativa, entre otros.

La importancia de construir megabases de datos georreferenciados de los centros educativos del país radica no solo en la utilidad de tener representaciones de ciertos indicadores, sino en su potencial para motivar la investigación y ahondar en el análisis

espacial. Tanto la base de primaria como la de secundaria agrupan un conjunto de variables para el período 2000-2009 que antes estaban dispersas en múltiples bases. De este modo, las personas interesadas encontrarán información de once temas distintos a nivel de centros educativos para un período de diez años.

Estas bases son un primer resultado y se espera que, con el tiempo, el período de referencia que abarcan, así como el número de temáticas incluidas, vayan creciendo. Actualmente contienen información de los Departamentos de Análisis Estadístico y de Evaluación Académica y Certificación del MEP, pero se espera que el sistema se consolide a partir de nuevos datos a nivel de centros educativos, provenientes de otras dependencias e instituciones.

El propósito de este documento es servir como guía para el manejo de las megabases y como fuente de consulta sobre la terminología empleada en las distintas variables, sus alcances y limitaciones. El texto se organiza en tres partes, incluyendo la presente introducción. El segundo apartado describe las etapas del proceso de construcción de las megabases, e incluso expone situaciones específicas en las que fue preciso modificar las bases que contenían los registros iniciales, dejando constancia, en cada caso, del nombre del centro educativo, el nivel, la temática, el año y la base de datos en la que se realizó la variante. La tercera sección detalla la estructura de las megabases, es decir, cómo están ordenadas y cuáles son los temas y subtemas incluidos en ellas.

Por la información provista se agradece a Eliécer Ramírez y Dixie Brenes, del Departamento de Análisis Estadístico del MEP; Félix Barrantes y Lilliam Mora, de la Dirección de Gestión y Evaluación de la Calidad del MEP; Patricia Solano y Alexander Campos, del Área de Censos de Población y Vivienda del INEC; al Grupo Nación y, en especial, a Rosendo Pujol, Leonardo Sánchez, Eduardo Pérez y Juan José Castro, del Programa de Investigación en Desarrollo Urbano Sostenible (ProDUS) de la Universidad de Costa Rica

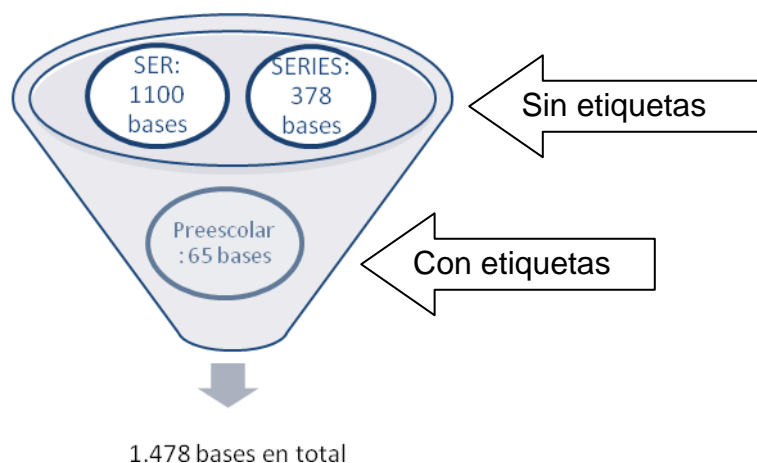
## Aspectos metodológicos

En el año 2009, en el marco de la preparación del *Tercer Informe Estado de la Educación*, el Programa Estado de la Nación se propuso desarrollar un conjunto de nuevos instrumentos para profundizar en el análisis de la educación en Costa Rica. Con ese objetivo, solicitó información a nivel de centros educativos a varias dependencias del MEP y a otras instituciones ligadas al sector.

Se recibió un total de 1.478 bases de datos en formato \*.DBF, provenientes del Departamento de Análisis Estadístico del MEP. Estas bases estaban en dos carpetas (SER y SERIES), dentro de la cuales había nueve subcarpetas para cada uno de los años del período 2000-2008. Estas bases no contaban con un diccionario de variables, ni con etiquetas. Al mismo tiempo, se recibió una carpeta para el nivel preescolar con 65 bases en SPSS (formato \*.SAV) con sus respectivas etiquetas, las cuales también se encontraban en las carpetas SER y SERIES sin etiquetar (bases duplicadas). Esto redujo el número de bases sin etiquetar a 1.413 (figura 1).

**Figura 1**

### Primera solicitud de información al Departamento de Análisis Estadístico



Fuente: Elaboración propia.

Posteriormente se recibió del mismo Departamento un total de 75 bases en formato Excel (\*.XSL) que sí contaban con un diccionario de variables. De estas, 54 correspondían al año 2009 (solo incluían preescolar, escuelas diurnas y colegios tradicionales). A su vez, el Departamento de Evaluación Académica y Certificación aportó las diez bases del *ranking* de colegios en las pruebas de bachillerato para el período 2000-2009.

Como se indicó antes, la información que generan y publican estas dependencias se encuentra desagregada a un nivel “máximo” de región educativa, una clasificación que ha cambiado con el tiempo, debido a la creación de nuevas direcciones regionales.

Ambas situaciones, la dispersión de la información en múltiples bases y su presentación a un nivel agregado que ha perdido comparabilidad con el tiempo, evidenciaron la necesidad de contar con una base que reuniera un conjunto de temas para el período 2000-2009 a nivel de centros educativos, con posibilidades de ampliación. Esto implicó abandonar la idea de tener múltiples bases para cada tema por año.

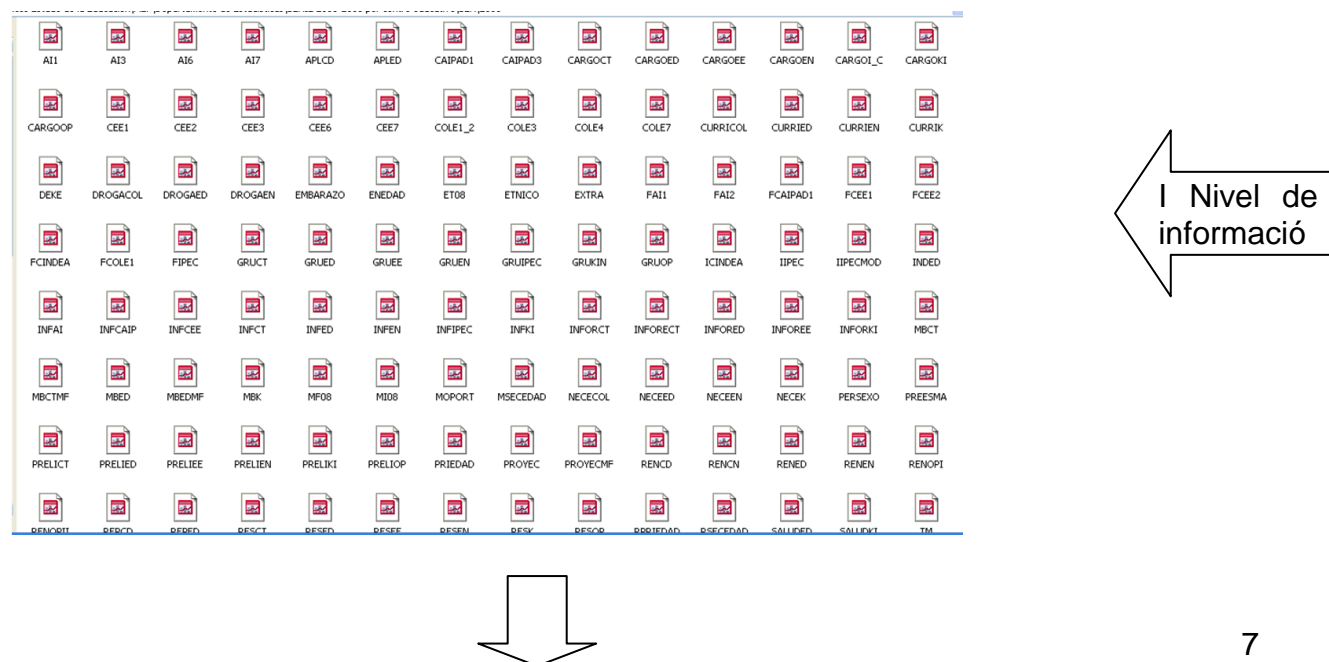
Para alcanzar esa meta, el proceso se realizó en diferentes etapas, que se describen a continuación.

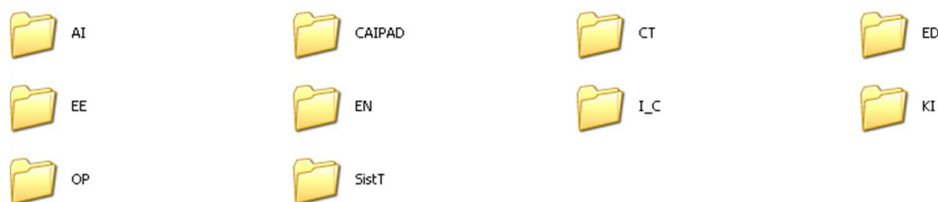
### Etapa I. Agrupación de bases de datos dispersas

Las bases fueron agrupadas por nivel educativo y modalidad, es decir: preescolar, educación especial, escuelas diurnas, escuelas nocturnas, secundaria tradicional y secundaria no tradicional como Institutos Profesionales de Educación Comunitaria (IPEC); Centros Integrados para la Educación de Jóvenes y Adultos (Cindea); Colegio Nacional de Educación a Distancia (Coned); Centro de Atención Integral para Adultos con Discapacidad (Caipad), entre otras. No obstante, existen bases que aglutinan los diferentes niveles, como las relativas a violencia en los centros educativos, embarazos en la educación tradicional o alumnos extranjeros; estas fueron reunidas en una carpeta denominada "SistT" (Figura 2). Cada una de las bases incluidas en las carpetas SER y SERIES posee una variable "nivel" que se utilizó para generar la clasificación antes indicada; esta variable posee un código para cada uno de los niveles y modalidades de la oferta educativa del MEP. Se encuentra codificada de la siguiente forma: 1, preescolar; 2, educación especial y Caipad; 3, primero y segundo ciclos; 4, tercer ciclo y educación diversificada; 5, escuelas nocturnas; 6, Cindea e IPEC; 7, programa Nuevas Oportunidades.

Figura 2

### Clasificación de bases de datos del Departamento de Análisis Estadístico de acuerdo con el nivel educativo y la modalidad





## *Etapa II. Elaboración de diccionarios de variables*

El siguiente paso fue crear diccionarios o listados de variables con sus correspondientes significados o etiquetas. Esto se hizo para cada una de las bases que incluían escuelas diurnas y colegios tradicionales, las que, en conjunto con las bases para varios niveles, representan el 57,4% del total recibido (769 bases solo en escuelas y colegios). Varias razones llevaron a hacer esta selección: la continuidad de la información disponible para el período 2000-2009, los porcentajes que representaron estas dos modalidades en la matrícula del sistema educativo (78,9%) y dentro del total de instituciones en la educación regular (95,4%) en 2009, además del costo asociado a georreferenciar la totalidad de centros educativos en esta primera etapa.

Para la creación de los diccionarios de variables se utilizaron como guías los formularios de “Matrícula inicial y personal” y de “Matrícula final” que emplea el Departamento de Análisis Estadístico para recolectar los datos. Es importante acotar que esos cuestionarios pueden tener diferencias importantes en cuanto a la cantidad de información que se solicita y la forma en que ésta se registra en los diferentes niveles y modalidades; además pueden cambiar de un año a otro. En el cuadro 1 se resumen las principales temáticas que se recopilan mediante estos instrumentos.



**Cuadro 1. Principales temáticas incorporadas en los formularios de matrícula inicial y personal y matrícula final**

<b>Formulario y temática</b>	
<b>Matrícula inicial y personal</b>	<b>Matrícula final</b>
Matrícula inicial y número de secciones	Movimientos de matrícula: matrícula inicial, nuevos ingresos, traslados, abandonos, matrícula final
Alumnos repitentes	Matrícula final de estudiantes con adecuación curricular
Residencia y matrícula de los alumnos durante el curso lectivo.	Matrícula final de estudiantes integrados (discapacidad)
Alumnos extranjeros	Participación en programas y/o actividades de prevención del uso indebido de drogas
Alumnos que aprobaron todas las asignaturas (convocatoria)	Consumo de drogas
Personal total	Embarazos
Personal docente-administrativo y docente por grupo profesional (actualmente solo se registra personal docente)	Matrícula final de estudiantes en programas de educación abierta
Matrícula inicial en programas de educación abierta	Indicadores de violencia en los centros educativos
Infraestructura educativa y servicios	

Fuente: Elaboración propia.

### *Etapa III. Construcción de matriz de información*

Con los insumos obtenidos en la etapa previa, se procedió a elaborar una matriz en la cual se describe la información que contienen las bases, ordenada por filas. En las columnas de la matriz se indica cuáles bases, en un año determinado, tienen la información respectiva. Este ordenamiento permitió identificar cuáles bases son similares en cuanto a los datos disponibles para el período 2000-2009 y, consecuentemente, cuáles temáticas se podrían incorporar a las megabases. Las temáticas seleccionadas fueron las siguientes:

- Matrícula inicial por año cursado y sexo, según edad en años cumplidos.
- Repitentes por año cursado y sexo, según edad en años cumplidos.
- Matrícula inicial, matrícula final y deserción por sexo y año cursado. A partir del 2001 las bases incluyen estudiantes provenientes de otras sedes, traslados a otras sedes y abandonos (deserción).
- Alumnos extranjeros por sexo, según nacionalidad.
- Estudiantes con adecuación según tipo de adecuación, por año cursado y sexo.
- Rendimiento preliminar y definitivo (aprobados y reprobados), por año cursado y sexo. Al final del proceso solo se considero rendimiento definitivo.
- Violencia en centros educativos.
- Embarazos.

- Personal total y docente por grupo profesional (posteriormente se agrupó en aspirantes, autorizados y titulados); solo se incluyeron datos a partir del año 2005 donde las bases que se encuentran son más homogéneas.
- Infraestructura y servicios. En este caso, aunque la información que captan los formularios ha cambiado, hay un conjunto importante de variables que se mantienen para todo el periodo, o bien que, por su relevancia, se empezaron a captar en 2002 o 2003 (cuadro 2).

**Cuadro 2. Información<sup>a/</sup> sobre infraestructura educativa recopilada en el formulario de matrícula inicial y personal**

<b>Espacio físico, mobiliario y equipo</b>	<b>Período incluido</b>
Aulas académicas	2000-2009
Aulas de asignaturas especiales	2000-2009
Biblioteca <sup>b/</sup>	2000-2009
Comedor	2000-2009
Laboratorio de Informática	2002-2009
Laboratorio de Ciencias	2002-2008
Talleres	2003-2009
Sala de profesores	2002-2009
Inodoros	2000-2009
Lavatorios	2000-2009
Servicio sanitario accesible (Ley 7600)	2002-2009
Pupitres unipersonales	2000-2009
Mesas de pupitres	2000-2009
Sillas de pupitre	2000-2009
Televisión	2000-2009
VHS o DVD	2000-2009
Computadoras para administrativos	2002-2009
Computadoras para estudiantes <sup>c/</sup>	2002-2009
Pizarras	2002-2009

a/ Solo se incluyen las variables que se presentan con mayor regularidad en los años más recientes. No obstante, para años anteriores al 2009 se solicitó información adicional sobre otros espacios físicos, mobiliario y equipo.

b/ A partir de 2008 la solicitud de información corresponde a un espacio físico.

c/ A partir de 2008, se pide incluir las computadoras del laboratorio o laboratorios.

Fuente: Elaboración propia con base en los formularios de matrícula inicial y personal del MEP.

*Etapa IV. Problemas encontrados en la información de base*

Dentro de las megabases, el resultado final esperado es asignar un par de coordenadas geográficas a un centro educativo, es decir, que toda la información de ese centro se encuentre en una única fila de la base. Es por ello que cobró especial relevancia el uso del código secuencial de centros educativos elaborado por el Departamento de Análisis Estadístico, como variable de enlace entre las distintas bases. En cuanto al código presupuestario, aunque muy importante en el ámbito de los

planteles públicos, presenta el problema de que asigna una cero (“0”) a todos los centros privados.

El primer paso en esta etapa consistió en pasar de la agrupación de las bases por nivel (como se hizo en el proceso de creación de variables), a la agrupación de acuerdo con las temáticas, para tener, por ejemplo, una carpeta con la información para escuelas diurnas de la matrícula inicial por año cursado y sexo según edad en años cumplidos para el periodo 2000-2009. Una vez realizadas las distintas agrupaciones nivel-temática se procedió a verificar la existencia de un código secuencial único para cada institución. Este ejercicio arrojó distintos resultados, entre ellos que había duplicaciones en el código secuencial. Estos casos, que ya fueron expuestos al Departamento de Análisis Estadístico del MEP, obedecen a los motivos que se detallan a continuación.

### Centros con información duplicada

Se refiere a la existencia de dos filas con información idéntica dentro de una misma base, para un determinado centro educativo. En este caso se eliminó una de las filas duplicadas, para garantizar la existencia de una única fila por centro educativo. Esta situación se encontró en dos de las bases exploradas (cuadro 3).

**Cuadro 3. Centros educativos presentes en dos filas de la base, con información duplicada**

Nombre del centro	Nivel	Nombre de la base	Tema	Año
Corazón de Jesús	Primaria	curried	Adecuaciones curriculares	2000
Térraba	Secundaria	rsecedad	Repitencia	2008

### Centros con información en dos filas, pero una de ellas con variables de seguimiento iguales a cero

En estos casos la información del centro educativo se registra en dos filas separadas dentro de una misma base pero, a diferencia del caso anterior, las variables de seguimiento en una de las filas están compuestas por ceros (cuadro 4). Esas filas fueron eliminadas.

**Cuadro 4. Centros educativos presentes en dos filas de la base, variables de seguimiento igual a cero**

Nombre del centro	Nivel	Nombre de la base	Tema	Año
Telesecundaria Concepción de Pilas	Secundaria	msecedad	Matrícula inicial	2000
Colegio Técnico Profesional de Jicaral	Secundaria	msecedad	Matrícula inicial	2009

### Centros con errores en el código secuencial o en la variable “rama académica”

Son casos en los que se presenta duplicación en el código secuencial, debido a un error en la digitación del código o en la variable “rama académica”. En el segundo caso se trata de colegios que aparecen dos veces con ramas distintas y en ocasiones con variables de seguimiento iguales a cero (cuadro 5).

**Cuadro 5. Centros educativos con errores en el código secuencial o la variable “rama académica”**

Nombre del centro	Nivel	Nombre de la base	Tema	Año
La Esperanza	Primaria	priedad	Matrícula inicial	2002
Navajuelar	Primaria	priedad	Matrícula inicial	2002
Colegio Técnico Profesional San Sebastián	Secundaria	rsecedad	Repitencia	2001
Colegio Técnico Profesional Calle Blancos	Secundaria	rsecedad	Repitencia	2001, 2002
Colegio Técnico Profesional de Venecia	Secundaria	rsecedad	Repitencia	2001
Colegio Técnico Profesional Ricardo Castro Beer	Secundaria	rsecedad	Repitencia	2004
Experimental Bilingüe Santa Cruz	Secundaria	rsecedad	Repitencia	2004
Telesecundaria Jardín de Páramo	Secundaria	rsecedad	Repitencia	2004
Telesecundaria Río Nuevo	Secundaria	rsecedad	Repitencia	2004
Liceo Dos Ríos	Secundaria	infct	Infraestructura	2002

Cuando se determinó la existencia de un error en el código secuencial, se buscó el código correcto; si éste ya existía, se comparó la información y se eliminó aquella con datos iguales a cero en las variables de seguimiento.

### Centros de educación secundaria con una sección nocturna

Corresponde a instituciones de secundaria, tanto técnicas como académicas, que imparten lecciones en horario diurno y que, adicionalmente, cuentan con una sección nocturna (cuadro 6). Se registran sobre todo en el período 2000-2002.

**Cuadro 6. Duplicidad en el código secuencial en colegios con secciones nocturnas. 2000 y 2001**

Nombre del centro	Nombre de la base	Tema	Años
Paraíso	msecedad	Matrícula inicial	2000, 2001
Maurilio Alvarado Vargas	rsecedad	Repitencia	2000
Experimental Bilingüe de La Cruz	ctde	Deserción	2000, 2001
Colegio Técnico Profesional de Buenos Aires	infcot	Infraestructura	2000
Colegio Técnico Profesional de Osa			
Quepos			
Colegio Técnico Profesional Guaycará			
Colegio Técnico Profesional de Siquirres			
Colegio Técnico Profesional Batán			
Nocturno de Río Frío			
Pacífico Sur			

En estos casos se indagó en bases posteriores al 2002, y se encontró que las secciones nocturnas tienen un código secuencial distinto. Se procedió entonces a asignar a los colegios respectivos (secciones nocturnas), para los años 2000 y 2001, el código secuencial de años posteriores al 2002.

#### Centros de educación secundaria con matrícula académica y técnica

Se refiere al caso específico del Colegio Técnico Profesional San Isidro (cuadro 7), para el cual, durante 2000 y 2001, se registró la información de la matrícula y repitencia del tercer ciclo (séptimo, octavo y noveno años) como académica y la de la educación diversificada (décimo, undécimo y duodécimo) como técnica. Como alternativa se creó el código “rama 41”, cuyo significado es precisamente: tercer ciclo académico y educación diversificada técnica. De esta forma se puede distinguir entre matrícula académica y técnica a nivel nacional. Por ejemplo, si el usuario de la megabase de secundaria desea conocer la matrícula académica en 2000 o 2001, debe sumar la matrícula de todos los centros académicos (diurnos y nocturnos) más la matrícula en el tercer ciclo del colegio citado.

**Cuadro 7. Centros educativos con matrícula en las modalidades académica y técnica**

Nombre del centro	Nombre de la base	Tema	Año
Colegio Técnico Profesional San Isidro	msecedad	Matrícula inicial	2000, 2001
	rsecedad	Repitencia	2000, 2001

Cabe destacar que el código secuencial impone limitaciones para construir bases que agrupen diferentes niveles educativos. Es decir, una institución que ofrece los servicios de preescolar, primaria y secundaria puede presentar tres códigos secuenciales distintos, según sea el nivel de la base en que se encuentre. Por ejemplo, el centro

educativo José Fidel Tristán, ubicado en el distrito Merced, cantón de San José, tiene el código secuencial 93 en preescolar, el 10 en escuelas diurnas y el 2 en colegios tradicionales. Esta situación plantea el reto de contar con una base que agrupe todos los niveles preservando la condición de una sola fila por plantel.

Esta razón, unida al hecho de que existen instituciones que prestan diferentes servicios educativos en ubicaciones geográficas distintas (coordinadas geográficas diferentes) motivó la construcción de dos megabases, una para primaria y otra para secundaria, en lugar de una. La primera cuenta con 3.362 variables y agrupa información de 103 bases. La megabase secundaria consta de 3.007 variables y reúne los datos de 117 bases. Es importante recalcar que las bases originales empleadas para la construcción de estas megabases tienen más variables.

### Cambios en centros educativos

Un elemento adicional de suma importancia es la presencia de vacíos de información en las megabases. Se identificaron tres razones por las cuales esto ocurre, a saber:

- **Centros que desaparecen:** son instituciones que presentan información en los primeros años incluidos en las bases, pero que en los años siguientes no registran datos debido a que cerraron sus puertas. Por ejemplo, el Colegio John Dewey, ubicado en Cipreses en Curridabat, no tiene información a partir del 2004.
- **Centros que se crean o códigos que se reabren:** son centros que no están en los primeros años de la base pero aparecen posteriormente. Un ejemplo es el Liceo Rural Kjakuo Sulo, en Chirripó de Turrialba, que solo presenta información para 2009.
- **Centros con información no registrada:** son planteles que existen en el período considerado pero que, por diferentes motivos, no registran información para un año o grupo de años, o bien sobre un tema particular. Algunos reportan la matrícula inicial y personal, pero no tienen información de matrícula final. El cuadro 8 presenta la lista de colegios que poseen datos de matrícula inicial y personal, pero no de deserción.

**Cuadro 8. Colegios con información de matrícula inicial, pero con vacíos de información en el tema de deserción y abandono. 2008**

Instituto Julio Acosta García	Telesecundaria Barra de Tortuguero
Pindeco	Complementaria Cahuita
Costa Rica Academy	Gilante
Colegio Valle Azul	Santa Ana High School
Bilingüe San Francisco de Asís	Amigos de Monteverde
Bilingüe del Valle	Marian Baker
Sagrada Reina de los Ángeles	Liceo Rural Alto Guaymí
Europeo	Valle del Sol
Green Valley	Telesecundaria San Carlos Pacuarito
San Diego	Valley Forge Future
Colegio Santa Josefina	

La megabase de secundaria además combina información del Departamento de Evaluación Académica y Certificación del MEP, sobre los resultados de las pruebas de bachillerato. Estas bases por centro educativo cuentan con un código distinto al secuencial del Departamento de Análisis Estadístico y al código presupuestario del MEP. Es por ello que, como paso adicional, se elaboró una tabla de equiparación, de tal modo que a un único código secuencial de la base del Departamento de Análisis Estadístico corresponda un código único en la base de bachillerato.

Uno de los inconvenientes encontrados en las bases del *ranking* de colegios en las pruebas de bachillerato, es que el código de una institución puede variar entre un año y otro. Esto sucede principalmente en las telesecundarias, cuyos códigos se modificaron en 2006 y 2008. Los planteles de Centro de Atención Integral (CAI), Institutos Profesionales de Educación Comunitaria (IPEC), Centros Integrados para la Educación de Jóvenes y Adultos (Cindea) y Colegio Nacional de Educación a Distancia (Coned) fueron excluidos de la megabase. No se encontraron códigos equivalentes para la telesecundaria de Cabécar Tayni, ni para los colegios nocturnos de San Carlos (sede central), San Carlos (Monterrey) y Guatuso.

Para la unión de todas las bases se utilizó el programa Stata 11.

#### *Etapa V. Georreferenciación*

La georreferenciación realizada por el Programa de Investigación en Desarrollo Urbano Sostenible (ProDUS) inició paralelamente con la etapa I de este proceso. Esta labor consiste en asignar una coordenada geográfica a cada uno de los centros educativos, lo cual implicó un intenso trabajo de localización y verificación de la información.

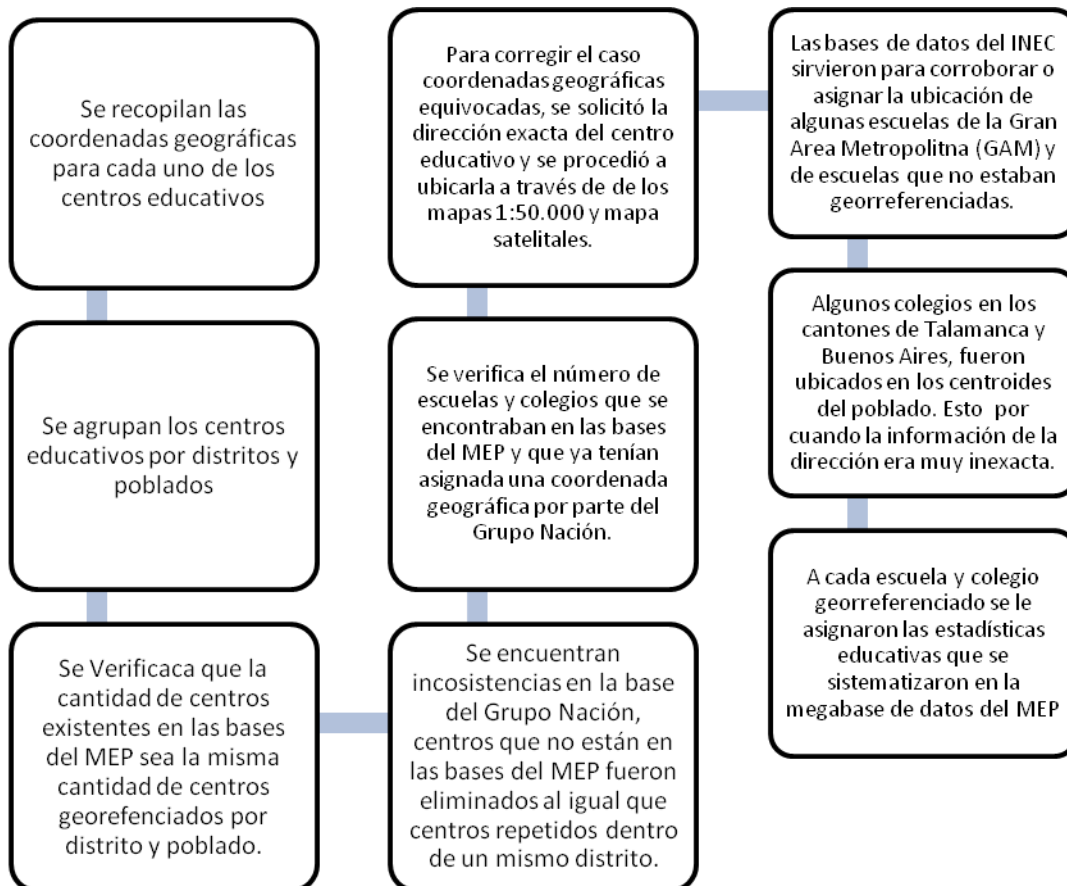
ProDus examinó cuatro bases de datos que registran información a nivel de centros educativos:

- Base de datos de ProDUS: tiene registrada la ubicación exacta de gran parte de los centros educativos de primaria y la totalidad de los centros de secundaria en los siguientes cantones: Pococí, Osa, Golfito, Corredores, Pérez Zeledón, San Ramón, San Carlos (cuatro distritos), Orotina, Alajuela y los otros 31 cantones de la Gran Área Metropolitana (GAM).
- Base de datos del INEC: corresponde a las zonas del país donde se ha realizado levantamiento de campo, especialmente dentro de la GAM, para la elaboración del Censo de Población y Vivienda del 2011. Tiene algunos centros educativos georreferenciados.
- Base de datos del Grupo Nación (mapas): contiene una lista de escuelas georreferenciadas.

- Bases de datos del MEP: incluye estadísticas de todos los centros educativos del país, tanto de primaria como de secundaria. Los datos no están georreferenciados.

El arduo proceso que se llevó a cabo para georreferenciar los centros educativos se explica brevemente en la figura 2.

**Figura 2. Procedimiento para georreferenciar las escuelas diurnas y colegios tradicionales de Costa Rica**



Fuente: Elaboración propia con base en ProDUS,

Con las fuentes de información antes indicadas fue posible georreferenciar la totalidad de colegios existentes en el año 2009, y el 88,9% de las escuelas (cuadro 9). El porcentaje de escuelas no georreferenciadas está constituido en un 95% por escuelas indígenas y centros unidocentes. Finalmente, a través del código secuencial estas bases fueron ligadas con las bases de primaria y secundaria.



**Cuadro 9. Estado de escuelas y colegios georreferenciados del país**

Nivel	Centros georreferenciados	
	Absolutos	Relativo
Escuelas	3.621	88,9
<i>Rango de matrícula:</i>		
Unidocente	1.114	84,1
Dirección 1	1.172	87,7
Dirección 2	664	93,9
Dirección 3	329	94,0
Dirección 4	248	97,6
Dirección 5	81	100,0
Sin Información	13	76,5
<i>Fuera GAM</i>	3.350	87,3
<i>Dentro GAM</i>	721	96,8
Colegios	810	100

Fuente: Probus.

#### *Etapa VI. Verificación de la información*

Una última etapa en este proceso fue la verificación de los procesamientos con respecto a las publicaciones del Departamento de Análisis Estadístico del MEP. Se encontraron varias discrepancias, algunas de ellas originadas por las modificaciones hechas a nivel de código secuencial, como la eliminación de centros con información duplicada; es el caso, por ejemplo, de los datos sobre repitencia en el centro educativo Terraba en 2008. Las diferencias halladas se presentan en los cuadros 10 y 11, y ya fueron expuestas al citado Departamento, para su respectiva valoración.

**Cuadro 10. Diferencias encontradas entre la megabase de educación primaria y las publicaciones del MEP**

<b>Temática y año</b>	<b>Megabase</b>	<b>Publicación</b>
Matrícula		
2006	521.488	521.505
Repitencia		
2001	37.566	37.570
Estudiantes con adecuación:		
<i>Significativa</i>		
2004	6.787	6.786
2007	9.265	9.271
2008	9.490	9.486
<i>No significativa</i>		
2000	45.922	45.979
2002	57.922	57.920
2003	63.145	63.142
2004	66.065	66.043
2007	72.846	72.835
2008	72.094	72.027
<i>Acceso</i>		
2002	8.008	8.003
2007	8.000	8.028
2008	8.118	8.116
Segundo idioma:		
<i>Italiano</i>		
2004	957	1.602
<i>Inglés (radio interactiva)</i>		
2006	4.054	4.044

**Cuadro 11. Diferencias encontradas entre la megabase de educación secundaria y las publicaciones del MEP**

<b>Temática y año</b>	<b>Megabase</b>	<b>Publicación</b>
Repitencia		
2004	28.676	28.677
2008	34.774	34.800
Estudiantes con adecuación:		
<i>Significativa</i>		
2008	2.162	2.158
<i>No significativa</i>		
2003	23.722	23.715
2007	37.264	37.459
2008	35.540	35.374
Acceso		
2000	1.232	1.206
2007	2.002	2.006
2008	1.844	1.843
Personal		
<i>Personal total</i>		
2006	28.997	28.988
2008	31.063	30.794
<i>Docentes administrativos</i>		
2006	3.021	3.020
<i>Docentes total</i>		
2006	21.747	21.739
2008	22.684	22.685
<i>Docentes titulados</i>		
2006	20.209	20.201
2007	21.084	21.085
2009	22.426	22.425

## Estructura de las megabases

Como se mencionó, la estructura de las megabases está organizada de tal modo que la información de un centro educativo se encuentra en una única fila, en tanto que la información general (características) del mismo y las variables de seguimiento se ubican en columnas.

Dentro de las columnas, la “información general” incluye datos como las coordenadas geográficas, el código secuencial, el nombre del centro educativo, el código distrital y el tipo de dependencia (pública, privada o privada subvencionada), entre otros. Por su parte, el conjunto de variables de seguimiento contiene información sobre diez

temáticas: matrícula, repitencia, rendimiento definitivo, deserción, personal, adecuaciones curriculares, alumnos extranjeros, embarazos, segundo idioma (solo primaria), violencia, resultados en pruebas de bachillerato (solo secundaria). El detalle para cada una de las bases se expone a continuación.

Debido a esta estructura particular, el usuario que desee realizar agregaciones geográficas con la información contenida en estas megabases deberá sumar la información en términos absolutos de los centros educativos de acuerdo al tipo de agregación de interés (por ejemplo: distrito, cantón, subregión) y proceder a recalcular los indicadores respectivos.

### *Información y características generales del centro educativo*

En esta categoría se encuentran las variables de ubicación geográfica del centro educativo. Se utilizan coordenadas CRTM05 para la base de primaria y la LAMBERT NORTE para secundaria. Además se construyó un código distrital a partir del número de provincia, número de cantón y número de distrito. Así, el primer dígito corresponde a la provincia, los dos segundos al cantón y los dos últimos al distrito (cuadro 11).

**Cuadro 11. Ejemplo de código distrital**

Número de provincia	Número de cantón	Número de distrito	=	Código del distrito de San Vito, Coto Brus
6	08	01		60801

No obstante, un centro educativo puede cambiar de ubicación geográfica por diferentes motivos, entre ellos el comercial (centros privados que trasladan sus sedes con la finalidad de tener una matrícula mayor) y los derivados de fenómenos naturales (Escuela de Cinchona), o bien por la creación de nuevos distritos, como en el caso de Los Guido de Desamparados (2003).

Además de las variables de información geográfica hay variables de índole administrativa y relativa a las características de la oferta que se imparte en cada centro.

### **Código secuencial y código presupuestario**

Se usa el código secuencial “codins”, desarrollado por el Departamento de Análisis Estadístico del MEP, como la variable de enlace entre las diferentes bases de ese departamento. Sin embargo, se mantiene el código presupuestario, el cual resulta útil únicamente para el caso de las dependencias públicas, con la finalidad de ampliar la base con información de otros departamentos que utilizan ese código.

### **Nombre del centro educativo**

Se refiere al nombre de la escuela o colegio, y puede variar de un año a otro. Sin embargo, se mantiene la información del último año registrado (previa verificación).

### **Año de creación**

Especifica el año en que se creó la institución. En esta variable se presentan diferencias (años distintos) que pueden deberse a errores humanos a la hora de llenar los formularios. En estos casos se mantiene el dato de mayor antigüedad. El criterio para hacerlo así es que puede haberse realizado una reapertura de código en un centro educativo que ya operaba en años anteriores.

### **Dependencia**

Indica si el centro educativo es público, privado o privado subvencionado. En la megabase de primaria se encuentra una variable de dependencia por año para el período en estudio, ya que hay centros educativos que cambian de categoría. Entre ellos existen centros privados que pasan a formar parte de las instituciones subvencionadas por el Estado.

### **Tipo de jornada (exclusivo de primaria)**

Informa si la jornada de trabajo del centro es única, doble o triple. Solo está disponible para los años 2005, 2006, 2008 y 2009.

### **Imparte el servicio de preescolar (exclusivo de primaria)**

Variable que indica si el centro de enseñanza además brinda el servicio de preescolar. Disponible en la megabase para los años 2000, 2001, 2003, 2006, 2008 y 2009.

### **Servicios de educación especial (exclusivo de primaria)**

Indica si el centro educativo ofrece el servicio de apoyo o aula recurso.

### **Pertenencia a programas (exclusivo de primaria)**

Variables que asocian al centro educativo con programas existentes a nivel de primaria. Los programas son:

- Programa de atención prioritaria
- Escuelas líderes
- Escuelas indígenas
- Programa de horario ampliado
- Escuelas de excelencia

### **Rama académica (exclusivo de secundaria)**

Se refiere a la rama a la cual pertenece el centro educativo: académica diurna, académica nocturna, técnica diurna, técnica nocturna, artística.

Se incluye una variable de rama para cada año, dado que algunos colegios pueden cambiar de modalidad, principalmente los académicos que han pasado a ser técnicos, como los colegios técnicos profesionales de Granilla y el INVU Las Cañas.

Aquí se encuentra el caso especial, antes comentado, del Colegio Técnico Profesional San Isidro, que durante 2001 y 2002 tuvo matrícula académica en el tercer ciclo y

matrícula técnica en la educación diversificada. Por tal razón, se observará que para esos años se recodificó la variable, como “código 41”. Se recomienda a los usuarios de la base que, en caso de calcular la matrícula por rama de actividad para los años en cuestión, procedan a sumar la matrícula de séptimo, octavo y noveno dentro de la rama académica diurna y la de décimo, undécimo y duodécimo dentro de la rama técnica.

### **Es telesecundaria (exclusivo de secundaria)**

Indica si el centro educativo es una telesecundaria o no.

### **Estructura curricular (exclusivo de secundaria académica pública)**

Corresponde al tipo de estructura curricular que tienen los colegios académicos públicos; por ejemplo si son científicos, humanistas, deportivos o bilingües, entre otros.

### *Variables de seguimiento*

#### **Matrícula inicial total**

Número total de estudiantes inscritos en los centros de enseñanza diurnos públicos, privados y privados subvencionados, al iniciarse el año lectivo. En primaria este dato excluye la matrícula de preescolar y aula integrada. La variable se encuentra desagregada por año cursado y para el sexo masculino (el dato de mujeres puede calcularse como la diferencia entre el dato total y el dato de hombres)

#### **Repitentes**

Los estudiantes repitentes en primaria son aquellos matriculados en el mismo año (grado) que cursaron el año anterior, o en el último año cursado en que estuvieron matriculados. La variable está desagregada por año cursado y para el sexo masculino. En cuanto al porcentaje de repitencia, es la proporción que representan los estudiantes que repiten con respecto al total de estudiantes matriculados al inicio del ciclo lectivo. En secundaria los registros administrativos solo incluyen centros diurnos en el período 2001-2008.

#### **Porcentaje de repitencia**

Según el artículo 36 del Reglamento de Evaluación de los Aprendizajes, modificado mediante el decreto 34886-MEP, a partir del año 2009 se establece que: “El estudiante de III Ciclo de la Educación General Básica o del Ciclo de Educación General Diversificada que, una vez realizadas las pruebas de ampliación, hubiese reprobado de forma definitiva algunas de las asignaturas que cursaba, se considerará reprobado en el nivel escolar que cursaba. Esto lo ubica en la categoría de estudiante repitente”.

#### **Deserción intra-anual**

La deserción intra-anual es la diferencia entre matrícula inicial y matrícula final. La matrícula final toma en cuenta a los estudiantes nuevos, es decir, aquellos que se incorporaron luego de la fecha en que se capturó el dato de matrícula inicial, a los alumnos que se trasladaron de un centro educativo a otro, y a los que abandonaron la escuela o colegio. Con este dato se calcula el porcentaje de deserción intra-anual, el

cual está recomendado para agregaciones geográficas pero no para centros educativos.

### **Abandono**

Esta variable registra la cantidad de alumnos que abandonaron el sistema educativo. Con este dato se calculó el porcentaje de abandono a nivel de centro de enseñanza, dado que es la cifra oficial de deserción que reportan los directores y que consta en los formularios de matrícula final entregados al MEP. En agregaciones geográficas grandes se tiende a subestimar la deserción, ya que muchos estudiantes que se registran como traslados no llegan al centro educativo de destino. La información está desagregada según grado académico y sexo. Disponible a partir del 2001.

### **Porcentaje de abandono**

Es la proporción que representan los estudiantes que abandonan el centro educativo con respecto a la matrícula inicial del mismo año. El Departamento de Análisis Estadístico del MEP recoge información sobre matrícula inicial de los respectivos formularios. Para calcular el porcentaje se utilizan los datos de matrícula inicial que se encuentran en las bases de deserción, ya que no todos los centros que reportan información a inicios de año lo hacen al finalizar este.

### **Rendimiento**

Para esta versión de la megabase se optó por el dato de rendimiento definitivo, que incluye los resultados de las distintas convocatorias. Estos resultados derivan en estudiantes aprobados y reprobados.

### **Aprobación**

Es el número de estudiantes que, al finalizar el ciclo lectivo, logran aprobar el año que cursaron. La cifra incluye los resultados de las convocatorias. El porcentaje de aprobación es la proporción que representan los estudiantes que aprobaron, con respecto a la matrícula final. En secundaria los registros administrativos solo incluyen centros diurnos en el período 2000-2008.

### **Reprobación**

Es el número de estudiantes que al finalizar el ciclo lectivo no logran aprobar el año que cursaron, la cifra incluye los resultados de las convocatorias. El porcentaje de reprobación es la proporción que representan los estudiantes que reprobaron, con respecto a la matrícula final. En secundaria los registros administrativos solo incluyen centros diurnos en el período 2000-2008.

### ***Matrícula inicial de estudiantes extranjeros***

Se refiere a la cantidad de estudiantes que se encuentran matriculados en el sistema educativo y que son extranjeros. Se encuentra desagregada por sexo.

### *Segundo idioma (exclusivo para primaria)*

Número de estudiantes matriculados que reciben lecciones de un segundo idioma, ya sea inglés (modalidad presencial y radio interactiva), francés o italiano. Para calcular el porcentaje de cobertura a nivel institucional se divide el número de alumnos que reciben un segundo idioma entre la matrícula inicial que reporta el centro educativo. Las publicaciones del Departamento de Análisis Estadístico sólo incluye la cobertura de segundo idioma para centros de primaria públicos. En el caso de secundaria según el mismo departamento se ofrece como mínimo un segundo idioma. Es importante indicar que, para realizar agregaciones geográficas, es necesario sumar las cifras absolutas, tanto para segundo idioma como para matrícula, y proceder a recalcularse ese porcentaje.

### *Alumnas embarazadas en educación tradicional*

Número de mujeres embarazadas matriculadas en el sistema educativo. Los registros que se incluyen en la base inician en el año 2002.

### *Casos de violencia en los centros educativos*

Estadísticas sobre el número de estudiantes expulsados del sistema educativo de manera temporal o definitiva. Las estadísticas que se reportan inician en el 2002.

### *Adecuaciones curriculares en educación tradicional*

El Departamento de Educación Especial del MEP define así las adecuaciones curriculares:

“Adecuación Curricular: Es el acomodo o ajuste de la oferta educativa a las características y necesidades de cada alumno, con el fin de atender las diferencias individuales de estos.

Adecuaciones de Acceso: Son las modificaciones o provisión de recursos especiales, materiales o de comunicación, dirigidas a algunos estudiantes (en especial aquellos con deficiencias motoras, visuales y auditivas) para facilitarles el acceso al currículo regular o, en su caso, al currículo adaptado.

Adecuaciones No Significativas: Son aquellas que no modifican sustancialmente la programación del currículo oficial. Constituyen las acciones que los docentes realizan para ofrecer situaciones de aprendizaje adecuadas, con el fin de atender las necesidades educativas de los estudiantes. Estas acciones incluyen la priorización de objetivos y contenidos, así como ajustes metodológicos y evaluativos de acuerdo con las necesidades, características e intereses de los educandos.

Adecuaciones Significativas: Consisten principalmente en la eliminación de contenidos esenciales y objetivos generales que se consideran básicos en las diferentes asignaturas y la consiguiente modificación de los criterios de evaluación. La aplicación de este tipo de adecuaciones requiere de un análisis exhaustivo, ya que no se trata de simples adaptaciones en la metodología o en la evaluación, sino que representan modificaciones sustanciales del currículo oficial” (Departamento de Educación Especial, 2009).



## *Infraestructura educativa*

El Departamento de Análisis Estadístico registra un conjunto de variables sobre infraestructura educativa, como las mencionadas en el cuadro 2. Es importante destacar que las categorías que aluden al estado esa infraestructura -bueno, regular o malo- se derivan de la percepción de las y los directores de los centros educativos. En el 2009, el formulario de matrícula inicial y personal empezó a solicitar información sobre el estado de espacios físicos destinados a aulas académicas, aulas para asignaturas especiales, comedor, laboratorio de Informática, taller de Artes Industriales (anteriormente solo “taller”) y sala de profesores. En el caso de las bibliotecas la categorización se realiza a partir del formulario de 2008.

## *Personal*

La información relativa al personal se incluye desde el 2005, ya que es a partir de ese año que se encuentran datos más homogéneos entre sí. Los funcionarios de cada centro educativo se clasifican según el tipo de cargo, en personal docente-administrativo, personal docente y personal administrativo y de servicios. Cabe indicar que las cifras que publica el Departamento de Análisis Estadístico no corresponden a personas físicas, sino a plazas. Es decir, si un educador imparte lecciones en dos colegios distintos, se contabiliza como parte del personal en ambas instituciones. Por otro lado, si un docente imparte dos materias en un mismo colegio, debe ser considerado como profesor de ambas asignaturas (dos plazas). La información comprende al personal que está laborando a la fecha indicada por el Departamento, incluyendo a aquellos que tienen una incapacidad o permiso, o se encuentran ausentes por otros motivos pero no han sido sustituidos en ese momento. Para 2009 la fecha definida por el MEP fue el 13 de marzo.

### **Personal docente-administrativo**

Incluye puestos como director, asistente de Dirección, bibliotecario asistente y una categoría de “Otros”. De acuerdo con el formulario, el director de Enseñanza General Básica se encuentra dentro de esta categoría.

### **Personal docente**

Corresponde a los docentes que se desempeñan en los siguientes puestos: Profesor de Enseñanza General Básica (PEGB), Unidocentes y Asignaturas Especiales. En las escuelas unidocentes el director está incluido como unidocente y el docente de apoyo como docente del grupo PEGB.

De acuerdo con el grupo profesional el personal docente se clasifica en:

### **Personal docente titulado**

“Son profesores titulados los que, de conformidad con esta ley, posean un grado o título profesional que los acredite para el ejercicio docente, extendido por las instituciones oficiales del país, o reconocido y equiparado por la Universidad de Costa Rica o por el Consejo Superior de Educación, según corresponda. Estos organismos

tendrán, además, la obligación de especificar las especialidades que, por sus estudios, puedan impartir los profesionales comprendidos en esta ley” (artículo 108, Ley de Carrera Docente).

En primaria incluye a los grupos PT1, PT2, PT3, PT4, PT5, PT6, VT1, VT2, VT3, VT4, VT5 y VT6, además de otros docentes titulados. En secundaria agrupa a los MT1, MT2, MT3, MT4, MT5, MT6, VT1, VT2, VT3, VT4, VT5 y VT6, además de otros docentes titulados.

### **Personal docente autorizado**

“Son profesores autorizados los que, sin poseer título o grado específico para el cargo que desempeñan, ostenten otros que sean afines, según será determinado, para cada caso, en la presente ley” (artículo 109, Ley de Carrera Docente).

En primaria incluye a los grupos PAU1, PAU2, VAU1 y VAU2, y en secundaria a las categorías MAU1, MAU2, VAU1 y VAU2.

### **Personal docente aspirante**

“Son profesores aspirantes los que, por sus estudios y experiencias, no pueden ser ubicados en ninguno de los dos grupos anteriores” (artículo 110, Ley de Carrera Docente).

### **Porcentaje de docentes titulados**

Es la proporción que representan los docentes titulados con respecto al total de puestos docentes.

### **Porcentaje de docentes no titulados**

Es la proporción que representan los docentes no titulados (autorizados y aspirantes) con respecto al total de puestos docentes.

### **Personal administrativo y de servicios**

Incluye los puestos de oficinista, agente de seguridad y vigilancia, misceláneo, cocinera, trabajador especializado, equipo interdisciplinario y otros.

### ***Bachillerato (solo secundaria)***

Entre las variables incluidas se encuentran el número de estudiantes que presentaron las pruebas de bachillerato, el número de estudiantes que las aprobaron y el porcentaje de promoción, que se define como la proporción que representan los alumnos aprobados con respecto al total que presentó la prueba.