

EL SISTEMA DE CONSULTA DEL FORMATO 911 EN EDUCACIÓN SUPERIOR: UNA PROPUESTA PARA EXPLOTAR LAS BASES DE DATOS

JUAN CARLOS ORTEGA GUERRERO

Instituto de Investigaciones en Educación, Universidad Veracruzana

GERMÁN ÁLVAREZ MENDIOLA

Departamento de Investigaciones Educativas, CINVESTAV-IPN

OLIVA ADELA ROSALES RODRÍGUEZ

Instituto de Investigaciones en Educación, Universidad Veracruzana

RESUMEN: En este trabajo hacemos una reflexión sobre la información disponible de las instituciones de educación superior (IES) en México. Se analiza el caso de la Estadística 911, instrumento utilizado por la Secretaría de Educación Pública para recabar datos sobre las IES, y se revisan los cambios que ha sufrido de 1997 a 2009. Con base en esto, se presenta un sistema de cómputo que une, normaliza y agrupa los cuatro formatos de la Estadística 911 sobre Educación Superior, desarrollado por el primer autor y colaboradores, en el marco de su tesis de doctorado en investigación educativa en la Universidad Veracruzana, y del proyecto de investigación “El papel del sector privado en las configuraciones sistémicas estatales de la educación superior privada en México” que cuenta

con financiamiento del CONACYT, coordinado por el segundo autor. El sistema desarrollado permite analizar la información mediante gráficas, cuadros estadísticos y cubos de información.

En la primera sección se describe brevemente el proceso de desarrollo de la Estadística 911; en la segunda parte se describe su estructura actual y los datos que se recaban por medio de ella; en la tercera parte se muestran las características de la interfaz del sistema desarrollado y finalmente en la última parte se discute sobre resultados obtenidos y trabajo a futuro.

PALABRAS CLAVE: Sistemas de Información, Educación Superior, ANUIES, Innovaciones Tecnológicas, Política Educativa.

Introducción

El sistema de educación superior en México se caracteriza por ser complejo y heterogéneo. Se integra por un gran número de instituciones de educación superior que se encuentran distribuidas por todo el territorio nacional. Las IES tienen diferentes fuentes de financiamiento, estructuras académicas, formas de organización, constitución y modalida-

des educativas. Lo anterior dificulta obtener información sobre el tamaño, infraestructura física y humana, oferta educativa, matrícula y eficiencia terminal de las IES a nivel nacional. El desarrollo de las IES requiere tener información confiable, oportuna y de fácil acceso; la demanda creciente de servicios y productos estadísticos y de información cualitativa sobre el sistema de educación superior (SES) rebasa las posibilidades de los mecanismos actuales, evidenciando la falta de un verdadero Sistema Nacional de Información sobre la educación superior (ANUIES, 2000).

A pesar de todos los esfuerzos y planes de organismos gubernamentales (SEP, 2007) no se tiene un sistema que concentre, sistematice y ponga a disposición pública la información. Actualmente hay varios proyectos encaminados a lograr este fin. La ANUIES está trabajando en una propuesta que ordena la información de la Estadística 911; por un lado, crea catálogos únicos de instituciones y, por otro, ofrecerá un conjunto de consultas tanto fijas como dinámicas que permitirán tener la información fácilmente. Otro esfuerzo por sistematizar información estadística es el que lleva a cabo la Dirección General de Evaluación Institucional (DGEI) de la UNAM, en el proyecto Estudio Comparativo de las Universidades Mexicanas (ECUM). Este proyecto ha desarrollado una herramienta de exploración de datos en línea denominado EXECUM. El proyecto tiene como objetivo principal aportar información comparable sobre las características y funciones de las universidades y de otras instituciones mexicanas de educación superior, en el marco general, tanto nacional como internacional, de producción de *rankings*. Utiliza diversas fuentes oficiales de información, como el Formato 911 para los temas de personal docente, matrícula y programas académicos (<http://www.ecum.unam.mx/>).

Ahora bien, dado que es deseable que esta información estadística se convierta en un insumo útil para el desarrollo institucional, será necesario que investigaciones futuras diluciden las causas políticas, organizacionales y técnicas que impiden tener un sistema de información deseable. Con ese conocimiento, se tendrán bases para ofrecer una propuesta de sistematización de la información que produzca datos en tiempo y forma, útiles para investigadores y tomadores de decisión. .

En la primera etapa del proyecto desarrollado para ayudar a solucionar la problemática arriba expuesta, se han depurado, normalizado y reunido en un sistema de cómputo los formatos de la Estadística 911 que dan cuenta de la información de las IES.

1. De dónde viene la Estadística 911

1.1 Historia

Hasta 1975 tres instituciones recolectaban y procesaban la información sobre las IES a nivel nacional de manera independiente: la Dirección General de Planeación Educativa (DGPE) dependiente de la SEP, la Dirección General de Estadística (DGE, posteriormente INEGI) de la Secretaría de Industria y Comercio y la Secretaría General de la Asociación Nacional de Universidades e Instituciones de Educación Superior (ANUIES). La descoordinación entre estas entidades generaba trabajo redundante e inconsistencia en la información reportada por las IES.

Para solucionar este problema la ANUIES y la SEP convocan en 1975 al Primer Seminario de Información y Estadística del cual emerge un cuestionario único, el denominado cuestionario estadístico 911, que se aplicó como prueba piloto en Guanajuato durante el ciclo escolar 1975-1976.

En 1980, la XVIII Reunión Ordinaria de la Asamblea General de la ANUIES propuso el Plan Nacional de Educación Superior en el que uno de sus programas consiste en la implementación del Sistema Nacional de Información de la Educación Superior, con los objetivos de crear o consolidar unidades de información en las IES e integrar un sistema nacional de datos sobre educación superior que conjunte dichos datos a nivel estatal y federal.

En 1984 se actualizó el cuestionario estadístico de educación superior SEP-ANUIES-SPP 911 para captar información sobre cursos de especialización, maestría y doctorado. En 1990 la 4ª Reunión Nacional de Información Estadística y dentro del Programa Nacional de Modernización Educativa planteó la necesidad de contar con datos estadísticos válidos y oportunos que sustentaran la planeación del sistema de educación superior. Para esto propuso mejorar la respuesta al cuestionario 911, ampliar la cobertura temática del mismo y fomentar el intercambio de ideas entre los responsables de la información estadística en las IES. Además propuso ampliar el cuestionario para captar información de los sistemas de enseñanza abierta.

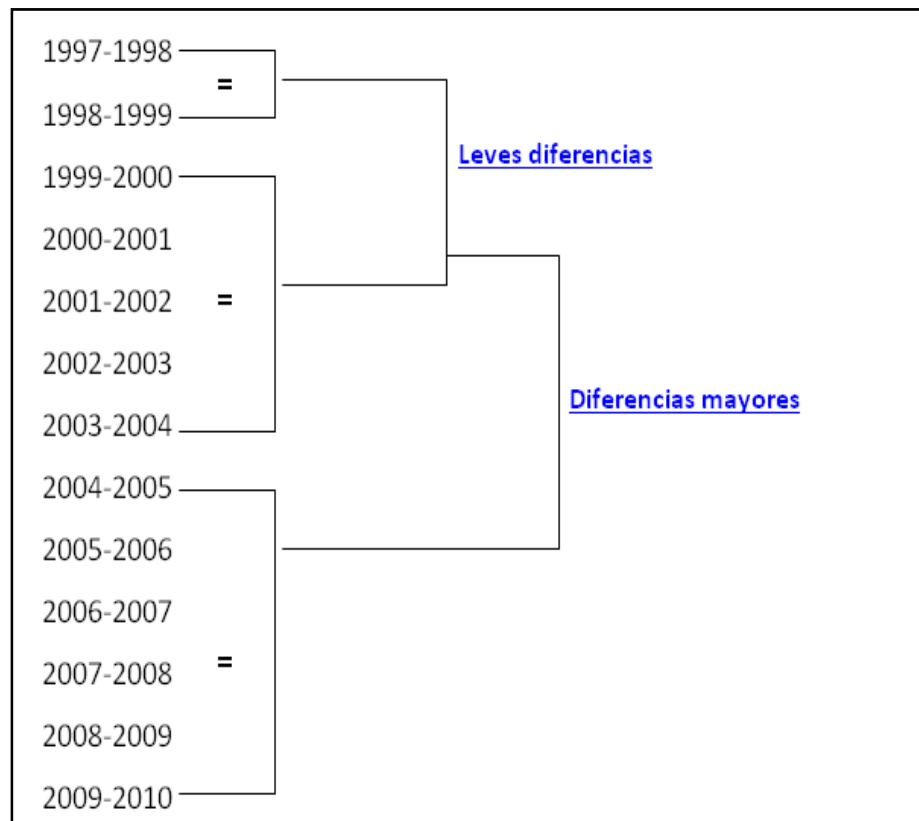
En 1994 el formato incluyó indicadores usados en los países integrantes de la Organización para la Cooperación y Desarrollo Económicos (OCDE).

No es sino hasta 1995, durante la 5ª Reunión Nacional de Información Estadística, que se propone captar la información por medio de un cuestionario electrónico en disquetes que contenía la misma información que en el medio impreso. Así mismo, se insistió en la necesidad de promover y consolidar un sistema único de información que cubriera todas las peticiones de los diferentes usuarios. Con el fin de conocer la demanda se consideró la posibilidad de incluir preguntas sobre el número de solicitantes a ingresar, no solo los inscritos como se venía preguntando hasta ese momento (RÍOS H., 2001).

En *La Educación Superior Hacia el Siglo XXI* (ANUIES, 2000) la ANUIES propuso que para 2006 "...las IES contarán con un sistema integral de información que permita la comparación entre indicadores institucionales e información de interés común" objetivo que parece aún lejano.

En la siguiente figura se muestran a grandes rasgos los cambios que ha sufrido el formato 911 de 1997 a la fecha.

Figura 1. Variaciones de la Estadística 911 en el tiempo



1.2 Estructura actual de la 911

El Cuestionario Estadístico de Educación Superior, Estadística 911, recaba estadísticas de todos los niveles educativos: básica, media y superior. Es elaborado conjuntamente por la SEP, la ANUIES y la SHCP (antes SPP).

Por medio de este cuestionario se recopila información de muy variados ámbitos: fuente de ingresos; número de alumnos por área académica, género, grado escolar, fecha de ingreso, egresados y titulados; número de escuelas, facultades, posgrados, institutos y centros de investigación; personal académico, directivo, administrativo, de intendencia; escolaridad del personal, características de los inmuebles, número y tipo de carreras ofertadas, modalidades educativas manejadas. Todo lo anterior a diferentes niveles de agregación. Para la Educación Superior maneja los siguientes formatos:

- 911.10 – Estadísticas de Educación Superior por Institución. Módulos de docencia, investigación, extensión y difusión, administración
- 911.9 - Estadísticas de Educación Superior por Escuela. Módulo de docencia
- 911.9A - Estadísticas de Educación Superior por Carrera. Módulo de docencia
- 911.9B - Estadísticas de Educación Superior por Posgrado. Módulo de docencia

El método para recabar los datos del cuestionario 911 sigue siendo básicamente el formato impreso aunque las IES entregan también a la SEP un archivo electrónico.

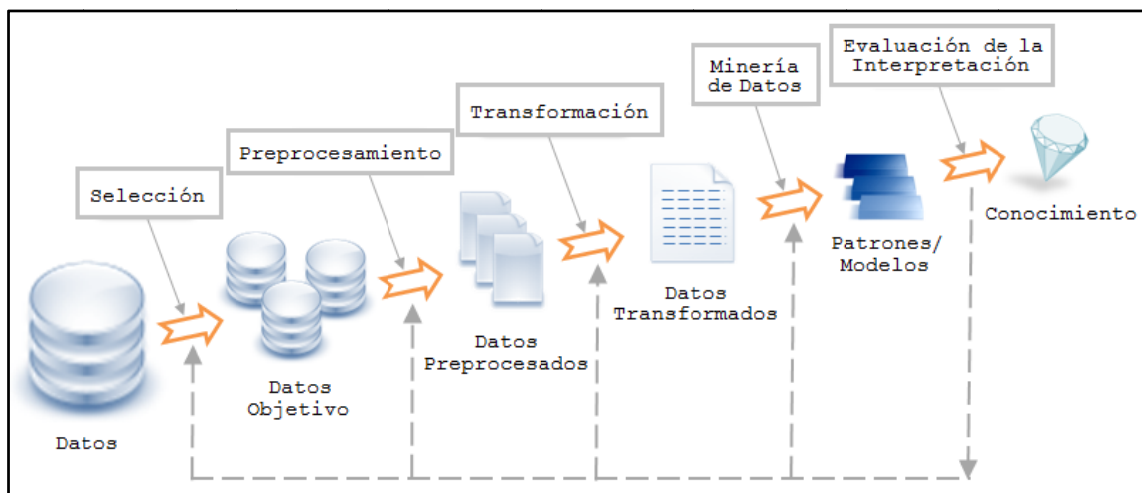
La información recolectada a través de cuestionario 911 se usa ampliamente pero generalmente de manera fraccionada, sin articulación entre sus diferentes secciones.

Este cuestionario constituye la base empírica sobre la que se centra este proyecto de investigación, debido a que alrededor de él se han concretado los esfuerzos de recolección de información de las IES en México.

2. Proceso de integración

Para integrar la información se tomó como base el enfoque de minería de datos para la adquisición del conocimiento o *KDD (Knowledge Discovery Data)* que propone varias etapas para que a partir de datos se obtenga información con la que se genera nuevo conocimiento.

Figura 2. Etapas del proceso KDD



Se procesaron e integraron los archivos de los años 1997-1998 a 2009-2010 proporcionados por la SEP en base de datos de tipo DBF. En las tablas siguientes se resumen los volúmenes de información integrada y depurada.

Tabla 1. Registros integrados de 911.10 Instituciones

Año	Registros en base de datos original	Registros con errores	Registros integrados
1997	1,067	0	1,067
1998	1,166	0	1,166
1999	1,315	0	1,315
2000	1,416	0	1,416
2001	1,505	0	1,505
2002	1,663	0	1,663
2003	1,765	0	1,765
2004	1,774	0	1,774
2005	2,265	0	2,265
2006	2,394	0	2,394
2007	2,479	0	2,479
2008	2,573	0	2,573
2009	2,822	0	2,822
Totales	24,204	0	24,204

Tabla 2. Registros integrados de 911.9 Escuelas

Años	Registros en base de datos original	Registros con errores	Registros integrados
1997	2,520	0	2,520
1998	2,657	0	2,657
1999	2,903	0	2,903
2000	3,054	0	3,054
2001	3,176	0	3,176
2002	3,434	0	3,434
2003	3,623	0	3,623
2004	3,475	0	3,475
2005	4,091	0	4,091
2006	4,303	0	4,303
2007	4,462	0	4,462
2008	4,582	0	4,582
2009	4,984	0	4,984
Totales	47,264	0	47,264

Tabla 3. Registros integrados de 911.9A Carreras

Años	Registros en base de datos original	Registros con errores	Registros integrados
1997	6,767	0	6,767
1998	7,609	0	7,609
1999	8,182	0	8,182
2000	8,955	0	8,955
2001	9,583	0	9,583
2002	10,746	0	10,746
2003	11,519	5	11,514
2004	12,301	52	12,249

2005	14,442	132	14,310
2006	15,958	171	15,787
2007	16,950	192	16,758
2008	17,940	216	17,724
2009	20,478	386	20,092
Totales	161,430	1,154	160,276

Tabla 4. Registros integrados de 911.9B Posgrados

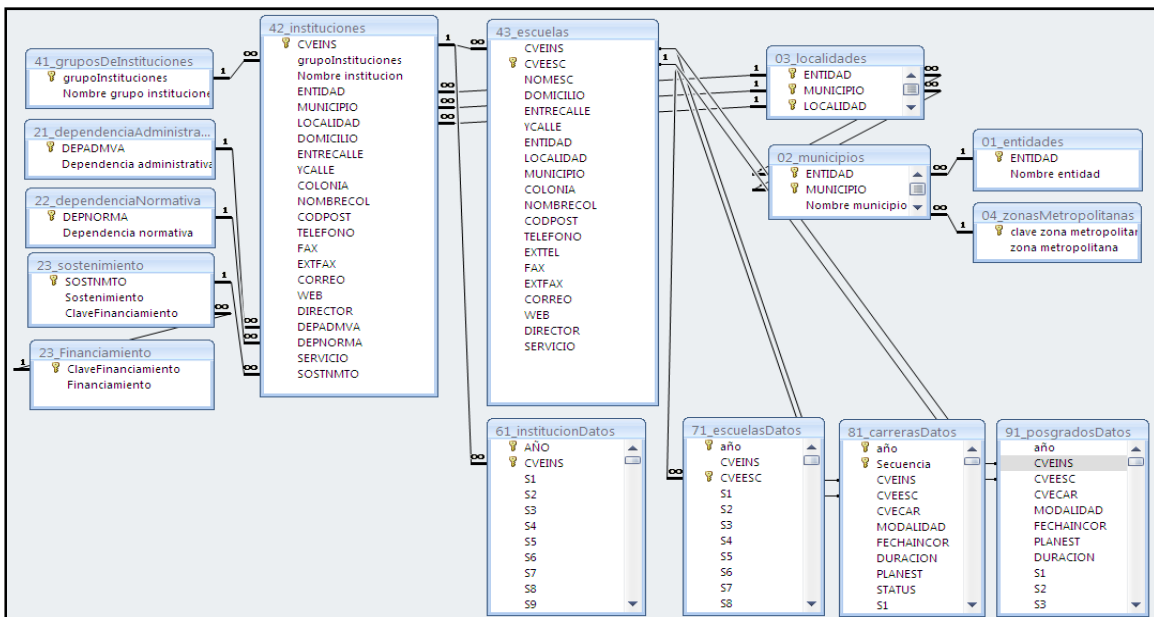
Años	Registros en base de datos original	Registros con errores	Registros integrados
1997	3,334	4	3,330
1998	3,404	0	3,404
1999	3,641	0	3,641
2000	4,057	0	4,057
2001	4,527	0	4,527
2002	4,947	0	4,947
2003	5,350	0	5,350
2004	5,785	32	5,753
2005	6,086	8	6,078
2006	6,647	13	6,634
2007	6,375	18	6,357
2008	7,109	16	7,093
2009	7,722	28	7,694
Totales	68,984	119	68,865

Como se puede apreciar en general la información es consistente, sin embargo hay pequeñas anomalías y errores a partir de los años 2003 y 2004 en los formatos de Carreras y Posgrados. Estos errores tienden a incrementarse en estos formatos debido principalmente a programas educativos reportados doble o triple.

En el proceso de integración de cada base de datos se llevaron a cabo los siguientes pasos, que corresponden a las etapas de selección, preprocesamiento y transformación del enfoque KDD:

- Revisión del archivo para verificar su formato. Debido a las diferentes versiones de las bases de datos y cambios en el número y nombre de variables.
- Depuración de los registros. Se eliminaron registros que presentaban información duplicada y aquellos sin información.
- Redefinición de campos y tipo de datos. A lo largo de los años ha habido cambios en los formatos, principalmente cambiaron los nombres de variables y su tamaño. Se hizo corresponder el nombre y el tamaño de las variables en todos los años. En algunos casos se agregaron o eliminaron.
- Verificación relacional. Para resolver los problemas de redundancia e inconsistencia entre las tablas generadas a partir de las bases de datos de la SEP, se diseñó un modelo de entidad-relación (ver figura 3).
- Adición de la información al sistema. Con la información ya verificada y depurada se generó un sistema de bases de datos relacional (BDR) que asegurara la consistencia de la información contenida en el sistema. Para esto se crearon, entre otros, catálogos de entidades, municipios, localidades, zonas metropolitanas, dependencias administrativas, dependencias normativas, tipos de posgrado, instituciones.

Figura 3. Diagrama entidad–relación de la base de datos.



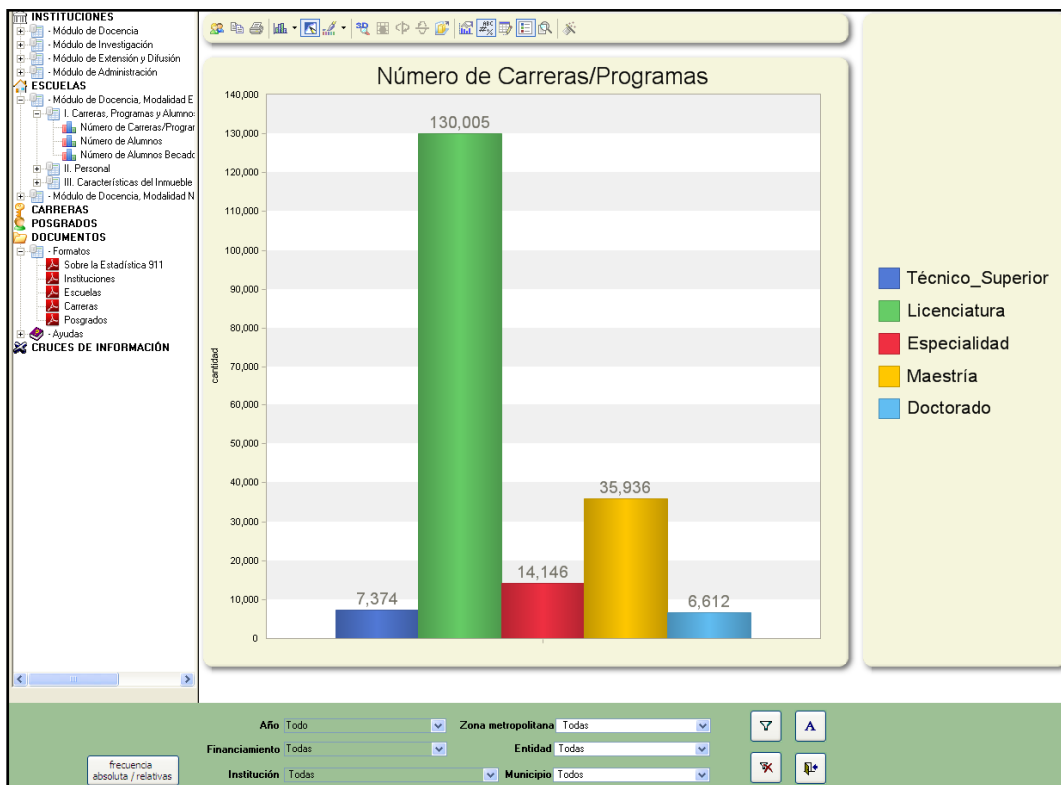
3. Características e interfaz del sistema

El sistema funciona en un entorno cliente-servidor compatible con sistema operativo Windows, permite hacer conteos simples y gráficas de cada una de las variables de la Estadística 911; elaborar cubos de información y exportar todo a diferentes formatos, tanto gráficos como de hojas de cálculo.

Debido a lo amplio y complejo de los formatos de la Estadística 911, alrededor de 500 variables en el caso de los formatos de instituciones y escuelas, el sistema incluye una sección que muestra todos los formatos con el nombre de cada variable, un documento sobre la integración de la información y un amplio sistema de ayudas.

En los conteos simples es posible seleccionar la información por año, financiamiento, institución, zona metropolitana, entidad y municipio. El resultado se muestra en forma de gráfica (Figura 4), pudiéndose escoger entre diferentes tipos de gráfica: pastel, barras, líneas, etcétera. Los usuarios pueden ver las tablas de datos subyacentes y exportar tanto datos como gráficas a otros programas como Word, Excel, Power Point o SPSS.

Figura 4. Conteos simples. Número de carreras/programas por grado académico



Los cubos de información OLAP (OnLine Analytical Processing) (Cood, 1970) permiten hacer cruces entre variables y seleccionar la información mostrada por cualquier criterio. Es posible hacer conteos de las variables o presentarlas en porcentajes por línea o columna. Se pueden generar variables que suman varios datos y obtener diferentes niveles de detalle. También es factible representar los resultados en forma de gráfica y exportar a Excel.

Figura 5. Cubo de información. Número de IES por Entidad, año y sector

Nombre entidad	ANO - Financiamiento				ANO - Financiamiento				ANO - Financiamiento			
	1997				1998				1999			
	Desconocido	PRIVADO	PUBLICO	Total	Desconocido	PRIVADO	PUBLICO	Total	Desconocido	PRIVADO	PUBLICO	Total
AGUASCALIENTES		3	6	9			4	6	10			5
BAJA CALIFORNIA		22	16	38			16	17	33			21
BAJA CALIFORNIA SUR		2	8	10			2	8	10			2
CAMPECHE	1	2	9	12	1		2	10	13	1		4
COAHUILA		26	16	42			27	19	46			27
COLIMA		3	4	7			3	4	7			4
CHIASPAS		24	11	35			29	12	41	1		32
CHIHUAHUA		17	18	35			18	18	36			22
DISTRITO FEDERAL	3	113	27	143	3	113	27	143	153	3		153
DURANGO		12	11	23			12	11	23			12
GUANAJUATO		30	13	43			37	16	53			47
GUERRERO		9	15	24			10	16	26			11
HIDALGO		14	10	24			14	10	24			14
JALISCO	1	28	13	42	1		29	16	46	1		34
IMEX	3	35	20	58	3		48	29	80	4		60
MICHOACAN		9	15	24			9	15	24			11
MORELOS		17	7	24			19	7	26			20
NAYARIT		5	3	8			6	5	11			6
NUEVO LEON		38	8	46			40	10	50			44
OAXACA		5	17	22			10	17	27			13
PUEBLA	1	87	23	111	1		99	25	125	1		108
QUERETARO	1	5	8	14	1		6	10	17	1		10
QUINTANA ROO		5	8	13			5	9	14			5
SAN LUIS POTOSI		10	12	22			12	12	24			14
SINALOA	1	19	14	34	1		20	16	37	1		23
SONORA		7	16	23			8	20	28			10
TABASCO		3	9	12			4	10	14			4
TAMAULIPAS		31	15	46			31	16	47			32
TLAXCALA		8	6	14			8	6	14			8
VERACRUZ	1	44	19	64	1		47	21	69	1		53
YUCATAN	1	26	8	35	1		26	9	36	1		30
ZACATECAS		2	8	10			2	10	12			2
Total general	13	661	393	1067	13	716	437	1166	15	841	46	

4. Resultados obtenidos y trabajo a futuro

El sistema descrito facilita la obtención de datos. Actualmente es usado por investigadores en diversos trabajos. Uno de ellos, en la Universidad Veracruzana bajo la coordinación de la Dra. Cristina Miranda Alvarado, analiza el *Desarrollo de las IES privadas en el estado de Veracruz* usando para ello el número de instituciones, escuelas, carreras y posgrados que se ofrecen; la matrícula con que cuentan; el número de estudiantes que egresan; los inmuebles; el personal administrativo y académico que labora en ellas y su producción artística y eventos culturales. Otra investigación sobre la *educación superior privada en México en la primera década del siglo XXI*, del Dr. Germán Álvarez del CINVESTAV, analiza los ritmos de crecimiento de este tipo de instituciones.

La consistencia de los datos contenidos en el sistema se está verificando y se ha confrontado con otras publicaciones, entre ellas los anuarios de la ANUIES y el trabajo de Manuel Gil (2009) sobre la cobertura de la educación superior.

Se está trabajando en una versión en WEB que permita el acceso a cualquier persona interesada en esta información.

Referencias

- ANUIES (2000) La Educación Superior en el Siglo XXI, [en línea]. México. Recuperado el 10 de octubre de 2009, de http://www.anuies.mx/servicios/d_estrategicos/documentos_estrategicos/21/sXXI.pdf
- Codd, E.F. (1970) A Relational Model of Data for Large Shared Data Banks. Communications of the ACM 13 (6): pp. 377–387.
- Gil, M. et al (2009) Cobertura de la educación superior en México. Tendencias, retos y perspectivas. México: ANUIES
- Ríos, H (2001) Las estadísticas de la educación superior, Revista de la Educación Superior en Línea. Num. 118 [en línea]. México. Recuperado el 1 de noviembre de 2009, de www.anuies.mx/servicios/p_anuies/publicaciones/revsup/res118/art9.htm
- SEP (2007) Programa Sectorial de Educación 2007 – 2012, [en línea]. México. Recuperado el 15 de noviembre de 2009, de http://upepe.sep.gob.mx/prog_sec.pdf

Agradecimientos

El tiempo para hacer este trabajo ha sido otorgado por la Universidad Veracruzana en apoyo al primer autor para que desarrolle su investigación en el marco del Doctorado en Investigación Educativa de la misma universidad.

Este trabajo ha contado con el apoyo del CINVESTAV a través del proyecto *El papel del sector privado en las configuraciones sistémicas estatales de la educación superior privada en México: políticas públicas, mercados y diferenciación institucional*, coordinado por el Dr. Germán Álvarez Mendiola y financiado por el CONACYT.