

## Integración de información a almacenes de datos de seguridad ciudadana

### *Integration of Information into Public Safety Data Warehouses*

Iván M. Cárdenas Tandrón

Universidad Central de Las Villas, (UCLV)

[ivanmct@uclv.edu.cu](mailto:ivanmct@uclv.edu.cu),

#### **Resumen**

La seguridad ciudadana constituye uno de los principales retos a que están sometidos los Estados en el mundo actual. La implementación de estrategias, proyectos y acciones dirigidas a factores claves de seguridad ciudadana, constituye una de las formas de contribuir progresivamente a mejorar su comportamiento. Se hace necesario relacionar las informaciones contenidas en las distintas fuentes de información empleando herramientas informáticas que propicien su integración, transformación y almacenamiento. Por tal motivo, es sumamente importante establecer los procesos de Integración de la Información brindada por los diferentes Órganos de Seguridad para garantizar de esta forma una visión integradora que permita solucionar el tema de la Inseguridad en la población y bajar los Índices Delictivos por medio del conocimiento de la realidad en las naciones. La definición o establecimiento de un patrón de integración, es un punto vital para garantizar el éxito de este tipo de problemática. Por ende, se especifican algunas puntos referentes a los Procesos de Integración de Fuentes de Información a Almacenes de Datos de Seguridad Ciudadana, que incluye como fuente motora de la aplicación del Almacén de Datos, la necesidad de ampliación del Dominio de Análisis, la identificación de nuevas fuentes de Información que cubran con el dominio a alcanzar, la clasificación de las fuentes identificadas, y en base a la clasificación de cada fuente, la forma de integración.

**Palabras clave:** almacén de datos, *data warehouse*<sup>1</sup>, etl, fuente, información, integración, seguridad ciudadana.

#### ***Abstract***

*The security of the citizen constitutes one of the main problems of the states of the actual world. The implementation of the strategies, projects, and actions focused to resolve key aspects of the citizen security, constitute a well fundament way to contribute progressively to improve its behavior. It is very important to relate the information from the sources related to each topic, using informatics tools that make the integration possible, including the transformation and the storing. That's why, it is very important to establish the processes of information integration given by the security institutions to guaranty by this way, an overlook vision that allow to resolve the insecurity of the citizen, and to low the criminal indexes, using the knowledge of the reality of the nations. The definition of a pattern of integration is a vital point to ensuring the success of this type of problem. Therefore, this document specifies some points concerning to the process of integration of information sources to Public Safety Data Warehouse, which includes as a principal motor of the implementation of this kind of Data Warehouse, the need of expansion of the Analysis Domain, the identification of new sources of information covering the domain to reach, the classification of identified sources and based on the classification of each source, the form of integration.*

---

<sup>1</sup> Almacén de Datos

**Keywords:** citizen security, data storage, data warehouse, etl, information, integration, public safety, source.

## **Introducción**

La seguridad ciudadana constituye uno de los principales retos a que están sometidos los Estados en el mundo actual. Por una parte, las organizaciones criminales se convierten cada día más poderosas, letales y globales, mientras que por otra parte, los ciudadanos se sienten más inseguros y son cada vez más víctimas de forma directa e indirecta de las acciones de la delincuencia.

Por esas razones, la lucha contra la criminalidad es una de las principales prioridades de muchos estados.

La implementación de estrategias, proyectos y acciones dirigidas a factores claves de seguridad ciudadana, constituye una de las formas de contribuir progresivamente a mejorar su comportamiento. Para ello, se requiere ante todo de una adecuada integración, procesamiento y análisis de información de interés que permita una proyección ajustada a lo que se requiere y necesita.

En la mayoría de los casos la estructura actual de la información de seguridad ciudadana es descentralizada y dispersa, lo cual dificulta sensiblemente la elaboración de análisis que agrupen los diversos y multifactoriales intereses y requerimientos que conllevan los problemas de la criminalidad.

Se hace necesario relacionar las informaciones contenidas en las distintas fuentes de información empleando herramientas informáticas que propicien su integración, transformación y almacenamiento, de tal manera que el análisis de esta información integrada facilite la toma de decisiones en la formulación de medidas, así como en el seguimiento y control de éstas desde una perspectiva proactiva y de corresponsabilidad ciudadana.

En la actualidad no es posible pensar que los analistas recolecten y analicen la información existente en las diferentes instituciones del Estado de manera manual. Las tecnologías de la información han evolucionado y existen tecnologías capaces de integrar, almacenar y distribuir información con tiempos de respuesta aceptable.

Una solución altamente eficiente y validada para el almacenamiento de las Estadísticas Delictivas y otro tipo de Información afin son los Almacenes de Datos que convierten los datos operacionales de una organización en una herramienta profesional, haciéndolos disponibles a los usuarios que lo necesiten para el análisis y toma de decisiones [Thomsen, 2002]. Según Bill H. Inmon, padre reconocido del *Data Warehousing*, “El Data Warehouse (DWH) es una colección de datos orientados al tema, integrados, no volátiles e historizados que facilitan la toma de decisiones” [Inmon, 2005].

Los datos guardados en el almacén deberán unirse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. La información se organizará por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Las peticiones de información serán más fáciles de responder dado que toda la información reside en el mismo lugar. En materia de seguridad ciudadana

resulta imprescindible realizar análisis de tendencias. El Almacén de Datos para la Seguridad Ciudadana debe ser cargado con los distintos valores que toma una variable en el tiempo para permitir comparaciones.

El éxito de una solución de este tipo para contribuir a la Seguridad Ciudadana, y que también se evidencia en otros tipos de aplicaciones similares en un Almacén de Datos viene dado, en gran medida, por un estudio amplio de las Instituciones y Órganos de Seguridad involucradas en el tema, a partir del conocimiento extraído de las fuentes de datos externas y de los responsables de su mantenimiento. No se trata de una mera copia de datos.

### **Metodología de diseño**

La metodología de diseño utilizada se basa en la definición de un modelo de integración de información a almacenes de datos de seguridad ciudadana. El modelo incluye un esquema de desarrollo que impulsa este tipo de solución, así como la selección y clasificación de las fuentes de información según su función, e importancia para el almacén.

### **Integración de Fuentes de Información**

Las tareas principales de un proceso de Integración de Fuentes de Información lo constituyen las siguientes:

1. Se requiere de Nueva Información para ampliar el dominio de análisis. Identificación de Necesidades de Información (Ampliación del Dominio de Análisis)
2. Identificación y clasificación de fuentes fiables de obtención de la Información necesaria identificada para el análisis.
3. Integración de las Fuentes de Información (Órganos de Seguridad y/o otras Instituciones del Estado). Las principales tareas son [Zorrila, 2007]:
  - a. **Extracción**, extracción de datos de las fuentes.
  - b. **Limpieza**, para detectar y eliminar errores, rellenar atributos vacíos, resolver inconsistencias, etc.
  - c. **Transformación**, para transformación los datos al contexto DWH, estandarizar códigos y formatos de representación, definir dominios, utilizar conversiones y combinaciones para generar nuevos campos, corregir datos.
  - d. **Carga**, para definir procesos de carga y sincronización.
4. Análisis y explotación sistemática, por parte de los analistas, del nuevo dominio informativo ampliado que puede conllevar a la necesidad de adquisición de nueva información.

En el ámbito de la Seguridad Ciudadana, son muchas las variables que pueden ser combinadas para obtener resultados que ayuden a establecer patrones de desorden, identificar causas de los altos índices de las incidencias delictivas y realizar análisis para evaluar los resultados de los Órganos de Seguridad y de esta forma tener una visión integradora de todas las variables y sus combinaciones. Existen técnicas precisas para la explotación de los amaneces de datos. En el caso puntual que nos aborda, la Inteligencia de Gobierno, o Inteligencia de Negocios para entornos empresariales, puede ayudar mucho.

Para la identificación de fuentes de información es necesario responderse las siguientes preguntas:

- ¿Alguna de los Órganos de Seguridad proveedores de información, ya integrados, contiene la información que se necesita?
- ¿Ya se está recibiendo de alguna de estos Órganos de Seguridad, y demás fuentes que se necesita pero no está integrada al Almacén?
- ¿Existen en los Órganos de Seguridad sistemas que permitan obtener la información digitalizada, o habría que crearlos?
- ¿Existen condiciones objetivas para que la Información esté disponible en tiempo y espacio (Conectividad, Latencia de Digitalización, otras)?

### **Identificación de Fuentes de Información**

Las principales Fuentes de Información en el entorno de Seguridad Ciudadana son [Espinosa, 2007]:

1. Los Sectores **Justicia y Policía**, que disponen de la más completa información, pero no siempre está organizada y sistematizada.
2. El Sector **Salud**, que atiende las lesiones de víctimas de acuerdo a su severidad, pero pocas veces identifican la causa o móvil por la cual consultan. Clasifican los diagnósticos por el daño físico u orgánico a través del código internacional de enfermedades.
3. La **Policía**, registra los delitos contra la vida e integridad personal, contra la propiedad y contra el estado. Acompaña a los fiscales en los levantamientos de cadáveres y recolectan datos acerca de la víctima, las características del modo, tiempo y lugar, y en general las que pueden contribuir a la investigación criminal. Como son los primeros en llegar al sitio del suceso, los datos relacionados con la hora y fecha de ocurrencia de los hechos, registrados por la Policía, son los más confiables.
4. Las **lesiones personales**, son de conocimiento de la Policía, solo cuando el hecho es denunciado y en algunos casos solo informan los que generan una incapacidad mayor de 30 días. Es una fuente poco confiable en este evento.
5. **Medicina Legal**, prácticas de autopsias a las víctimas de muertes violentas (homicidios, suicidios, muertes de tránsito y otras no intencionales) pero no siempre tienen una cobertura adecuada en los países. También realizan peritajes médico-legales cuando son ordenados por una autoridad judicial. La cobertura de otras formas de violencia está supeditada a las denuncias y requerimientos de los fiscales.
6. **La Fiscalía**, tiene la función de dirigir, realizar y coordinar las investigaciones en materia legal. Su principal objetivo es enjuiciar a los presuntos responsables de los hechos, mediante la recolección de pruebas y la investigación alrededor de los mismos. El sistema de información está orientado a distribuir los casos denunciados a los fiscales para avanzar en las investigaciones.
7. **Institutos de Estadísticas**, que contienen la información más fiable acerca de la demografía de las Distribuciones Político Administrativas que se abordan, en base a Censos de Población y Viviendas y otro amplio conjunto de variables importantes para el tema de la Seguridad Ciudadana, que combinadas con los demás Órganos de Seguridad pueden dar un resultado excelente.

### **Clasificación de Fuentes de Información**

Es necesario clasificar las Fuentes de Información en base a la clasificación de cada una, planificar y desarrollar los mecanismos de integración en cada caso.

Las categorías propuestas son las siguientes [BACH, 2006]:

1. **No cooperativas:** Al exportan un archivo con un formato establecido o permitir una consulta directa con SQL o a través de Servicios Webs, no se garantiza que el destino de la información la integre. A diferencia de las Fuentes Cooperativas que a través de mecanismos de replicación, u otros mecanismos, se establecen intercambios mucho mas fiables, seguros y responsables.
  1. **Fuentes de Información Congelada (Snapshots<sup>2</sup>):** Copia completa de la Información congelada en un instante de tiempo.
  2. **Fuentes Específicas:** Ej. Sistemas Legados, Autónomos, etc. En este conjunto se encuentran las fuentes que brindan la información a partir de archivos intermedios, u otros mecanismos que no implican funcionalidades internas puntuales en base al receptor de la información.
  3. **Fuentes Consultables: Suministra Interfaces para Consultas (SQL, Servicios Webs, etc.).**
2. Cooperativas
  1. **Fuentes de Replicación:** Mecanismos de Publicación/Suscripción.
  2. **Fuentes Call Back<sup>3</sup>:** Se invocan códigos externos de ETL cuando ocurren cambios en la información.
  3. **Fuentes de Cambios Internos:** Se activan acciones internas cuando ocurren los cambios (Triggers).

Luego, se pretende entonces tener un cuadro como se muestra en la figura con todas las fuentes de información.

		Fuente 1	Fuente 2	Fuente 3
<b>NO COOPERATIVA</b>	Snapshot			
	Fuentes Específicas (archivos, etc.)			
	Fuentes Consultables (SQL, WS)			
<b>COOPERATIVA</b>	Fuentes de Replicación			
	Fuentes de Call Back			
	Fuentes de Cambios Internos			

Fig. 1. Clasificación de Fuentes de Información.

Las dos últimas clasificaciones permitirían incluso tener el estado de las estadísticas y ocurrencias delictivas a tiempo real. Sobre la integración de Información a tiempo real, vale destacar que la intención de un Almacén de Datos no es operativa, es decir, que la tendencia del uso de almacenes de datos está enfocada a estudios posteriores a las

<sup>2</sup> Del termino en inglés “Foto”

<sup>3</sup> Retroalimentación

ocurrencias de los hechos, y no a la obtención de resultados a tiempo real. La tendencia actual es a procurar tener todas las fuentes de información como consultables, a partir de conexiones directas desde los Ministerios y Órganos Superiores del Estado, y se deben definir cortes en el tiempo para la obtención de la información.

Luego de Identificadas las fuentes, de tener suficiente experiencia y de entender los mecanismos de transferencia de información debe analizarse si es necesario cambiar de categoría alguna fuente dado el caso de necesitar acceder a la información a demanda, a tiempo real, etc.

		Fuente 1	Fuente 2	Fuente 3
NO COOPERATIVA	Snapshot	(X)		
	Fuentes Específicas (archivos, etc.)	↓	(X) ↓	(X) ↓
	Fuentes Consultables (SQL, WS)	(X)	(X)	(X)
COOPERATIVA	Fuentes de Replicación			
	Fuentes de Call Back			
	Fuentes de Cambios Internos			

Fig. 2. Las fuentes pueden ser consultadas a demanda.

Puede incluso pensarse en tener resultados con latencia mínima.

		Fuente 1	Fuente 2	Fuente 3
NO COOPERATIVA	Snapshot	(X)		
	Fuentes Específicas (archivos, etc.)	↓	(X)	(X)
	Fuentes Consultables (SQL, WS)	(X)	↓	↓
COOPERATIVA	Fuentes de Replicación		↓	↓
	Fuentes de Call Back		(X)	(X)
	Fuentes de Cambios Internos		(X)	(X)

Fig. 3. Integración preparada para Latencia Mínima o Tiempo Real.

Cualquiera de las variantes acarrearía un cambio en los mecanismos y procesos de Integración.

Por todo esto, la clasificación de las fuentes de información constituye uno de los principales puntos en el Proceso de Integración, ya que de ello dependen los mecanismos utilizados para su funcionamiento.

## **Condiciones para la Integración de las Fuentes**

Para el éxito de este tipo de solución deben cumplirse las siguientes condiciones [Espinosa, 2007]:

1. **Voluntad Política**, de los gobernantes, autoridades locales y nacionales, de utilizar la información par a la toma de decisiones.
2. **Coordinación con las Instituciones, fuentes de Información** (Policía, Medicina Forense, Fiscalía, Salud Publica, otras).
3. **Tecnología Apropriada (Formación de Capital Humano y acompañamiento técnico y pedagógico)**, Software, análisis de información, estrategias de intervención, evaluación de estrategias, etc.

## **Conclusiones**

El análisis de los diferentes factores que intervienen en la Inseguridad de los Ciudadanos en todos los pueblos del mundo, es un aspecto que debe priorizarse. El estudio de tales factores, a través de sistemas de información integrados, parece ser la solución a tales problemas, para luego, con la experiencia y el conocimiento adquirido, sea posible la toma de medidas y estrategias acertadas que den una solución verdadera a la violencia, la inseguridad, y permitan el desarrollo de sociedades limpias de criminales y de incidentes que atenten contra el desarrollo y la integridad de la humanidad. La Seguridad Ciudadana es tarea de todos.

## **Referencias Bibliográficas**

[Thomsen, 2002] Thomsen, Erick, OLAP Solutions, Building Multidimensional Information Systems, Second Edition. Jhon Wiley & Sons, Inc. USA, (2002)

[Inmon, 2005] W.H. Inmon, Building Data Warehouse, Fourth Edition. USA. Jhon Wiley&Son, Inc, (2005).

[Espinosa, 2007] Espinosa del Vallin, Rafael, Evaluación de la Metodología indicadores y estadísticas sobre seguridad ciudadana en América Latina, Santiago de Chile, (2007)

[BACH, 2006]Bach Pedersen, Torben, Extract, Transform and Load. Aalborg University, (2006).

[Zorrilla, 2007] Zorrilla, Marta, Data Warehouse y OLAP, Universidad de Cantabria, (2007).