

 Servicio Andaluz de Salud CONSEJERÍA DE SALUD	Estrategia Explotación de Datos Catálogo de Normas	Ver 0.2
	Oficina de Calidad Subdirección de Tecnologías de la Información	



Oficina de Calidad
 Subdirección de Tecnologías de la Información

Estrategia Explotación de Datos Catálogo de Normas

	Estrategia Explotación de Datos Catálogo de Normas	Ver 0.2
	Oficina de Calidad Subdirección de Tecnologías de la Información	

CONTROL DE CAMBIOS DEL DOCUMENTO

Autor	Versión	Referencia de cambios	Fecha
Jesús Amaya	Inicial		10/12/2013
Jesús Amaya, Francisco Barroso	Corrección, Adaptación	Adaptación a cambios tecnológicos	01/10/2015
Oficina Técnica BI	Nueva versión	Redacción " <i>de novo</i> " de normas. Adaptación a nuevas prácticas de gestión de proyectos y a cambios tecnológicos. Inclusión de enfoques de responsabilidad de cumplimiento, consecuencias de incumplimiento, responsabilidad de control de cumplimiento y arbitrio.	15/11/2016

Nombre fichero	Nrm_Explotacion_Datos_v02.docx
Fecha de creación	15/11/2016
Última actualización	21/11/2016

	Estrategia Explotación de Datos Catálogo de Normas	Ver 0.2
	Oficina de Calidad Subdirección de Tecnologías de la Información	

Índice

Control de cambios del documento.....	2
Objeto del documento	4
Alcance	5
Catálogo de Normas.....	6
Normas de carácter general.....	7
Normas referidas a la fase de definición Requisitos de los sistemas de explotación	7
Normas referidas a la fase de Análisis de los sistemas de explotación.....	7
Normas referidas a la fase de Diseño de los sistemas de explotación	7
Normas referidas a la fase de Construcción de los sistemas de explotación	9
Normas referidas a la fase de Explotación de los sistemas de explotación.....	12

	Estrategia Explotación de Datos Catálogo de Normas	Ver 0.2
	Oficina de Calidad Subdirección de Tecnologías de la Información	

Objeto del documento

Este documento contiene las normas relacionadas con los aspectos claves a la hora de afrontar un proyecto de explotación de datos. El objetivo que persigue es la uniformidad de criterios y la identificación y estandarización de procesos. Así, en este documento se catalogan las normas de la Oficina de calidad de la Subdirección de Tecnologías de la Información del Servicio Andaluz de Salud, relativas a:

Las generalidades de los sistemas de explotación de datos

La definición de requisitos de los sistemas de explotación de datos

El análisis de los sistemas de explotación de datos

El diseño de los sistemas de explotación de datos

La construcción de los sistemas de explotación de datos

La explotación de los sistemas de explotación de datos

	Estrategia Explotación de Datos Catálogo de Normas	Ver 0.2
	Oficina de Calidad Subdirección de Tecnologías de la Información	

Alcance

Este documento está dirigido a:

Los jefes de proyectos de explotación de datos.

Los técnicos de proyectos de explotación de datos.

Los proveedores de servicios TIC para proyectos de explotación de datos.

Los consultores funcionales de cada una de las áreas implicadas en el Data Warehouse (en adelante DW)

	Estrategia Explotación de Datos Catálogo de Normas	Ver 0.2
	Oficina de Calidad Subdirección de Tecnologías de la Información	

Catálogo de Normas

El Data Warehouse Corporativo o Almacén de Datos Corporativo del SAS cuenta con información procedente de diferentes fuentes. Con el objeto de respetar la integridad y la uniformidad de la información y estructura del Data Warehouse, los actores a los que se dirige el contenido de este documento, deben seguir una serie de normas de obligado cumplimiento. Estas normas irán asociadas a cada una de las fases del ciclo de vida de los proyectos de explotación de datos.

En este catálogo de normas, también se han incluido algunas recomendaciones, a modo de buenas prácticas, que fomenten la calidad de los productos de los proyectos de explotación de datos.

El incumplimiento de las normas designadas como "*normas críticas*" podrá conllevar la interrupción parcial o total del desarrollo del proyecto y su modificación eventual sin coste adicional de desarrollo para la STIC.

Si el incumplimiento de normas concretas en proyectos específicos fuera necesario, deberá solicitarse a la OTBI a través del jefe de proyecto. Dicha solicitud deberá acompañarse de una justificación razonada para el incumplimiento de la norma, y solo podrá materializarse después de la aprobación formal y explícita de la OTBI.

Este catálogo se cumplimenta con la información incluida en el *Anexo al documento de Estrategia Explotación de Datos, Catálogo de Normas*, sobre la criticidad de las normas, la responsabilidad del cumplimiento, las consecuencias del incumplimiento, la responsabilidad del control del cumplimiento y el arbitrio en caso de disconformidad de alguno de los actores obligados por las normas.

	Estrategia Explotación de Datos Catálogo de Normas	Ver 0.2
	Oficina de Calidad Subdirección de Tecnologías de la Información	

Normas de carácter general

NG001. Los proyectos de explotación de información se orientarán a un uso analítico flexible, permitiendo la realización de análisis dinámicos.

Esta norma es **crítica**, su incumplimiento conllevará la interrupción del desarrollo del proyecto y su modificación eventual sin coste adicional para la STIC.

NG002. El Data Warehouse corporativo será la única fuente de información de los diferentes sistemas de explotación que se implanten.

Normas referidas a la fase de definición de Requisitos

NR001. Los consultores funcionales designados por la STIC validarán los requisitos funcionales en un plazo máximo de 15 días hábiles a partir de su comunicación formal por parte de la OTBI. Si la validación formal de los consultores funcionales no se produce en este plazo, los requisitos funcionales se considerarán validados.

RR001. Recomendación: No es recomendable incorporar entidades al sistema que no procedan de peticiones expresas de los consultores funcionales.

NR002. Los consultores funcionales designados por la STIC deberán identificar y definir los indicadores que se vayan a implementar en el proyecto de explotación de datos.

NR003. Los consultores funcionales designados por la STIC serán los encargados de identificar las jerarquías que se deban incorporar al sistema de explotación, según sus necesidades de información.

Normas referidas a la fase de Análisis de los sistemas de explotación

NA001. La inclusión de campos no explotables en cualquier sistema de explotación deberá comunicarse al jefe de proyecto de explotación de datos, junto con la justificación razonada para su inclusión, y solo podrá materializarse después de la aprobación formal y explícita del jefe de proyecto.

Normas referidas a la fase de Diseño de los sistemas de explotación

ND001. Las Tablas de Dimensiones, tablas que se usan para categorizar los hechos, se denominarán con un nombre que comience por el prefijo "**TD**".

 Servicio Andaluz de Salud CONSEJERÍA DE SALUD	Estrategia Explotación de Datos Catálogo de Normas	Ver 0.2
	Oficina de Calidad Subdirección de Tecnologías de la Información	

ND002. Las Tablas de Hechos, tablas que contienen los diferentes hechos o medidas, se denominarán con un nombre que comience por el prefijo "**TH_**".

ND003. Las Tablas de Relación, tablas que se utilizan para relacionar dos atributos del modelo lógico, normalmente para modelar una relación de muchos a muchos, se denominarán con un nombre que comience por el prefijo "**TR_**".

ND004. Las Columnas de Códigos, campos que contienen las claves de negocio, se denominarán con un nombre que comience por el prefijo "**COD_**".

ND005. Las Columnas de Fechas, campos con formato de fecha, se denominarán con un nombre que comience por el prefijo "**FEC_**".

ND006. Las Columnas de Identificadores, campos que se utilizan para almacenar las claves primarias y subrogadas que se generen durante el proceso de carga, se denominarán con un nombre que comience por el prefijo "**ID_**".

ND007. Las Columnas de Descripciones, campos de tipo descripción, se denominarán con un nombre que comience por el prefijo "**DESC_**".

ND008. Las Columnas IND, campos de tipo flag, se denominarán con un nombre que comience por el prefijo "**IND_**".

ND009. Las Columnas que contengan importes, campos que se utiliza para los hechos de tipo importe dinerario, se denominarán con un nombre que comience por el prefijo "**IMP_**".

ND010. Las Columnas utilizadas para los hechos de tipo número, se denominarán con un nombre que comience por el prefijo "**NUM_**".

ND011. Las Secuencias se denominarán con un nombre que comience con el prefijo "**SQ_**" seguido del nombre de la secuencia (***SQ_Nombre de la Secuencia***).

ND012. La tablas de hechos únicamente contendrá los elementos:

Clave primaria: Clave que identifica unívocamente cada registro de la tabla de hechos.

Claves foráneas: Claves que se utilizarán para los enlaces con las dimensiones asociadas a la tabla de hecho.

Hechos: Columnas que almacenan la información que se pretende medir. Los hechos son la base para el cálculo de los indicadores.

El conjunto de dimensiones definirán la granularidad de la tabla de hechos.

	Estrategia Explotación de Datos Catálogo de Normas	Ver 0.2
	Oficina de Calidad Subdirección de Tecnologías de la Información	

ND013. Se podrán incluir reglas de negocio en las tablas de hechos. La implementación de estas reglas se realizará mediante un atributo de tipo numérico, cuyo nombre estará relacionado con la regla de negocio y que tome el valor 1 cuando se cumpla la regla de negocio, y 0 en caso contrario. Es decir se creará un campo de tipo flag (**IND_**).

Esta norma es **crítica**, su incumplimiento conllevará la interrupción del desarrollo del proyecto y su modificación eventual sin coste adicional para la STIC.

ND014. En el Modelo Físico se recogerán las *tablas de agregados* necesarias para agilizar los tiempos de respuesta de las consultas que se realicen sobre el Data Warehouse que presenten bajos rendimientos de ejecución (tiempos de ejecución prolongados), por su complejidad o por la extensión de la información consultada.

ND015. Las tablas de agregados se denominarán añadiendo el sufijo “**_AGR**” al nombre de la tabla original de la que procede. Ej.: Tabla origen: “*TH_Altas*”. Tabla de agregados: “*TH_Altas_AGR*”.

ND016. Los consultores funcionales designados por la STIC validarán el modelo lógico en un plazo máximo de 15 días hábiles a partir de su comunicación formal por parte de la OTBI. Si la validación formal por parte de los consultores funcionales no se produce en este plazo, el modelo lógico se considerará validado.

Esta norma es **crítica**, su incumplimiento conllevará la interrupción del desarrollo del proyecto y su modificación eventual sin coste adicional para la STIC.

RD001. Recomendación: Para la definición de los elementos del modelo lógico es recomendable utilizar términos que sean familiares para los usuarios del negocio.

El modelo de los sistemas a implementar deberán ser modelos en estrella o copo de nieve.

Esta norma es **crítica**, su incumplimiento conllevará la interrupción del desarrollo del proyecto y su modificación eventual sin coste adicional para la STIC.

Normas referidas a la fase de Construcción de los sistemas de explotación

NC001. La base de datos que se usará para albergar el “*Data Warehouse*” corporativo estará basada en Oracle 11.2.

NC002. Para la implementación de los diferentes datamarts, se usará “*Microstrategy 10.4*” sobre Base de datos “*Oracle*”. El uso de otra tecnología deberá ser aprobado expresamente por la OTBI.

 Servicio Andaluz de Salud CONSEJERÍA DE SALUD	Estrategia Explotación de Datos Catálogo de Normas	Ver 0.2
	Oficina de Calidad Subdirección de Tecnologías de la Información	

NC003. Los procesos ETL (Extracción, Carga y Transformación) para cargar datos desde las bases de datos operacionales hacia el DW se llevarán a cabo mediante ODI ("*Oracle Data Integrator*") versión 12c.

NC004. No se usará código genérico ("*PL/SQL*") salvo bajo aprobación expresa y formal de la OTBI.

NC005. El modelo de explotación final se albergará en la base de datos de explotación:
Esquema de Base de Datos: DWSAS_NombreMódulo.

NC006. Las tablas temporales e intermedias necesarias para obtener las tablas con los datos finales del esquema de explotación se ubicarán en la base de datos de carga, Esquema de Base de Datos: DWSAS_ETL_NombreMódulo, que se usará como área de "Staging" durante el proceso de carga.

NC007. Los proyectos de carga usarán como información de auditoría la proporcionada por "*ODI Console*", por lo que no será necesario implementar instrucciones de auditoría en los propios procesos de carga.

Esta norma es **crítica**, su incumplimiento conllevará la interrupción del desarrollo del proyecto y su modificación eventual sin coste adicional para la STIC.

NC008. El proceso general de carga se dividirá en subprocesos independientes. El módulo gestor de la carga deberá ser capaz de ejecutar los subprocesos de carga de forma aislada, sin necesidad de ejecutar el proceso general completo.

NC009. Las cargas deben ser programadas, secuenciales, periódicas con periodicidad fija, automáticas y desatendidas, sin la necesidad de ninguna intervención manual de operadores, y parametrizadas a estos efectos.

Esta norma es **crítica**, su incumplimiento conllevará la interrupción del desarrollo del proyecto y su modificación eventual sin coste adicional para la STIC.

NC010. Todas las tablas del DW, tanto las tablas de hechos como de dimensión, tendrán obligatoriamente el campo "**FEC_CARGA**" que reflejará la fecha de carga/modificación de cada registro. El formato de este campo será "*Datetime*".

Esta norma es **crítica**, su incumplimiento conllevará la interrupción del desarrollo del proyecto y su modificación eventual sin coste adicional para la STIC.

NC011. Todas las tablas del Data Warehouse deberán disponer de clave primaria.

Esta norma es **crítica**, su incumplimiento conllevará la interrupción del desarrollo del proyecto y su modificación eventual sin coste adicional para la STIC.

 Servicio Andaluz de Salud CONSEJERÍA DE SALUD	Estrategia Explotación de Datos Catálogo de Normas	Ver 0.2
	Oficina de Calidad Subdirección de Tecnologías de la Información	

NC012. No se versionará ninguna dimensión salvo por requerimiento funcional. Durante la fase de definición se indentificarán todas las dimensiones que deberán versionarse, indicando los campos que generarán nueva versión y el tipo de versión usada.

NC013. La implementación del versionado de las tablas de dimensión se llevará a cabo con la función "**DW_FU_CAMBIO_VER**", no estando permitido ningún mecanismo alternativo. Toda la documentación y fuentes asociados a dicha función, se encuentra disponible en la comunidad de Explotación de Unifica.

NC014. Para la implementación del versionado se utilizará el "*Módulo de conocimiento para dimensiones lentamente cambiantes*" proporcionado por ODI.

NC015. En el caso de las dimensiones versionables, el enlace se deberá realizar teniendo en cuenta que el periodo de validez de la dimensión debe contener a la fecha del hecho.

Esta norma es **crítica**, su incumplimiento conllevará la interrupción del desarrollo del proyecto y su modificación eventual sin coste adicional para la STIC.

NC016. Las transformaciones de la información se realizarán en un área de "*Staging*" del proyecto de carga, antes de cargarla en el Data Warehouse. Finalizada la carga en el Data Warehouse, los datos del area de "*Staging*" deberán ser eliminados.

NC017. Durante el proceso de carga de todas las dimensiones y tablas de hecho se deberán incorporar los valores -1 (No definido) y -2 (Dato erróneo).

Esta norma es **crítica**, su incumplimiento conllevará la interrupción del desarrollo del proyecto y su modificación eventual sin coste adicional para la STIC.

NC018. La carga del Data Warehouse debe realizarse a través de ficheros planos delimitados.

Esta norma es **crítica**, su incumplimiento conllevará la interrupción del desarrollo del proyecto y su modificación eventual sin coste adicional para la STIC.

NC019. Las tablas muy pobladas, con más de 1 000 000 de registros, deben particionarse en el "*Data Warehouse*".

NC020. No se permite el uso de cursores.

Esta norma es **crítica**, su incumplimiento conllevará la interrupción del desarrollo del proyecto y su modificación eventual sin coste adicional para la STIC.

NC021. Se deberá indicar toda la información necesaria para ejecutar los procesos de carga, tanto de forma manual como desatendida.

RC001. Recomendación: Se recomienda la desactivación de claves foráneas "*foreing key*", antes de las cargas, y su activación posterior a las mismas.

	Estrategia Explotación de Datos Catálogo de Normas	Ver 0.2
	Oficina de Calidad Subdirección de Tecnologías de la Información	

RC002. Recomendación: Se recomienda que la integridad referencial se mantenga mediante la lógica de los procesos de carga.

RC003. Recomendación: Se recomienda que los procesos de carga sean siempre incrementales.

RC004. Recomendación: Se recomienda que la carga inicial del DW se diseñe de modo que pueda ejecutarse por fases definidas en base a criterios temporales o estructurales, y que contemple la posibilidad de su ejecución modular.

RC005. Recomendación: Se recomienda que las claves primarias se obtengan a partir de secuencias autonuméricas.

RC006. Recomendación: Si se contempla la carga de archivos en "Oracle" existen dos opciones, mediante "SQL * Loader" y mediante Tablas Externas. En este caso se recomienda que la carga se realice utilizando tablas externas en lugar de "SQL * Loader".

RC007. Recomendación: En las tablas de hechos se recomienda el uso de "Bitmap index" para cada uno de los campos que enlazan con dimensiones de baja cardinalidad (no más de 20 valores discretos).

NC022. Si el desarrollo tienen parte de aplicativo, se entregará un manual de usuario que permita navegar y utilizar la herramienta de forma eficiente.

Normas referidas a la fase de Explotación de los sistemas de explotación

NE001. Solo se accederá una única vez a las fuentes originarias. No está permitido acceder más de una vez a la misma fuente de información para alimentar el Data Warehouse con una información dada.

NE002. Se identificarán todos los roles de acceso al sistema de explotación posibles.

Esta norma es **crítica**, su incumplimiento conllevará la interrupción del desarrollo del proyecto y su modificación eventual sin coste adicional para la STIC.

NE003. Los permisos de acceso se concederán a través de roles, nunca a usuario nominales, siendo el proceso de alta el que asigne el usuario a un determinado rol.

Esta norma es **crítica**, su incumplimiento conllevará la interrupción del desarrollo del proyecto y su modificación eventual sin coste adicional para la STIC.

NE004. Los roles irán asociados a los ámbitos sobre los que se apliquen.

NE005. Los roles se definirán de forma que sean reutilizables entre diferentes ámbitos.

	Estrategia Explotación de Datos Catálogo de Normas	Ver 0.2
	Oficina de Calidad Subdirección de Tecnologías de la Información	

NE006. Se evitará la duplicidad de roles entre diferentes ámbitos.

NE007. Los roles se definirán enfocados al negocio y no a las funcionalidades.

NE008. Para cada rol habrá que definir los privilegios asociados, es decir, el conjunto de funcionalidades que tiene disponible cada rol.

NE009. Se inhabilitarán las credenciales de acceso a los usuarios que no usen el sistema de explotación durante un período prolongado de tiempo, por defecto un año.

NE010. Los informes identificados como periódicos, se planificarán y ejecutarán de forma desatendida.

NE011. La prioridad de ejecución de los informes será establecida por los consultores funcionales.

RE001. Recomendación: Para la ejecución de informes se recomienda asignar prioridades si el sistema de reporting lo permite, de forma que los informes considerados prioritarios no se vean ralentizados por informes secundarios.

RE002. Recomendación: Se recomienda que los informes se basen en plantillas genéricas predefinidas en cada proyecto.

RE003. Recomendación: Para el formateo de las salidas de informes se recomienda utilizar estilos predefinidos (títulos, subtítulos, cabecera de tablas, importes...) , evitando el uso de formatos directos.

RE004. Recomendación: Para el formateo de atributos se recomienda utilizar estilos, evitando el uso de formatos directos.

RE005. Recomendación: Se recomienda que los informes relacionados entre sí desde el punto de vista funcional sean navegables mediante el uso de enlaces.

NE012. Todos los objetos relacionados con la explotación de datos que no se utilicen en un periodo de tiempo superior al año y medio serán propuestos para su eliminación del sistema en producción, para almacenarse en un sistema de respaldo.

NE013. La documentación de ayuda, tanto de los informes como de los posibles indicadores que se muestren en ellos, deberá tener una descripción asociada y debe ser accesible desde la propia herramienta de explotación.