

MEMORIA TÉCNICA DE LA ACTIVIDAD

“Banco de datos Estadísticos de Andalucía (BADEA)”

0. IDENTIFICACIÓN.....	2
1. INTRODUCCIÓN.....	3
2. ÁMBITO DE ESTUDIO.....	7
3. RECOGIDA O CAPTURA DE DATOS.....	8
4. FLUJO O PROCESO DE TRABAJO.....	9
5. PLAN DE DIFUSIÓN.....	13
6. CALIDAD.....	14





0. IDENTIFICACIÓN

- **Código y denominación de la actividad:** 14.00.08 Banco de datos Estadísticos de Andalucía (BADEA)
- **Organismo responsable:** Instituto de Estadística y Cartografía de Andalucía (IECA)
- **Unidad ejecutora:** Servicio de Gestión de la Información
- **Organismos colaboradores y convenio:** Ninguno



1. INTRODUCCIÓN

- **Objetivos:**

- Poner a disposición de los productores de estadística del Sistema Estadístico y Cartográfico de Andalucía, en adelante SECA, un repositorio de información centralizado.
- Ofrecer los procesos y mecanismos necesarios para difundir de forma segura.
- Permitir a los usuarios interactuar con los datos y exportarlos en distintos formatos, facilitando su uso a través de servicios interoperables.
- Facilitar la reutilización de los datos al permitir cruces de información procedente de distintas fuentes.
- BADEA utiliza nomenclaturas y clasificaciones estandarizadas que permiten la comparabilidad tanto en el territorio como en el ámbito de estudio de las distintas actividades.

- **Marco conceptual:**

Algunos elementos que se pueden encontrar en BADEA son:

- Dimensión: es un concepto a través del cual se puede clasificar la información, por ejemplo el sexo, la edad, etc.
- Categoría: Cada dimensión consta de varias categorías, por ejemplo en la dimensión sexo, las categorías son hombre, mujer y ambos sexos.
- Medida: es lo que se está cuantificando, por ejemplo, la población, los vehículos, los inmigrantes, etc.
- Cubo: Se trata de una tabla donde se almacena toda la información histórica de una o varias medidas con sus dimensiones asociadas, un ejemplo de cubo podría ser 'Cubo de población', que recogerá la cifra de población (medida), para los distintos años, territorios, por sexo y por edad (todas estas variables son las dimensiones). Cada actividad estadística tiene sus propios cubos y los responsables de cada actividad se encargan de alimentarlos periódicamente.
- Cubo-periodo: Es la parte del cubo referida a un periodo concreto, un cubo tendrá tantos cubos-periodos como periodos diferentes tenga.
- Consultas: Son las tablas a través de las cuales se le da difusión a la información contenida en los cubos, cada consulta puede alimentarse de uno o de varios cubos. Siguiendo con el ejemplo anterior, una consulta podría ser 'Población por provincia y sexo' y otra 'Población por provincia y grupos de edad'. Ambas consultas tirarían del 'Cubo de Población mencionado anteriormente.



- Especificación de carga: Se trata de un diseño de registro del cubo en el que se indica en qué formato se ha de cargar la información, si el fichero tiene o no cabeceras, el símbolo decimal, el símbolo de separación de campos, el orden de las columnas de datos, etc.
 - Árbol de consultas: Es la estructura que permite organizar las consultas de forma ordenada en base al criterio definido por el productor. Hay tres tipos de árboles de consultas Consultas individuales , Informes o Privadas.
 - Consultas individuales: en los árboles de consultas individuales, para visualizar las consultas hay que ejecutarlas una a una.
 - Informes: en el árbol tipo Informes las consultas aparecen ya abiertas y se muestran una a continuación de otra, este tipo de árboles se suele utilizar más en consultas que no sean de gran tamaño.
 - Privadas: este tipo de árboles se utiliza para organizar consultas de uso interno, que no se van a publicar en internet, por ejemplo consultas de comprobación, o de cálculos internos.
- **Marco jurídico:**
 - Ley 4/1989, de 12 de diciembre, de Estadística de la Comunidad Autónoma de Andalucía.
 - Ley 9/2023, de 25 de septiembre, por la que se aprueba el Plan Estadístico y Cartográfico de Andalucía 2023-2029 y sus programas estadísticos y cartográficos de desarrollo.

- **Antecedentes:**

Hasta 2007, cada actividad estadística generaba sus resultados en un soporte propio y con diferentes grados de documentación. Esa dispersión de información, tanto en lo que respecta a soporte como a formato, dificultaba en gran medida localizar y dar respuesta a las demandas de información estadística.

Por este motivo surgió la necesidad de crear un repositorio único, que utilizase clasificaciones comunes estandarizadas en la medida de lo posible y ofreciese una herramienta de difusión común para todos los productos.

Aunque BADEA es realmente un banco de datos y una herramienta de difusión, se incluye en el Programa Estadístico Anual en 2007 como actividad estadística, permaneciendo así hasta la fecha.

El primer paso en la creación de BADEA consistió en hacer un análisis exhaustivo de la información que se producía en el IECA y que podría ser susceptible de incorporarse. Se trató de incluir la mayor cantidad de información posible, a continuación había que definir el nivel de detalle al que se iba a almacenar la información, el suficiente para satisfacer las necesidades de difusión y para atender a determinadas peticiones a medida (las más comunes). El siguiente paso fue adaptar la información existente a la estructura de datos requerida por BADEA, haciendo hincapié en homogeneizar y estandarizar las clasificaciones usadas. Una vez cargada la información en BADEA se procedió a crear las consultas e integrarlas en los árboles de consultas en base al plan de tabulación deseado



para cada actividad y finalmente se procedió a la publicación a través de BADEA, sustituyendo esta por las que se venían haciendo en hojas de cálculo u otros formatos dispuestos para su difusión.

- **Justificación y utilidad:**

BADEA agrupa toda la información producida en un repositorio común, lo que lo convierte en un referente donde depositar la información generada, tanto para los servicios productores del IECA como de otros organismos del SECA, dejando a un lado las aplicaciones departamentales usadas anteriormente.

Facilita los procesos de carga y difusión a los productores de datos, puesto que una vez que está diseñada la estructura de la actividad, el proceso de carga para periodos futuros se simplifica enormemente, añadiendo además los controles necesarios para perseverar el secreto estadístico y mantener ocultos los datos hasta el momento de su publicación según calendario.

Facilita a los usuarios el acceso a la información, ya que ofrece un marco único para todas las estadísticas. Además, permite interactuar con los datos a través de filtros, construir series temporales, descargarse la información en distintos formatos, generar gráficos y mapas con la información de las tablas

Facilita la reutilización de la información estadística a través del servicio API Restful que permite acceder mediante peticiones URL a toda la información disponible en BADEA.

Utiliza los mismos estándares de codificación y definición de variables para todas las actividades, lo que facilita y garantiza la comparabilidad.

A nivel de explotación, toda la información se encuentra recopilada en el mismo lugar, por lo que el dato almacenado será único aunque sea utilizado en diferentes consultas o tablas, con lo que se evita cometer errores de duplicidades de datos y se fomenta la reutilización. BADEA permite hacer operaciones con la información almacenada, aunque provenga de distintas fuentes u organismos, lo que facilita el trabajo al productor al evitarle hacer cálculos innecesarios.

- **Restricciones y alternativas:**

- *Restricciones externas:* Al ser una herramienta bastante compleja, el desarrollo y las mejoras deben realizarlas empresas muy especializadas, lo que en ocasiones hace dependiente a BADEA de especialistas externos muy concretos.
- *Restricciones internas:* Hay algunas actividades estadísticas que por su naturaleza y grado de especialización no se adaptan al esquema de datos de BADEA. También se presentan problemas de rendimiento cuando hay un volumen grande de información que procesar, sobretodo por la existencia de límites establecidos para evitar que dichas consultas puedan colapsar la red, y que impiden en algunos casos poder interactuar (filtrar) las consultas prediseñadas.
- *Alternativas:* En cuanto a la restricción externa se ha avanzado mucho en la documentación de la programación de todos los procesos, aunque todavía hay casos en los que se muestra como concepto de error de carga 'Error inesperado'. Uno de los objetivos de los próximos evolutivos



va encaminado en subsanar este problema. En cuanto a las actividades que no se pueden incluir en BADEA, se trata de actividades que se gestionan en otros bancos de datos específicos por lo que en principio, no se ve inconveniente en que siga siendo así. En lo que respecta a la limitación que tienen algunas consultas en las que se ha eliminado la posibilidad de filtrado, se está trabajando en diferentes alternativas para ofrecer a los usuarios la información a través de servicios interoperables, evitando que pasen necesariamente por la ejecución de la consulta a través de BADEA.

- **Comparabilidad territorial:**

No se tiene conocimiento de que otras oficinas estadísticas hayan desarrollado trabajos similares que abarquen a toda la producción. A nivel de proyectos concretos, sabemos que tanto el Instituto Nacional de Estadística (INE) como el Instituto Canario de Estadística han abordado la creación de bancos aunque en estos casos no se cargan generalmente los microdatos sino tablas ya fruto de una elaboración externa de los datos.



2. ÁMBITO DE ESTUDIO

- **Objeto de estudio:** Actividades estadísticas del SECA.
- **Resolución, escala o desagregación del objeto de estudio:** La desagregación temporal máxima es el mes, la desagregación territorial máxima en la mayoría de los casos es el municipio, aunque hay algunas variables que se cargan a nivel de sección censal.
- **Fenómenos o variables:** Las variables que se contemplan son muy diversas y abarcan todos los aspectos que se contemplan en el Plan Estadístico y Cartográfico: territoriales, demográficos, económicos, sanitarios, sociales, educativos, económicos, etc.



3. RECOGIDA O CAPTURA DE DATOS

- **Sujeto informante:** Organismos productores de estadísticas del SECA.
- **Tipología de datos a suministrar:** Se trata de un banco de datos que suministra información de distinta tipología (población, empresas, hospitales, centros de enseñanza, asalariados, parados, presupuestos, ...).
- **Periodicidad:** Continua.
- **Método de obtención:** La información se obtiene de la carga que hacen los productores de estadísticas que están adscritos al banco de datos, a través de ficheros de carga.

Actualmente abarca la práctica totalidad de la producción estadística que se lleva a cabo en el IECA, y además, la producción de las siguientes Consejerías:

- Salud y Consumo
- Cultura y Deporte
- Universidad, Investigación e Innovación
- Economía, Hacienda y Fondos Europeos. Agencia Tributaria de Andalucía
- Empleo, Empresa y Trabajo Autónomo
- Fomento, Articulación del Territorio y Vivienda. Agencia de Vivienda y Rehabilitación de Andalucía (AVRA)
- Justicia, Administración Local y Función Pública y Agencia de Servicios Sociales y Dependencia de Andalucía (ASSDA)
- Inclusión Social, Juventud, Familias e Igualdad
- Industria, Energía y Minas

Está en proceso de inclusión la consejería de Desarrollo Educativo y Formación Profesional



4. FLUJO O PROCESO DE TRABAJO

- **Preparación y tratamiento base de la información:**

Una vez que el productor ha generado la información conforme al diseño de registro del cubo que la soporta, el sistema dispone de una serie de mecanismos desarrollados para ir cargando los distintos periodos.

Previamente el administrador de BADEA ha creado el cubo en la actividad que corresponda.

El acceso a BADEA se realiza a través de usuario y contraseña y cada usuario tiene acceso a determinadas funcionalidades en función de su perfil (cargar datos, crear consultas, dar aprobación técnica a los datos, dar aprobación de publicación...). En el caso de los usuarios productores, estos solo tendrán acceso a las actividades que tenga asignadas. Las urls de acceso a BADEA son las siguientes:

<https://badea.ieca.junta-andalucia.es/intranet/admin/login>

<https://www.juntadeandalucia.es/institutodeestadisticaycartografia/badea/intranet/admin/login>

Cuando el usuario accede a BADEA, la aplicación solicitará que elija la actividad estadística sobre la que desea trabajar, y a partir de ese momento, todas las acciones que realice estarán orientadas a esa actividad.

Para realizar la carga de un cubo concreto, en primer lugar hay que crear los Cubos-periodos. De esta forma estamos creando en el cubo el espacio para a continuación poder cargar los datos de dicho periodo dentro del cubo.

Una vez creado el cubo-periodo ha de crearse una Especificación de carga para definir el diseño de registro del fichero de carga; aquí se indicará si se trata de un fichero excel o texto (son las dos opciones de fichero de carga que ofrece BADEA). Hay que indicar también si el fichero tiene cabeceras o no, el símbolo usado para decimales y el usado para separación de campos, etc. En el fichero de carga las primeras columnas corresponderán a las dimensiones, y a continuación estarán las medidas, en el mismo orden en el que aparecen en el cubo.

Finalmente, desde el mismo espacio donde hemos creado la especificación de carga, hay una opción que permite seleccionar el fichero de carga y ejecutar la carga. Si el proceso no detecta ningún error, aparecerá en el estado de la carga 'Finalizado' mientras que si hubiese errores en el fichero de carga, el estado de la carga lo indica y en el log nos da información sobre dicho error.

Una vez cargados los datos en el cubo (uno o varios periodos), la forma de visualizar la información contenida en el cubo es a través de consultas que ha diseñado el productor de los datos, y que organizará dentro de un árbol de consultas, haciendo este árbol de índice del plan de tabulación.



Para diseñar las consultas, BADEA dispone de un módulo denominado ‘Asistente de consultas’ basado en tecnología JPIVOT, (para conocer más acerca de la creación de consultas, consultar el manual de BADEA).

En la imagen siguiente se muestra un ejemplo de consulta

Sistema de Información Multiterritorial de Andalucía (SIMA)

Nacimientos por sexo del recién nacido y residencia materna
Unidad de medida: **Nacimientos** Periodicidad: **Anual**
Fuente: **Instituto de Estadística y Cartografía de Andalucía. Movimiento Natural de la Población**
Anual: **2022**

Lugar de residencia	Sexo		
	Hombres	Mujeres	Ambos sexos
Andalucía	32.459	30.508	62.967
Almería	3.511	3.304	6.815
Cádiz	4.462	4.152	8.614
Córdoba	2.852	2.726	5.578
Granada	3.679	3.378	7.057
Huelva	1.927	1.972	3.899
Jaén	2.284	1.968	4.252
Málaga	6.310	5.838	12.148
Sevilla	7.434	7.170	14.604

Metainformación
Para ampliar información: [Movimiento Natural de la Población](#)

Una vez obtenida la tabla, en la parte superior de la misma aparecen una serie de marcadores. Entre ellos destaca el botón para filtrar la consulta, que permite modificar las categorías iniciales seleccionadas por el productor pero además, se pueden generar mapas, mostrar los códigos territoriales, intercambiar ejes, generar gráficos y exportar a distintos formatos además de obtener una url a través de la cual los reutilizadores pueden conectarse a los datos proporcionados en dicha tabla.

- **Garantía del secreto estadístico y protección de datos personales.**

Los controles para garantizar el secreto estadístico y la protección de datos de carácter personal recaen sobre los productores de las estadísticas, que son los que tienen que establecer el nivel de detalle al que desean cargar los datos en BADEA y el nivel de detalle al que desean difundir. BADEA ofrece algunas herramientas para facilitar este trabajo a los productores.

A través del asistente de consultas, es posible ocultar filas y columnas vacías, sustituir los campos vacíos con algún símbolo o letra, por ejemplo ‘se’ (secreto estadístico), poner máscara a los datos que cumplan una determinada condición, por ejemplo, ponerle un * a los valores en los que el tamaño muestral sea inferior a una determinada cantidad, controlar el nivel de detalle al que el



usuario puede llegar a través del filtrado, por ejemplo, en determinadas consultas se puede permitir que el usuario pueda filtrar a nivel de municipio pero no a nivel de sección censal.

Aparte de estas herramientas, el productor lógicamente podrá cargar los datos agrupados a un cierto nivel de manera que no se puedan difundir datos que no cumplan los criterios metodológicos de representatividad o de secreto estadístico, o que vulneren la normativa de protección de datos de carácter personal.

- **Codificación, estándares, nomenclaturas y clasificaciones utilizadas:**

Uno de los objetivos de BADEA es poner a disposición de los productores una serie de clasificaciones estandarizadas que permitirán la comparabilidad de las estadísticas y fomentará la reutilización del dato, al utilizar nomenclaturas comunes.

Entre otros los muchos que se utilizan, se muestra un ejemplo de sistemas de codificación y nomenclaturas utilizados:

- Clasificación Internacional de Enfermedades (La CIE 8, CIE 9 y la CIE 10)
- Clasificación Nacional de Actividades Económicas (CNAE 74, CNAE 93 Rev1, CNAE 2009)
- Clasificación Nacional de Educación (CNED-2000)
- Clasificación Nacional de Ocupaciones (CNO-79, CNO-94 y CNO-2011)
- Clasificación de Bienes y Servicios. COICOP
- Dimensiones territoriales: en general los códigos que dispone el INE para municipios, provincias, comunidades autónomas, países, distritos y secciones censales, además se utilizan otras dimensiones supramunicipales consideradas oficiales, como son la clasificación POTA (Plan de Ordenación del territorio de Andalucía), GDR (Grupos de Desarrollo Rural) y Mapa sanitario (Distritos Sanitarios y Zonas básicas de Salud).

- **Mantenimiento, conservación y actualización:**

Para actualizar la información, el productor tiene que crear los periodos adicionales que va a cargar en el cubo (es el proceso explicado anteriormente de crear uno o varios cubos-periodos, para hacer hueco en el cubo y así poder incorporar los nuevos datos), y a continuación podrá hacer la carga.

En el proceso de carga BADEA establece una serie de controles a la hora de hacer la carga de los datos. Uno de ellos es que sean correctos los códigos que aparecen en el fichero de carga referidos a las distintas dimensiones, por ejemplo si en la dimensión sexo el código 01 representa a los hombres y el 06 a mujeres, la aplicación mostrará un error si en la columna del fichero de carga que hace referencia al sexo aparece el código 02, puesto que 02 es un código inexistente en la dimensión sexo. Por otra parte también controla que no existan registros duplicados en el nuevo fichero de carga, o entre el fichero de carga y lo que había cargado con anterioridad. Cuando la



aplicación encuentra errores en el fichero de carga, lo especifica en el estado de la carga y en el log indicará el tipo de error que ha encontrado.

Por otra parte, los cubos-periodos llevan asociados unos atributos llamados estado del dato, con las siguientes categorías:

A cargar: Cuando se ha creado el cubo-periodo, pero aún no se ha realizado la carga para dicho periodo.

Revisar carga: cuando la carga ha finalizado sin errores, pero los datos están pendientes de revisión por parte del productor. Cuando los datos están en este estado, al ejecutar las consultas los usuarios productores que no tengan esta actividad asignada, no podrán visualizar estos nuevos datos, tampoco se podrán visualizar en internet. Si la carga ha dado errores, el estado de los datos seguirá siendo A cargar.

Aprobado técnico: es el estado que reciben los datos una vez que los productores han dado el visto bueno, estos datos ya si podrán visualizarse por parte de otros productores en consultas de otras actividades que estén tirando de estos cubos, pero al no tener la aprobación de publicación, aún no se podrán ver en la calle, es decir, se podrán visualizar en la intranet, pero no en internet. El aprobado técnico está pensado para que sea un superior jerárquico el que lo aplica una vez revise los datos.

Aprobación de publicación: una vez que los datos han pasado las revisiones anteriores, se le comunica al servicio de difusión (en algunos casos puntuales, los productores tienen permiso para publicar, pero no es lo habitual) para que concedan la aprobación de publicación, en este caso el estado de los datos pasará a ser 'Publicado' y los datos serán visibles también en internet.

Existe un estado adicional que es el de Publicado editable, se ponen los datos en este estado, cuando los datos publicados se quieren modificar, pero se desea que sigan apareciendo en internet mientras los datos nuevos están en proceso de carga y revisión. Una vez terminadas las revisiones, se le da la aprobación de publicación a los datos nuevos y estos sustituirán en internet a los anteriores, esto es muy útil, por ejemplo en el caso de datos provisionales que se quieran sustituir por otros definitivos.



5. PLAN DE DIFUSIÓN

- **Producto:** BADEA.
- **Tipo de resultados y formatos:** La información se ofrece en tablas html, mapas o gráficos, que a su vez se pueden descargar en distintos formatos de software libre y comercial (pdf, Pc-Axis, csv, xls, json, etc). Además es posible consumir la información a través de formatos interoperables.
- **Periodicidad:** Continua.
- **Usuarios:** Actualmente no se realiza ninguna actividad encaminada a evaluar la satisfacción de los usuarios, ni la calidad percibida por ellos, si bien es verdad que a través del formulario de contacto que ofrece el IECA en su página web, o por teléfono, se recogen las sugerencias y se atienden las dudas que puedan surgir.



6. CALIDAD

- **Respecto al productor de los datos:** La gestión en la calidad de los datos corresponde a los productores del dato, por lo tanto se detallará en las memorias técnicas de las distintas actividades adscritas a BADEA.
- **Respecto a los procesos:** en BADEA aún no hay implementado ningún mecanismo para medir el grado de cumplimiento de los procesos de gobierno del dato.
- **Respecto a los resultados:**
 - *Relevancia y utilidad:* al tratarse de una herramienta para almacenar y difundir información, no tiene sentido hablar de relevancia y utilidad en los datos, en este caso, la relevancia y utilidad podría ser lo que hemos descrito como objetivos dentro de esta memoria técnica.
 - *Precisión y confiabilidad:* Trasladando Precisión y confiabilidad al ámbito de BADEA, se podría hablar de si los cambios en el estado de los datos se reflejan de forma inmediata en los datos publicados tanto en intranet como en internet y en este sentido habría que decir que en muy raras ocasiones se ha detectado falta de sincronización de las bases de datos internas con intranet o internet. En cualquier caso no se dispone de un sistema de indicadores o de medición de frecuencia de esta anomalía. Y no se recuerda ningún caso en el que datos sin tener la aprobación de publicación hayan salido publicados en internet, lo cual da bastante confianza a los productores de datos.

Por otra parte, en el menú de BADEA, hay una zona de mantenimiento, encaminada a detectar y corregir fallos en dimensiones, consultas y árboles de consultas... Informa del número de errores que hay en cada una y cuales son.

- *Nivel de estandarización o conformidad:* Se siguen los principios del “Código de Buenas Prácticas de las Estadísticas Europeas”.
- *Esquema de calidad:* Hasta la fecha, no se ha construido ningún esquema de calidad en BADEA.