

Estadísticas de población de Andalucía basadas en registros administrativos

Nota metodológica

Tabla de contenido

1. Introducción.....	2
2. Finalidad y objetivos del estudio.....	3
3. Fuentes de información.....	4
4. Integración, conciliación y tratamiento de la información de base.....	8
4.1. Integración de las fuentes con la BDLPA.....	8
4.2. Conciliación de inconsistencias.....	8
4.3. Incorporación de nuevas variables derivadas del tratamiento de otras fuentes.....	9
5. Productos de difusión derivados del estudio.....	11
5.1. El fichero muestral de microdatos integrados.....	11
5.2. Indicadores estadísticos de población de Andalucía por secciones censales.....	13
5.2.1. Procedimiento de cálculo de los indicadores.....	13
5.2.2. Organización y definición de los indicadores.....	13
5.2.3. Canales de difusión de los indicadores.....	17

1. Introducción

Los registros administrativos constituyen el instrumento mediante el que la Administración deja constancia de los hechos e interacciones que, en el ejercicio de sus competencias, realiza con sus administrados, sean estos personas físicas o jurídicas. La tramitación de cualquier acto administrativo genera información que queda almacenada en el correspondiente registro y esta es susceptible de ser aprovechada para usos estadísticos.

De hecho, las características que presentan los registros administrativos (están descritos legalmente, habitualmente son de obligado cumplimiento, afectan a toda la población en su ámbito, cuentan con formularios para la recolección de datos, etc.) hacen que sean una fuente primordial para la realización de muchas de las operaciones estadísticas que se realizan actualmente.

Pero la explotación con fines estadísticos de la información contenida en los registros administrativos adolece de limitaciones (cambios en la normativa que regula el registro, su origen nunca es estadístico, las definiciones o las unidades empleadas pueden ser distintas de las estadísticas, etc.) que deben ser conocidas para aprovechar al máximo su potencial como fuente estadística.

La importancia del uso de los registros administrativos en el ámbito estadístico queda reflejada en el *Código de Buenas Prácticas de las Estadísticas Europeas*¹, promovido por Eurostat, que cita el uso de fuentes administrativas en varios de sus principios con el fin de limitar la carga de información solicitada a los informantes y mejorar el uso eficiente de los recursos disponibles.

Actualmente son múltiples los ejemplos de aprovechamiento estadístico de registros administrativos que facilitan información esencial sobre determinados fenómenos (estadísticas de movimiento laboral registrado, estadísticas de nacimientos, defunciones y matrimonios, etc.), pero en la mayoría de los casos esta información se procesa y difunde conforme al modelo tradicional de producción estadística basado en una arquitectura operativa de compartimentos estancos, es decir, de manera aislada. Se obtiene así una información útil, pero parcial, que no permite la exploración de las interrelaciones entre distintas dimensiones y ámbitos temáticos.

Para superar esta limitación y facilitar una información más rica que ayude a conocer y entender la realidad desde una perspectiva multidimensional se hace necesario trabajar con información integrada proveniente de distintas fuentes.

Tal enfoque es el que recoge la *Comunicación de la Comisión al Parlamento Europeo y al Consejo sobre el método de elaboración de las estadísticas de la Unión Europea: una visión para la próxima década*². En esta comunicación se reconoce que las necesidades de información de los usuarios relacionadas con fenómenos multidimensionales complejos e interdependientes aumentan continuamente, por lo que para dar respuestas adecuadas a los retos actuales es preciso que los sistemas estadísticos evolucionen hacia un modelo de producción integrada. Es decir, es necesario transformar la arquitectura operativa estadística para acceder y combinar eficientemente la información procedente de distintas fuentes (en

1 <https://ec.europa.eu/eurostat/documents/3859598/5922097/10425-ES-ES.PDF>

2 <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52009DC0404&from=ES>

particular, las fuentes de origen administrativo) con el fin de desarrollar, elaborar y difundir conjuntos de datos estadísticos (y geoespaciales) integrados y coherentes.

Así pues, los registros administrativos y su integración desempeñan un papel determinante para superar los análisis en los que solo se tienen en cuenta a las personas desde un punto de vista unidimensional (estudio de un solo fenómeno) y efectuar la transición a una perspectiva multidimensional (personas que viven, trabajan, estudian, etc.), sin tener que requerir la información directamente a los informantes a través de encuestas específicas, evitándoles así las molestias e inconvenientes que esto conlleva, junto con las limitaciones de la representatividad y la extrapolación de los resultados a la población total.

Actualmente la adopción de este modelo de reutilización e integración de información de base administrativa relativa a las personas se está abordando en el Instituto de Estadística y Cartografía de Andalucía (IECA) gracias, por una parte, a la mejora en los procedimientos informáticos que permiten almacenar, enlazar y procesar cada vez más información a un coste operacional cada vez menor y, por otra, al desarrollo y consolidación de la infraestructura de información estadística “**Base de Datos Longitudinal de Población de Andalucía**” (BDLPA). Las fuentes principales que nutren este sistema de información provienen de las inscripciones en los padrones municipales de Andalucía, los boletines estadísticos del Movimiento Natural de Población y los censos de población y viviendas. La integración de esta información permite la incorporación de variables recopiladas en una fuente a las de otra, que o bien no disponía de dichas variables o bien las tenía recogidas con un menor nivel de detalle.

Asimismo, el desarrollo de este sistema se ha fortalecido por su adscripción a las iniciativas³ de la Consejería de Transformación Económica, Industria, Conocimiento y Universidades en materia de estrategia para la integración de la información administrativa de base poblacional disponible en los diferentes organismos de la Administración autonómica. El objeto de esta estrategia es el análisis conjunto de la información y su posterior reutilización desde un punto de vista estadístico, de manera que permita profundizar en el conocimiento de la realidad socioeconómica de Andalucía, evaluar el éxito de las políticas públicas a partir de datos contrastados y ofrecer las soluciones sectoriales más eficaces. La estrategia reconoce que la función de integración debe ser desempeñada por el IECA, como órgano técnico con la capacidad de acceso a las bases de datos administrativos y con competencia para fijar las claves identificativas necesarias que permitan la fusión de toda la información poblacional a nivel individual en el marco de la BDLPA.

Así pues, este modelo de producción basado en la reutilización de información recopilada en registros administrativos sectoriales y su integración en la BDLPA va a permitir generar una nueva gama de productos estadísticos de perspectiva transversal, en los que se prioriza la investigación de estructuras y relaciones de asociación entre determinadas características en un momento dado, ampliando y complementando así la producción estadística clásica con un coste menor que el que supondría diseñar nuevos circuitos de recogida de información definidos ad hoc.

2. Finalidad y objetivos del estudio

En el contexto descrito se enmarca la actividad “**Estadísticas de población de Andalucía basadas en registros administrativos**” (EPABRA). Este estudio es el resultado de un proyecto

3 <https://juntadeandalucia.es/organismos/consejo/sesion/detalle/182495.html>

experimental de integración y conciliación de fuentes cuya finalidad es la interconexión y el aprovechamiento estadístico de la información procedente de varios registros administrativos a nivel individual que, partiendo de la BDLPA como infraestructura vertebradora del proyecto, aporte información sobre las distintas situaciones individuales recogidas en ellos: demanda de empleo, afiliación a la Seguridad Social o a MUFACE, y percepción de pensiones contributivas, todo ello a nivel individual y con cobertura exhaustiva.

Para ello, el objetivo general que se fija en esta actividad es fusionar coherentemente información a nivel individual proveniente de diversos ficheros de origen administrativo para aportar un enfoque multidimensional en el estudio de la situación individual de las personas residentes en Andalucía.

De manera específica, la actividad “Estadísticas de población de Andalucía basadas en registros administrativos” persigue:

- a) Construir con periodicidad anual un fichero maestro o marco poblacional integrado, con carácter *experimental*, que combine de forma coherente información procedente de las distintas bases de datos poblacionales de origen administrativo a las que tiene acceso el IECA, con fecha 1 de enero del año de referencia.
- b) Propagar la geocodificación de la información individual contenida en la BDLPA a las variables temáticas aportadas por las fuentes administrativas que se integren.
- c) Ampliar y mejorar la información contribuida por las distintas fuentes de forma aislada mediante la reutilización de información ya existente.

3. Fuentes de información

El eje central de la información que se ha recopilado y combinado en este proyecto para caracterizar la situación individual de la persona es la BDLPA, por lo que la población de referencia está constituida por los residentes en Andalucía (empadronados en algún municipio andaluz) en una fecha concreta (1 de enero del año de referencia). Por tal motivo pueden presentarse diferencias con respecto a los volúmenes poblacionales recogidos en los diferentes ficheros administrativos sectoriales, posiblemente debidas a la inscripción en los mismos de residentes fuera de la Comunidad Autónoma o de personas que a la fecha de referencia no han actualizado aún su situación de empadronamiento en algún municipio andaluz.

A la información individual de la BDLPA se le ha incorporado una serie de características presentes en el resto de ficheros administrativos utilizados. Concretamente, las fuentes de información empleadas en la construcción de este fichero marco poblacional integrado han sido:

Fuentes de información básica personal

- Base de Datos Longitudinal de Población de Andalucía.
 - Ficheros de stocks y flujos padronales.
 - Movimiento Natural de la Población (MNP).
 - Censos de población y viviendas de los años 1991, 2001 y 2011.
- Ficheros de rendimientos académicos y de avance de matriculaciones en títulos universitarios oficiales suministrados por las universidades públicas andaluzas.
- Ficheros de matriculaciones y resultados en enseñanzas regladas competencia de la Consejería Educación y Deporte.

Fuentes de información temática personal

- Fichero de afiliaciones en alta laboral y cuentas de cotización a la Seguridad Social.
- Fichero de titulares registrados en la Mutualidad General de Funcionarios Civiles del Estado (MUFACE).
- Registro de personas demandantes de empleo inscritas en las oficinas del Servicio Andaluz de Empleo.
- Registro de Prestaciones de la Seguridad Social.

Base de Datos Longitudinal de Población de Andalucía (BDLPA)

La información referida a las características demográficas básicas, así como la información residencial de cada persona, procede de la Base de Datos Longitudinal de Población de Andalucía.

Esta base de datos se origina a partir de la integración de información de stocks, flujos y variaciones procedente del sistema de coordinación padronal, que junto con los datos provenientes de los Registros Civiles referidos a nacimientos, defunciones y matrimonios (MNP) y la contenida en los censos de población y viviendas, dan lugar a un marco integrado longitudinal para las estadísticas de población y territorio en Andalucía.

La información almacenada en la BDLPA se ciñe a un modelo orientado al seguimiento longitudinal de las personas y sus atributos, con objeto de constituir biografías o itinerarios vitales individuales que informen de los episodios y cambios de estado en sus características demográficas y residenciales.

Para la construcción del fichero marco poblacional de la actividad “Estadísticas de población de Andalucía basadas en registros administrativos”, se extrae del fichero longitudinal de episodios vitales y residenciales derivado de la BDLPA un corte transversal con fecha 1 de enero del año de referencia. Este corte sirve como marco de información básica para la población residente en Andalucía en esa fecha de referencia, y sobre él se ha integrado a nivel individual la información temática de los ficheros administrativos sectoriales empleados.

La información básica, que por defecto prevalece frente a la aportada por las fuentes administrativas temáticas, comprende, junto con la identificación del individuo, las variables sexo, año de nacimiento, nacionalidad, municipio de nacimiento y año de llegada al municipio de residencia, así como la información de localización territorial del lugar de residencia.

No obstante, el criterio de superioridad jerárquica de la BDLPA frente a las restantes fuentes se ha alterado a favor de las fuentes temáticas en aquellos casos en que ha sido necesaria tal actuación para la conciliación de las inconsistencias que hayan podido surgir en la información integrada a nivel individual.

Fichero de afiliaciones en alta laboral y cuentas de cotización a la Seguridad Social

La información referente a las personas afiliadas a la Seguridad Social en alta laboral que trabajan en Andalucía proviene del fichero que suministra trimestralmente la Tesorería General de la Seguridad Social al IECA con fecha de referencia último día del trimestre.

En dicho fichero se encuentran los datos de afiliaciones y cuentas de cotización de la Seguridad Social de Andalucía, que aportan información relativa a los trabajadores afiliados en los

distintos regímenes del Sistema de la Seguridad Social en situación de alta laboral y situaciones asimiladas al alta con fecha de referencia el último día laborable del trimestre del envío.

Los afiliados a la Seguridad Social en alta laboral son aquellos que están dados de alta en alguno de los regímenes del Sistema de la Seguridad Social, tanto por cuenta ajena como por cuenta propia. El número de afiliaciones a los distintos regímenes no se corresponde necesariamente con el número de trabajadores, ya que un trabajador puede encontrarse en alta laboral (afiliado a la Seguridad Social) en dos regímenes distintos a la vez (pluriactividad) o en el mismo régimen pero en distintas actividades (pluriempleo). Estos se contabilizarán como afiliados a la Seguridad Social doble o múltiplemente, según las situaciones de afiliación que presenten. En definitiva, el número de afiliaciones en alta laboral será un número mayor al número de trabajadores en alta a la Seguridad Social.

Es necesario remarcar que el fichero de afiliaciones contiene las personas afiliadas que trabajan en Andalucía, lo cual implica que dicho fichero puede contener personas no residentes en la Comunidad Autónoma y que, por consiguiente, no van a formar parte del marco poblacional integrado. Recíprocamente, los residentes en Andalucía afiliados a la Seguridad Social que trabajan fuera de la Comunidad Autónoma no constan en el fichero suministrado, por lo que forman parte del marco poblacional integrado pero no aparecen en él en situación de afiliación en alta laboral.

La información temática aportada por este fichero al marco poblacional integrado comprende el tipo de contrato, el régimen de cotización, la actividad económica y el número de afiliaciones de cada persona afiliada.

Fichero de titulares registrados en la Mutualidad General de Funcionarios Civiles del Estado (MUFACE).

La información se refiere a mutualistas titulares de la Mutualidad General de Funcionarios Civiles del Estado (MUFACE) residentes en Andalucía. El fichero correspondiente es enviado al IECA trimestralmente con fecha de referencia el último día del trimestre.

Dado que el fichero se refiere exclusivamente a mutualistas titulares, no están recogidos en el mismo otros beneficiarios como pueden ser las personas perceptoras de prestaciones de viudedad u orfandad.

Los titulares incluidos en el fichero son clasificables en función de su situación administrativa en tres grandes grupos asimilables a la situación de activos, situación de excedencia o suspensión y situación de jubilación o incapacidad permanente. La información de la situación administrativa, así como el sector de actividad asociado, forman parte de los datos individuales aportados al marco poblacional integrado.

Fichero de personas demandantes de empleo inscritas en las oficinas del Servicio Andaluz de Empleo (SAE).

El IECA recibe mensualmente, con fecha de referencia último día laborable del mes, un fichero que contiene todas las demandas registradas en alguna de las oficinas del Servicio Andaluz de Empleo.

Se entiende como demandantes a las personas que se inscriben en los servicios públicos de empleo para la búsqueda de un empleo o para mejorar el que ya poseen (intermediación

laboral), para recibir otros servicios ajenos al empleo (orientación para el autoempleo, formación ocupacional, etc.) o por la obligatoriedad de inscribirse para percibir una prestación contributiva o un subsidio. Estos datos, procedentes del fichero estadístico, son continuamente actualizados por las acciones de los gestores o periódicamente con cruces de registros públicos de trabajadores afiliados a los regímenes de la Seguridad Social, pensionistas, estudiantes, militares profesionales, etc.; asimismo se actualizan con el resto de la información del Servicio Público de Empleo Estatal (SEPE) relativa a prestaciones por desempleo, formación profesional para el empleo o contratos.

Al igual que en el caso del fichero de afiliaciones en alta laboral a la Seguridad Social, una fracción del colectivo de personas inscritas en el fichero de demandantes de empleo no consta en la información de empadronamiento disponible en la BDLPA. Por consiguiente, se trata de personas demandantes que no residen en Andalucía o que, a la fecha de referencia, no han actualizado aún su situación de empadronamiento en algún municipio andaluz, por lo que no forman parte del marco poblacional integrado.

Entre los demandantes es de interés el estudio del colectivo de los que son clasificados como parados registrados, que son aquellas personas cuya demanda de empleo está pendiente de satisfacer el último día del mes de referencia y cumple con las condiciones establecidas por la Orden Ministerial, de 11 de marzo de 1985, para ser consideradas paradas registradas. Esta consideración viene determinada por la no pertenencia a los denominados colectivos de exclusión: demandantes de otros servicios, demandantes ocupados, trabajadores eventuales agrarios beneficiarios del subsidio especial por desempleo, otros demandantes de empleo no ocupados y otros demandantes excluidos del paro registrado.

El fichero de personas demandantes de empleo inscritas en las oficinas del SAE suministra al marco poblacional integrado información individual relativa al tipo o clasificación de la demanda, la ocupación preferente solicitada, el último sector de actividad en el que se ha trabajado, el ámbito territorial de búsqueda de empleo y el tiempo que lleva inscrita la demanda.

Registro de prestaciones de la Seguridad Social.

El IECA recibe anualmente un fichero extraído del Registro de prestaciones de la Seguridad Social, con perceptores de pensiones contributivas que, o bien son gestionadas por las direcciones provinciales de la Seguridad Social en Andalucía, o bien los perceptores residen en Andalucía aunque sus pensiones son gestionadas por otras direcciones provinciales. Este fichero contiene las prestaciones en alta a 31 de diciembre y las prestaciones dadas de baja durante el año de referencia de los datos.

Los datos recogidos hacen referencia a las pensiones contributivas abonadas por el Sistema de la Seguridad Social en función de la contingencia protegida: incapacidad permanente, jubilación, viudedad, orfandad y favor de familiares. Esto implica que desde la perspectiva analítica del pensionista debe tenerse presente que una misma persona puede ser perceptora de varias pensiones correspondientes a contingencias que no son mutuamente excluyentes, como por ejemplo en los casos en que concurren la pensión de jubilación junto con la de viudedad.

La información aportada por este fichero de pensiones contributivas al marco poblacional integrado consiste en número de pensiones percibidas por la persona pensionista, la tipología de las mismas y el importe bruto mensual acumulado.

4. Integración, conciliación y tratamiento de la información de base

4.1. Integración de las fuentes con la BDLPA

El núcleo de información personal del sistema BDLPA permite identificar de manera individual a las personas residentes en el territorio andaluz. Los procedimientos de carga y mantenimiento que posibilitan esta identificación automática se han extendido a la información personal de los ficheros correspondientes a afiliados a la Seguridad Social, pensionistas contributivos, titulares de MUFACE y movimiento laboral registrado por el SAE (en particular, demandantes de empleo), creando de manera unitaria una base de datos conjunta que incluye solamente datos personales seudonimizados. Es decir, aunque la información de trabajo específica de cada fuente no está integrada, sí lo está la información identificativa seudonimizada referente a las personas en cada fichero.

Este procedimiento automático de enlace de datos identificativos se encuentra actualmente en desarrollo evolutivo para extender su ámbito de aplicación a otros ficheros administrativos de base poblacional que próximamente se incorporarán a la BDLPA. Para el tratamiento de esos ficheros administrativos aún no integrados, se han replicado manualmente los procedimientos automatizados de identificación y enlace de datos personales para explotar como información de contraste e imputación los datos contenidos en ellos.

El objetivo que se persigue con la construcción del fichero marco poblacional es la integración efectiva de la información temática de cada uno de los ficheros a nivel de registro individual, usando como eje de integración la información personal identificativa seudonimizada derivada del procedimiento de enlace. No obstante, hay que tener en cuenta que los procedimientos de enlace e integración siempre se ven afectados en cierta medida por la incertidumbre derivada de las incoherencias existentes en la información personal recogida en las distintas fuentes, así como por las diferencias en la cobertura de las distintas fuentes, ya que no todas las personas se encuentran de manera simultánea en todos los ficheros.

4.2. Conciliación de inconsistencias

El resultado inicial del proceso de identificación y enlace genera inevitablemente la aparición de conflictos de consistencia en la información temática integrada, debidos a que a un mismo individuo se le asocian características procedentes de distintas fuentes independientes que pueden representar situaciones aparentemente incoherentes o, incluso, contradictorias.

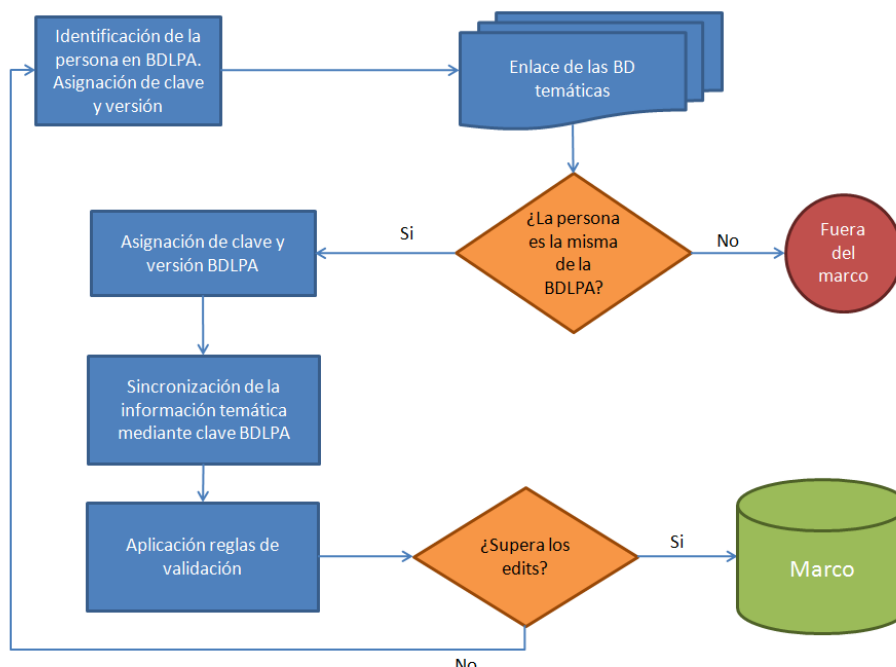
El origen de los conflictos de consistencia detectados ha tenido, principalmente, dos motivos:

- Incoherencias en la información común entre los distintos ficheros.
- Diferencias metodológicas y de calendario en la extracción de la información de cada uno de los ficheros.

Para la detección de las inconsistencias se han creado reglas de validación o “edits” a las que han sido sometidos los datos durante el proceso de integración de las fuentes. Estos edits se clasifican en dos categorías:

- Errores: combinaciones de valores imposibles o contrarios a la normativa reguladora de los registros administrativos vigente en la fecha de referencia. Han sido sometidos a actuaciones correctivas para conciliar los valores en conflicto, consistentes en la búsqueda de una combinación de valores, entre la información individual disponible de la persona afectada por la inconsistencia, que permita corregir el error. En caso de no hallar combinaciones alternativas coherentes, se ha procedido a imputar un valor válido.
- Avisos: combinaciones de valores inusuales en las variables que pueden deberse a errores presentes en alguna de las fuentes o a excepciones contempladas en los marcos normativos reguladores. Los avisos no justificados por excepciones normativas han sido sometidos a revisión y, en su caso, se han corregido mediante la búsqueda de otra combinación de valores en la información disponible de la persona que desactiven el aviso.

Esquemáticamente, el proceso seguido para la inicialización del fichero marco poblacional, su enlace a nivel individual con las fuentes administrativas sectoriales, el posterior tratamiento de las inconsistencias detectadas en la información integrada y la consolidación final del marco poblacional integrado experimental se representa en la siguiente figura:



4.3. Incorporación de nuevas variables derivadas del tratamiento de otras fuentes

El fichero marco poblacional integrado se ha enriquecido con la incorporación de cuatro variables que incrementan la potencia analítica del conjunto de datos: el máximo nivel de estudios alcanzado, los estudios oficiales en curso a la fecha de referencia del marco

poblacional, la identificación del hogar/vivienda familiar de residencia y el grado de urbanización de la unidad territorial de residencia (delimitada en términos de celda cuadrada de 1 km de lado).

Nivel de estudios alcanzado y estudios en curso

La información padronal referente al nivel de estudios no ha sido tradicionalmente tenida en cuenta para su explotación por los sistemas estadísticos. Este hecho ha implicado que no se le haya prestado una atención especial a la calidad de esta información, por lo que actualmente sigue presentando ciertas deficiencias que la invalidan para su aprovechamiento estadístico directo. No obstante, se ha constatado cierta mejoría en los últimos años debida sobre todo a la incorporación de datos procedentes del Ministerio competente en materia de educación.

Con objeto de mejorar la calidad e incorporar la información padronal sobre el nivel de estudios alcanzado al marco poblacional integrado, así como la información relativa a los estudios en curso, se ha empleado para su corrección o, en su caso, incorporación, la información relacionada con el nivel formativo individual que se encuentra disponible en el resto de fuentes administrativas implicadas en este proyecto de integración, así como la disponible en otras fuentes que no han contribuido con la incorporación de sus datos en el marco integrado, pero han desempeñado un rol de información auxiliar para el contraste y el control de la información integrada (censos de población y viviendas, contratos y colocaciones registradas en el Servicio Andaluz de Empleo, ficheros de rendimientos académicos y de avance de matriculaciones en títulos universitarios oficiales suministrados por las universidades públicas andaluzas, ficheros de matriculaciones y resultados en enseñanzas regladas competencia de la Consejería Educación y Deporte, etc.).

Para ello se ha contrastado la información y se han conciliado las posibles incoherencias que han surgido acerca del nivel de estudios reportado por fuentes administrativas distintas, con base en una hipótesis central de trabajo: el nivel de estudios no se pierde con el tiempo, es decir, siempre puede ir a más pero nunca a menos. Como resultado, la variable nivel de estudios alcanzado se ha incorporado en el marco poblacional integrado con una categorización de 10 niveles. Además, la variable clasifica como “No procede” a las personas para las que no es obligatorio especificar en el Padrón el nivel de estudios poseído (los menores de 16 años). Del mismo modo se ha establecido una clasificación de estudios en curso formada por 10 ítems, aplicable también a la población con 16 años o más.

Identificación del hogar/vivienda familiar

Se ha diseñado un procedimiento orientado a la identificación de los hogares/viviendas en los que residen los individuos que componen el stock de población extraído de la BDLPA.

En esta base de datos se encuentran almacenados, para cada persona, los datos relativos a la dirección postal del domicilio de residencia (denominado hueco en la operativa de BDLPA), así como los relativos a la hoja padronal. La combinación de esta información padronal se ha empleado para derivar un constructo o entidad equiparable a la vivienda/hogar, de forma que quede determinado por una clave identificativa única que se asigna a todos los individuos que son miembros del mismo. Como resultado, las personas que se encuentran recogidas en el marco poblacional integrado disponen de una clave identificativa de vivienda/hogar y aquellos individuos que compartan dicha clave son considerados miembros de la misma vivienda.

Las relaciones biunívocas entre el hueco y la hoja padronal, es decir, aquellas en las que a cada hueco le corresponde una única hoja padronal y en las que a la vez esta no se encuentra asociada a ningún otro hueco, dan lugar a la identificación directa de un constructo vivienda/hogar.

Desafortunadamente existe una proporción de casos en los que esta relación biunívoca no se da. En tales situaciones se ha recurrido a información auxiliar para reducir la cardinalidad de dichas relaciones. Como resultado se ha alcanzado una identificación segura de la vivienda/hogar para prácticamente el 90% de la población recogida en el marco poblacional integrado. El 10% restante está asignado a viviendas familiares que tienen un mayor grado de incertidumbre en cuanto a su delimitación, o residen en establecimientos colectivos.

Grado de urbanización de la celda de residencia

El grado de urbanización es una clasificación que busca caracterizar la intensidad del asentamiento de la población residente en un territorio usando las celdas cuadradas de 1 km² de la malla estadística de referencia. La información de localización territorial del lugar de residencia de las personas contenidas en la BDLPA ha permitido asignar a éstas el grado de urbanización con el que se clasificaron las respectivas celdas de residencia, con fecha de referencia del año anterior a la fecha correspondiente de la publicación de EPABRA.

A partir de la densidad de población se definen distintas categorías de áreas basándose en criterios de contigüidad geográfica, densidad y umbrales de población, según la metodología descrita por Eurostat⁴ para la asignación de las tipologías territoriales en las celdas de la malla estadística en tres tipos⁵:

- Centros urbanos. Conglomerado de celdas contiguas de 1 km² con una densidad de al menos 1.500 habitantes por km² y que conjuntamente agrupen un mínimo de población de 50.000 habitantes.
- Agrupaciones urbanas. Conglomerado de celdas contiguas de 1 km² con una densidad de al menos 300 habitantes por km² y que conjuntamente agrupen un mínimo de población de 5.000 habitantes.
- Celdas de malla rurales. El resto de celdas no clasificadas en alguno de los apartados anteriores.

5. Productos de difusión derivados del estudio

5.1. El fichero muestral de microdatos integrados

A partir del marco poblacional integrado, y tras su depuración estadística para suprimir inconsistencias, se ha elaborado un primer producto de difusión consistente en una serie de ficheros muestrales de microdatos integrados, con fecha 1 de enero del correspondiente año de referencia, preparados para suministrar a los usuarios la información necesaria que puedan requerir en sus respectivos proyectos de análisis e investigación.

Del fichero marco poblacional integrado se ha extraído una muestra de viviendas familiares (no se han considerado elegibles las viviendas calificadas como establecimientos colectivos). Cada

4 http://www.juntadeandalucia.es/institutodeestadisticaycartografia/gradourbanizacion/metodologia/Notas_Metodologicas.pdf

5 Artículo 4 ter, punto 2 del Reglamento (UE) 2017/2391.

vivienda seleccionada aporta a la muestra todas las personas que son miembros de la misma. El procedimiento de selección de las viviendas se ha realizado mediante muestreo estratificado utilizando como estratos las zonas geográficas resultantes de delimitar en cada provincia las unidades territoriales formadas por los municipios pertenecientes a cada uno de los seis niveles de tamaño municipal contemplados: capital de provincia; municipios con más de 100.000 habitantes (no capital de provincia); municipios de 50.001 a 100.000 habitantes; municipios de 20.001 a 50.000 habitantes; municipios de 10.001 a 20.000 habitantes; y municipios de hasta 10.000 habitantes. En cada uno de los estratos se ha realizado un muestreo aleatorio simple de las unidades muestrales, con una fracción de muestreo del 10%, aproximadamente, en cada estrato.

En la siguiente tabla se presenta la relación de variables contenidas en el fichero muestral de microdatos integrados. Esta relación se encuentra en continua ampliación, conforme se incorporan nuevas fuentes administrativas satelitales en la BDLPA, por lo que en futuras ediciones se podrá incrementar su contenido:

Variable	Contenido
B. Información básica	
B01	Año de referencia de la información (a 1 de enero)
B02	Clave numérica identificadora de persona
B03	Año de nacimiento
B04	Sexo
B05	Nacionalidad
B06	Provincia de residencia
B07	Tamaño del municipio de residencia
B08	Grado de urbanización de la celda de 1 km de residencia
B09	Nivel educativo
B10	Clave numérica identificadora de vivienda
B11	Relación entre el lugar de nacimiento y el de residencia
B12	Año de llegada al municipio de residencia
B13	Estudios en curso
P. Información de pensiones contributivas	
P01	Número de prestaciones percibidas
P02	Tipología de la prestación
P03	Importe mensual bruto acumulado
A. Información de afiliación a la Seguridad Social	
A01	Número de afiliaciones
A02	Régimen de afiliación a la Seguridad Social
A03	Actividad económica del régimen o de la cuenta de cotización
A04	Tipo de contrato
D. Información de demandantes de empleo	
D01	Clasificación de la demanda
D02	Ocupación preferente en la que el demandante desea trabajar
D03	Tiempo que lleva inscrita la demanda
D04	Ámbito territorial de búsqueda de empleo
D05	Actividad económica en la que ha trabajado anteriormente
M. Información de titulares de MUFACE	
M01	Situación administrativa
M02	Actividad económica en la que trabaja (o ha trabajado)

Para profundizar en el contenido y los detalles de los ficheros muestrales de microdatos integrados de las Estadísticas de población de Andalucía basadas en registros administrativos, se recomienda acceder al área de descarga de microdatos de esta actividad⁶ y consultar el documento “Descripción del contenido de los microdatos” correspondiente a la fecha de referencia de interés.

5.2. Indicadores estadísticos de población de Andalucía por secciones censales

5.2.1. Procedimiento de cálculo de los indicadores

El desarrollo del producto de difusión “Indicadores estadísticos de población de Andalucía por secciones censales” se ha basado en la geocodificación a nivel de sección censal de la residencia padronal de toda la población de Andalucía contenida en el marco poblacional integrado del proyecto EPABRA, usando para ello las coordenadas espaciales asociadas a sus datos de dirección postal horizontal. La información individual de esta población geocodificada se ha agregado en cada sección censal para generar un conjunto de veintiuna tasas e indicadores demográficos y socioeconómicos. El ámbito temporal de estos indicadores es de periodicidad anual y se corresponde con el de las fechas de referencia del proyecto EPABRA. Su ámbito territorial está formado por los sistemas zonales correspondientes a los seccionados censales de Andalucía vigentes en cada una de dichas fechas de referencia.

5.2.1.1. Acotación de los valores extremos de los indicadores

Para mitigar el efecto distorsionador de los posibles valores extremos presentes entre los valores cuantitativos obtenidos directamente del cálculo de los indicadores, se han establecido unas cotas superior e inferior para cada uno de ellos. Así, cuando el valor que toma un indicador en una determinada sección censal está por debajo de la cota inferior o por encima de la cota superior, se le asigna en esa sección censal el valor de dicha cota.

Para el cálculo de esas cotas se han ordenado las secciones censales según el valor del indicador correspondiente y se ha determinado el percentil 0,5 (cota inferior) y el percentil 99,5 (cota superior) tomando como unidad la sección censal y como variable el indicador correspondiente.

5.2.1.2. Umbral de riesgo de revelación de los indicadores

Con el fin de reducir el riesgo de revelación, se ha fijado un umbral mínimo de 150 habitantes en las secciones censales para publicar los valores correspondientes de los indicadores. Así, si en alguna sección censal se ve comprometida la preservación del principio del secreto estadístico según este criterio, se suprime el valor de todos sus indicadores y tasas.

5.2.2. Organización y definición de los indicadores

Las tasas e indicadores poblacionales calculados para cada sección censal se encuentran organizadas en cuatro áreas temáticas:

6 <http://www.juntadeandalucia.es/institutodeestadisticaycartografia/descarga/poblacion/poblacion.htm>

- Indicadores demográficos
- Indicadores de pensiones contributivas
- Indicadores de relación con la actividad económica
- Indicadores de hogares

El área de **Indicadores demográficos** está formada por los indicadores Edad media de la población, Edad mediana de la población, Porcentaje de menores de edad, Porcentaje de personas mayores, Índice de envejecimiento, Índice de ancianidad, Índice de dependencia y Densidad nocturna de población. Sus definiciones y unidades de medida son las siguientes:

Edad media de la población

Unidad de medida: Años.

Es la media aritmética de las edades de la población objeto de estudio.

Edad mediana de la población

Unidad de medida: Años.

Es la mediana de las edades de la población objeto de estudio.

Porcentaje de menores de edad

Unidad de medida: %.

Representa el número de personas menores de 18 años por cada cien personas de la población total.

Porcentaje de personas mayores

Unidad de medida: %.

Representa el número de personas de 65 años o mas por cada cien personas de la población total.

Índice de envejecimiento

Unidad de medida: %.

Representa el número de personas de 65 y más años por cada cien personas menores de 16 años.

Índice de ancianidad

Unidad de medida: %.

Representa el número de personas de 85 y más años por cada cien personas de 65 y más años.

Índice de dependencia

Unidad de medida: %.

Representa el número de personas menores de 16 años o mayores de 64 por cada 100 personas en edad de actividad laboral (de 16 a 64 años).

Densidad nocturna de población

Unidad de medida: Habitantes por kilómetro cuadrado (hab/km²).

Representa el número de personas residentes por cada kilómetro cuadrado de la superficie habitada de la sección censal.

Se considera superficie habitada de la sección censal la que se solapa con las cuadrículas de 250 m de lado de la malla de población generada en la actividad Distribución espacial de la población en Andalucía.⁷

El área de **Indicadores de pensiones contributivas** está formada por los indicadores Porcentaje de pensionistas de jubilación con pensión inferior al 60% del importe mediano de las pensiones de jubilación, Índice de cobertura de las pensiones contributivas en los mayores, Ratio de afiliados a la Seguridad Social por pensionista y Mediana del importe bruto de las pensiones contributivas. Sus definiciones y unidades de medida son las siguientes:

Porcentaje de pensionistas de jubilación con pensión inferior al 60% del importe mediano de las pensiones de jubilación

Unidad de medida: %.

Representa el número de perceptores de pensiones contributivas de jubilación con un importe inferior al 60% de la mediana de los importes de las pensiones de jubilación de Andalucía por cada cien perceptores de pensiones contributivas de jubilación.

Índice de cobertura de las pensiones contributivas en los mayores

Unidad de medida: %.

Representa el número de personas de 65 y más años que perciben alguna pensión contributiva por cada cien personas de 65 y más años.

Ratio de afiliados a la Seguridad Social por pensionista

Unidad de medida: %.

Representa el número de personas afiliadas a la Seguridad Social en alta laboral por cada cien perceptores de pensiones contributivas.

Mediana del importe bruto de las pensiones contributivas

Unidad de medida: euros (€).

Es la mediana de los importes brutos de las pensiones contributivas percibidas por los pensionistas.

El área de **Indicadores de relación con la actividad económica** está formada por los indicadores Tasa de actividad registrada, Tasa de empleo registrado, Tasa de desempleo registrado, Tasa de desempleo registrado de larga duración y Porcentaje de parados registrados de larga duración. Para una correcta interpretación de sus definiciones, ha de tenerse en cuenta que:

- Se consideran personas registradas como empleadas las afiliadas a la Seguridad Social en alta laboral y los mutualistas activos de MUFACE.
- Se consideran personas registradas como desempleadas los demandantes de empleo no ocupados.
- Se consideran personas con actividad registrada las personas registradas como empleadas o desempleadas.

Las definiciones de las tasas e indicadores y sus unidades de medida son las siguientes:

7 <http://www.juntadeandalucia.es/institutodeestadisticaycartografia/distribucionpob/index.htm>

Tasa de actividad registrada

Unidad de medida: %.

Representa el número de personas registradas como empleadas o desempleadas por cada cien personas entre 16 y 64 años.

Tasa de empleo registrado

Unidad de medida: %.

Representa el número de personas registradas como empleadas por cada cien personas entre 16 y 64 años.

Tasa de desempleo registrado

Unidad de medida: %.

Representa el número de personas registradas como desempleadas por cada cien personas con actividad registrada.

Tasa de desempleo registrado de larga duración

Unidad de medida: %.

Representa el número de personas registradas como desempleadas durante un periodo superior a un año por cada cien personas con actividad registrada.

Porcentaje de parados registrados de larga duración

Unidad de medida: %.

Representa el número de personas demandantes de empleo en situación de paro registrado durante un periodo superior a un año por cada cien personas demandantes de empleo en situación de paro registrado.

Por último, el área de **Indicadores de hogares** está formada por los indicadores Porcentaje de población en hogares unipersonales, Índice de soledad, Porcentaje de hogares unipersonales y Tamaño medio del hogar. Sus definiciones y unidades de medida son las siguientes:

Porcentaje de población en hogares unipersonales

Unidad de medida: %.

Representa el número de personas que viven solas por cada cien personas de la población total.

Índice de soledad

Unidad de medida: %.

Representa el número de personas de 65 y más años que viven solas por cada cien personas de 65 y más años.

Porcentaje de hogares unipersonales

Unidad de medida: %.

Representa el número de hogares unipersonales por cada cien hogares.

Tamaño medio del hogar

Unidad de medida: personas.

Es la media aritmética de los tamaños, o número de personas integrantes, de los hogares.

5.2.3. Canales de difusión de los indicadores

La difusión de los Indicadores estadísticos de población de Andalucía por secciones censales se ha organizado a través de cuatro canales:

- Banco de datos estadísticos BADEA
- Visor web de mapas
- Servicios interoperables WMS y WFS
- Conjuntos de datos espaciales descargables

5.2.3.1. Banco de datos estadísticos BADEA

Las consultas publicadas en el Banco de datos estadísticos BADEA ofrecen el valor calculado de cada uno de los indicadores por secciones censales. La organización de las consultas se ha estructurado por años de referencia de los datos y conforme a las áreas temáticas de indicadores referidas en el apartado 5.2.2. En aquellas secciones en las que se ve comprometida la preservación del secreto estadístico, el valor de los indicadores se ha sustituido por los caracteres “se” (secreto estadístico).

5.2.3.2. Visor web de mapas y servicio interoperable WMS

Se han generado y desplegado un servicio de mapas WMS y un visor web de mapas que dan soporte a la representación y visualización espacial de los Indicadores estadísticos de población de Andalucía por secciones censales para cada uno de los años de referencia de los datos producidos en el proyecto EPABRA.

Tanto en el visor como en el servicio WMS se ofrecen los valores calculados para el conjunto de indicadores mediante consulta espacial por sección censal. A su vez, la visualización cartográfica de los datos se realiza mediante una representación por coropletas con rampa de color temático (cada área temática de indicadores se caracteriza por una rampa de color específica). Esta representación se basa en la clasificación de los valores de las tasas e indicadores en siete intervalos. Se añade a esta clasificación una categoría más para representar aquellas secciones censales en las que se ve comprometida la preservación del secreto estadístico.

Los extremos de los intervalos de clase se mantienen fijos para todos los años de referencia de los indicadores, a fin de reducir los sesgos en la comparación temporal de las representaciones cartográficas. Para tal fin, los extremos se han calculado a partir de los datos con fecha de referencia 1 de enero de 2016. El procedimiento empleado para su cálculo se ha basado en el método de clasificación por rupturas naturales de Jenks. Este es un método de agrupación de datos en clases basado en el algoritmo de Fisher de optimización de la varianza. Este algoritmo busca la mejor estructura de agrupamiento de tal forma que se minimice la varianza de los datos en cada clase y se maximice la varianza entre clases.

5.2.3.3. Servicio interoperable WFS y conjuntos de datos espaciales descargables

Se ofrece también la posibilidad de consumir los datos espaciales en clientes SIG de escritorio mediante un servicio interoperable WFS o mediante la descarga directa de un conjunto de datos espaciales en formato SHP. En ambos casos, la tabla de atributos que acompaña a los

datos espaciales contiene el valor calculado de los indicadores, aunque dicho valor se almacena como nulo en aquellas secciones censales en las que se ve comprometida la preservación del secreto estadístico. La correspondencia entre la denominación del indicador y el nombre de la columna de la tabla de atributos en la que se almacenan sus datos es la siguiente:

Columna	Denominación del indicador
Indicadores demográficos	
I15	Edad media de la población
I2	Edad mediana de la población
I16	Porcentaje de menores de edad
I17	Porcentaje de personas mayores
I1	Índice de envejecimiento
I4	Índice de ancianidad
I3	Índice de dependencia
I22	Densidad nocturna de población
Indicadores de pensiones contributivas	
I7	Porcentaje de pensionistas de jubilación con pensión inferior al 60% del importe mediano de las pensiones de jubilación
I8	Índice de cobertura de las pensiones contributivas en los mayores
I9	Ratio de afiliados a la Seguridad Social por pensionista
I20	Mediana del importe bruto de las pensiones contributivas
Indicadores de relación con la actividad económica	
I10	Tasa de actividad registrada
I11	Tasa de empleo registrado
I12	Tasa de desempleo registrado
I13	Tasa de desempleo registrado de larga duración
I14	Porcentaje de parados registrados de larga duración
Indicadores de hogares	
I5	Porcentaje de población en hogares unipersonales
I6	Índice de soledad
I18	Porcentaje de hogares unipersonales
I19	Tamaño medio del hogar

La tabla de atributos también contiene columnas para identificar el código de la sección censal y el nombre y código del municipio correspondiente, así como otras veintiuna columnas con nombre C# en las que se informa del número de orden del intervalo de clase en el que se clasifica el valor del respectivo indicador I# a efectos de su representación por coropletas (véase el apartado 5.2.3.2). En los casos en que la sección censal se ve afectada por riesgo de revelación, el valor de las correspondientes columnas C# es 9.