

Integración y conciliación de ficheros administrativos a nivel individual para la elaboración de estadísticas públicas en Andalucía

Autores: Ramos Poyatos, Juan de Dios; Cánovas Balboa, Rosa; Montañés Cobo, Víctor; Alfaro Sánchez, Luisa; Montesó Llano, Margarita; Venteo de Haro, Jose Juan; Parias Conradi, Gema; Viciano Fernández, Francisco; Molina Trapero, Jose; Hernández Rodríguez, Juan Antonio.

Instituto de Estadística y Cartografía de Andalucía (IECA)

Palabras clave: integración; registros administrativos; censo virtual; demografía; microdatos

1 Introducción

Los registros administrativos son 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 administrativo¹ genera información que queda almacenada en el correspondiente registro y ésta es susceptible de ser aprovechada para usos estadísticos.

De hecho, las características que presentan los registros administrativos (están descritos legalmente, normalmente 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 los registros administrativos cuentan también con limitaciones (cambios normativos, su origen nunca es estadístico, las definiciones o las unidades empleadas pueden ser distintas de las estadísticas, etc.) que es necesario conocer 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, promovido por Eurostat, que cita el uso de fuentes administrativas en varios de sus principios² con el fin de limitar la carga de información solicitada a los informantes y mejorar el uso eficiente de los recursos disponibles.

¹El Inventario de Fuentes de Información Administrativa de Andalucía (elaborado por el IECA) contabilizaba a finales de 2017 un total de 267 registros administrativos del ámbito competencial de la Administración de la Junta de Andalucía, susceptibles de ser usados para aprovechamiento estadístico.

²Código de Buenas Prácticas de las Estadísticas Europeas: Principio 8 de procedimientos estadísticos adecuados, indicadores 8.7 a 8.9; Principio 9 de carga no excesiva para los encuestados, indicadores 9.4 y 9.6; Principio 10 de relación coste/eficacia, indicador 10.3.

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.

Con el afán de 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, del año 2009. 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 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, es en esta situación en la que los registros administrativos y su integración juegan un papel determinante para poder superar los análisis en los que solo se tiene en cuenta a las personas desde un punto de vista unidimensional (estudio de un sólo 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 representatividad y la extrapolación de resultados a la población total.

Actualmente esto empieza a ser posible gracias a la mejora en los procedimientos informáticos que permiten almacenar, enlazar y procesar cada vez más información a un coste operacional cada vez menor.

En este contexto se enmarca el presente proyecto de integración de fuentes que nace como ejercicio de interconexión y aprovechamiento estadístico de la información procedente de varios registros administrativos a nivel individual que, partiendo de la Base de Datos Longitudinal de Población de Andalucía (BDLPA) como infraestructura vertebradora del proyecto, aporte información sobre las

distintas situaciones individuales de demanda de empleo, afiliación a la Seguridad Social y percepción de pensiones contributivas, todo ello a nivel de microdato y con vocación censal.

Concretamente, en este trabajo se abordará toda la problemática que el Servicio de Estadísticas Demográficas y Sociales del Instituto de Estadística y Cartografía de Andalucía (IECA) ha ido encontrando al encarar un proyecto de integración de fuentes de datos administrativos bajo el paraguas de la Base de Datos Longitudinal de Población de Andalucía (BDLPA), haciendo hincapié en los problemas relacionados con la construcción, enlace, sesgos, lógica y mecanismos de difusión que aparecen al tratar los ficheros administrativos de forma conjunta: ¿Qué colectivo de personas se estudia? ¿Qué hacer en caso de información contradictoria entre ficheros? ¿Cómo difundir microdatos? ¿Cómo corregir los “efectos administrativos” de los diferentes ficheros?

2 Descripción de los ficheros implicados

Este proyecto de integración y conciliación de fuentes tiene por objetivo 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. La información que se pretende recopilar y combinar para caracterizar la situación individual es la siguiente:

CAMPO	DESCRIPCIÓN
1. Información básica de la persona	
B01	Año de referencia de la información (a 31 de diciembre)
B02	Identificador (transversal) de la persona
B03	Año de nacimiento
B04	Sexo
B05	Nacionalidad
B06	Nivel de estudios
2. Información de la vivienda/hogar	
H01	Identificador transversal de la vivienda/hogar
H02	Provincia de la vivienda/hogar
H03	Tamaño del municipio de la vivienda/hogar (estratos)
H04	Tipología Rural-Urbana de la celda Inspire de 1 Km ² de la vivienda/hogar
3. Información de afiliación a la Seguridad Social	
S01	Número de afiliaciones de la persona
S02	Tipología de régimen de afiliación a la Seguridad Social
S03	Actividad económica (CNAE09, agrupación por Sección de Actividad)
S04	Tipo de contrato

CAMPO	DESCRIPCIÓN
4. Información de mutualistas de MUFACE	
M01	Situación administrativa
M02	Actividad económica (CNAE09, agrupación por Sección de Actividad)
5. Información de demandantes de empleo	
D01	Tipo de demanda de empleo
D02	Ocupación en la que el demandante desea trabajar preferentemente (CNO11)
D03	Tiempo que lleva inscrita la demanda
D04	Ámbito de búsqueda de empleo
D05	Actividad económica en la que ha trabajado anteriormente (CNAE09)
6. Información de pensiones contributivas	
P01	Número de prestaciones que percibe la persona
P02	Tipología de prestaciones percibidas
P03	Importe mensual bruto en céntimos de euro (acumulado si hay más de una prestación)

Tabla 1. Diseño de registro teórico con la información a generar.

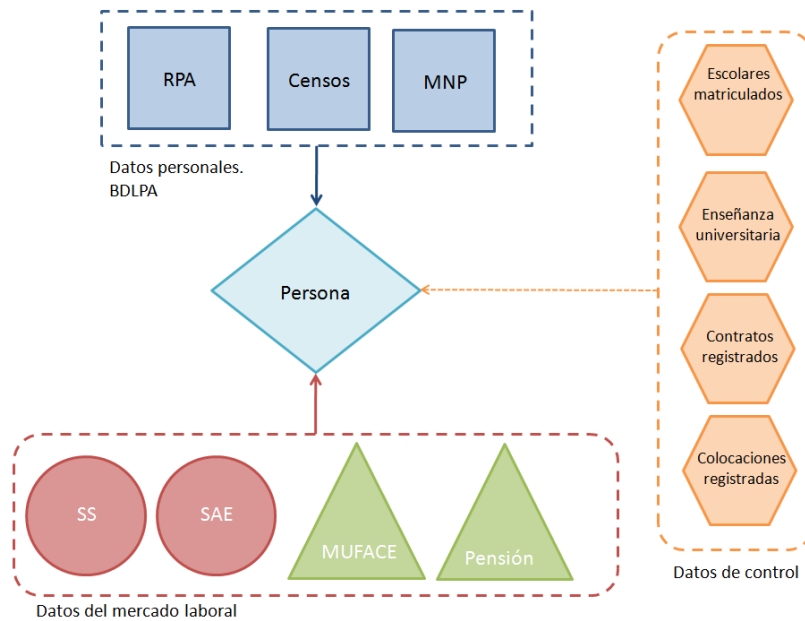
El eje central de la estructura de los datos es siempre la persona y a ésta se le van incorporando las características que presenta en el resto de ficheros analizados hasta completar la información a la que se quiere llegar. La información seleccionada, que tendrá fecha de referencia 1 de enero 2016, proviene de los siguientes ficheros:

– **Información personal**

- Base de Datos Longitudinal de Población de Andalucía (BDLPA)
 - Registro de Población de Andalucía (RPA).
 - Movimiento Natural de la Población (MNP).
 - Censos de población y viviendas de los años 1991, 2001 y 2011.

– **Información temática**

- Fichero de afiliaciones en alta laboral y cuentas de cotización a la Seguridad Social (SS).
- 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 (SAE).
- Registro de Prestaciones de la Seguridad Social (PEN).



A la información listada anteriormente hay que añadir la proveniente de otros ficheros que, si bien en esta edición del proyecto no se va a incluir dentro del fichero generado, ayudará a depurar y sincronizar la información proveniente del resto de ficheros.

– **Información de control**

- Fichero de escolares matriculados en enseñanzas no universitarias.
- Fichero de matriculados y de egresados académicos de las universidades públicas andaluzas.
- Contratos registrados en las oficinas de empleo (SAE).
- Colocaciones registradas (SAE).

A continuación se pasan a describir los distintos ficheros que sirven como fuente de la información a integrar, para conocer las particularidades de cada uno y exponer la problemática que conlleva un proyecto de integración como el presente.

2.1 Información personal

La información referida a las características demográficas básicas así como la información residencial de cada persona queda registrada en la Base de Datos Longitudinal de Población de Andalucía (BDLPA).

El origen de esta base de datos está en la integración de información proveniente del sistema de coordinación padronal, recogida en el Registro de Población de Andalucía (RPA), que junto con los datos provenientes de los Registros Civiles referidos a nacimientos, defunciones y matrimonios

(MNP), sirven para crear un marco integrado para las estadísticas de población y territorio en Andalucía.

El objetivo principal que se persigue al poner el acento en las personas, usando la información almacenada en la BDLPA, es el de obtener trayectorias vitales incluyendo en el análisis dos perspectivas importantes:

- Se aporta información de la persona así como de las modificaciones producidas en la residencia y en otras características demográficas.
- Se aporta información espacial sobre dónde residen las personas, siendo posible descender en el detalle hasta las coordenadas cartográficas.

La entrada de información usada en la BDLPA para reconstruir las biografías vitales de los residentes en Andalucía proviene de las siguientes fuentes estadísticas:

1. La información de última renovación padronal de 1996.
2. Los eventos recogidos en el circuito del Movimiento Natural de Población: partos, defunciones y matrimonios.
3. Las variaciones (flujos padronales) ocurridos en los padrones andaluces desde el nacimiento del sistema.
4. La información socio-demográfica recogida de la población residente en el censo de 2001.

El funcionamiento de este circuito de información comienza con los envíos de datos que realiza el Instituto Nacional de Estadística (INE) al IECA comunicando las variaciones municipales acumuladas en un único fichero semestral.

En este fichero de **flujos** semestrales se incluyen los nacimientos, defunciones, cambios de residencia e incidencias de gestión que se hayan producido en forma de altas, bajas o modificaciones en dicho periodo. Junto con esta información también se recibe semestralmente un fichero de **stocks** con los efectivos de población a 1 de enero y a 1 de julio de cada año.

A esto hay que añadir la información proveniente del Movimiento Natural de la Población que suministra datos de partos, defunciones y matrimonios una vez al mes y que sirven de indicador adelantado de la información padronal semestral para los casos de altas por nacimiento y bajas por defunción. Con la acumulación de todos estos datos se obtiene un fichero con la información del sistema desde el año 1996.

A los efectos de este proyecto, a partir de este fichero longitudinal (BDLPA) se extrae un corte transversal con fecha de referencia 1 de enero de 2016 que va a servir de marco poblacional sobre el que integrar a nivel individual la información de los ficheros temáticos antes mencionados. Esta

información básica, que por defecto prevalecerá frente a la aportada por las fuentes administrativas temáticas, comprende, junto con la identificación del individuo, las variables de sexo, fecha de nacimiento y nacionalidad, así como la información de localización territorial. No obstante este criterio de jerarquía entre fuentes puede verse alterado a favor de la fuente temática en los casos que el contraste de información a nivel individual conduzca a inconsistencias que precisen ser conciliadas (estos casos se analizarán en el apartado 3.1.)

2.2 Información temática

Ficheros de afiliaciones 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 envía trimestralmente la Tesorería General de la Seguridad Social al IECA.

En dicho fichero se encuentran los datos de Afiliaciones y Cuentas de Cotización de la Seguridad Social de Andalucía, suministrando 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.

Cabe reseñar también que el concepto de afiliado en este fichero excluye a los que cotizan a efecto de asistencia sanitaria exclusivamente; también excluye a los que están en situación de desempleo, los de convenios especiales y los que se encuentren en situaciones especiales sin efecto en cotizaciones.

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 último día del trimestre.

Los titulares incluidos en el fichero son clasificables en función de su situación administrativa en dos grandes grupos asimilables a la situación de activos y situación de jubilados. En el caso de los activos esta situación comprende las excedencias, funcionarios en prácticas, interinos, servicios activos, etc. Por lo que respecta a las jubilaciones, estas comprenden las de carácter voluntario, anticipadas, clases pasivas, etc.

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

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.

La demanda de una persona puede estar en tres situaciones: **alta** (con plena disponibilidad para los servicios solicitados), **suspendida** temporalmente, o **baja** (que no tiene disponibilidad porque no solicita ya los servicios o estos han sido satisfechos). Las demandas en alta tienen una caducidad y requieren ser renovadas periódicamente. La no renovación supone el paso a la situación administrativa de baja.

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.

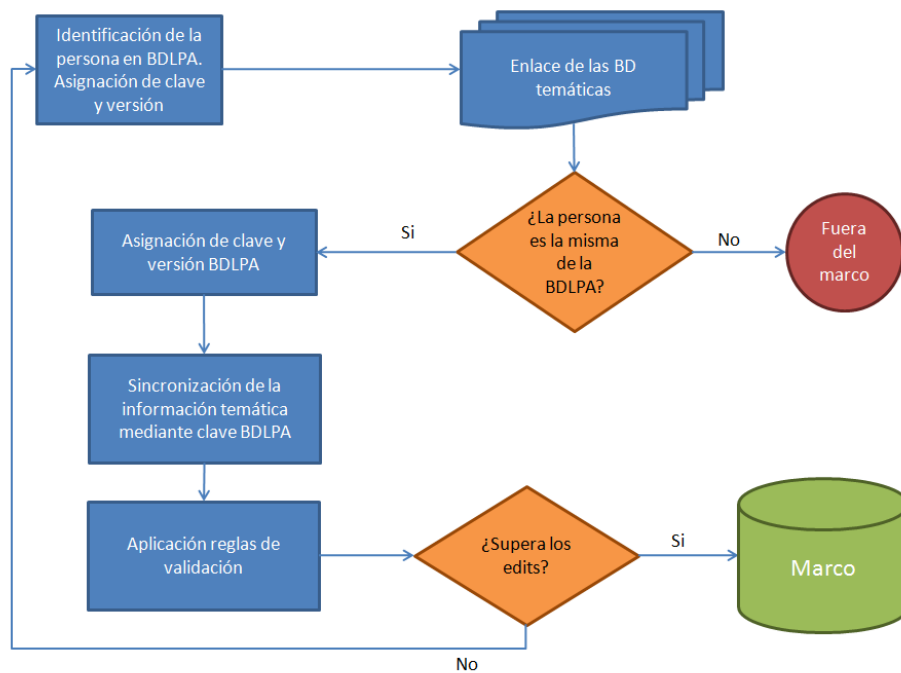
Registro de Prestaciones de la Seguridad Social.

El IECA recibe anualmente un fichero 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 pero 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 casos en los que concurren la pensión de jubilación junto con la de viudedad.

3 Descripción del proceso de integración

Actualmente 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 permiten esta identificación automática se han extendido a la información personal de los ficheros correspondientes a afiliados a la Seguridad Social, pensionistas contributivos y titulares de MUFACE, creando de manera unitaria una base de datos conjunta que incluye solamente datos personales. Es decir, aunque la información de trabajo específica de cada fuente no está integrada, sí lo está la información identificativa referente a las personas en cada fichero. Este procedimiento automático de enlace de datos identificativos se pretende extender para el resto de fuentes descritas en este proyecto.



El objetivo último que se persigue es la integración efectiva de la información temática de cada uno de los ficheros a nivel de registro individual, ya que la información personal identificativa ya estaría integrada de origen.

Este primer ejercicio de enlace e integración ha revelado problemas referentes a incoherencias en la información personal, así como problemas de cobertura de las fuentes ya que, como cabía esperar, no todas las personas se encuentran de manera simultánea en todos los ficheros.

La tabla 2 muestra la población base registrada en la BDLPA distribuida según se encuentren presentes en uno o varios de los registros administrativos que se pretenden integrar; en ella se observa que el 60,75% de la población residente se encuentra recogida además en algún otro fichero de los implicados en el proyecto.

BDLPA	SS	SAE	PEN	MUF	%
X					39,25%
X	X				27,97%
X		X			11,91%
X			X		14,77%
X				X	1,95%
X	X	X			2,78%
X	X		X		0,67%
X	X			X	0,08%
X		X	X		0,48%
X		X		X	0,00%

X			X	X	0,05%
X	X	X	X		0,08%
X	X	X		X	0,00%
X	X		X	X	0,00%
X		X	X	X	0,00%
					100,00%

Tabla 2. Distribución de las relaciones de concurrencia a nivel individual entre las distintas fuentes.

3.1 Problemas encontrados

En el desarrollo del proyecto se han ido presentando diversos problemas de consistencia en la información integrada, que al tratar de resolverlos han dado lugar a la redefinición de algunas de las premisas de partida así como de los objetivos perseguidos en el proyecto.

El origen de los problemas 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 tratar de detectar estos problemas se ha trabajado en la creación de unas reglas de validación o listado de **edits**³ al que deben someterse los datos durante el proceso de integración de las fuentes. Estos edits quedan definidos en dos tipos:

- **Errores.** Combinaciones de valores imposibles según la legislación o clasificación vigente. Actualmente hay definidos 16 errores.
- **Avisos.** Consisten en la notificación de valores poco habituales en las variables que pueden deberse a errores en una fuente y por tanto deben revisarse, aunque si no hay una corrección explícita, la combinación de valores puede ser explicada de conformidad con el correspondiente marco normativo. Actualmente se han definido 23 avisos.

A su vez, para cada uno de estos errores y avisos hay que explicitar las actuaciones encaminadas a eliminar el error o a explicar por qué se produce el aviso (si es legal o metodológicamente posible esa situación).

A continuación se describe con mayor detalle algunos de estos problemas.

³En el Anexo se presenta un relación de los edits (errores y avisos) aplicados.

3.1.1 Información personal incoherente

La principal fuente de errores ha sido la presencia de información común de una misma persona en los distintos ficheros pero de manera incoherente. Un ejemplo claro sería la **fecha de nacimiento**.

Aunque la información personal por defecto es la aportada por los datos de la BDLPA, hay situaciones en las que la fecha de nacimiento padronal no casa con la que procede de otra fuente. Esta situación puede generar en algunos casos un **error** (por ejemplo, el *E01* definido como una persona que cobra una pensión de viudedad y no ha cumplido los 14 años, edad legal para poder casarse) y en otros casos un **aviso** (por ejemplo, el *A09* definido como una persona que continúa afiliada a la Seguridad Social con más de 70 años, según BDLPA, y la fecha de nacimiento aportada por Seguridad Social da una edad inferior a ese límite).

En el caso del error, claramente se trata de una situación imposible que hay que resolver, y en el caso del aviso es una situación que, aunque a priori es posible, al ser tan infrecuente hay que revisar por si se tratara de un error.

Las posibles soluciones a estos errores pasan por comprobar los valores de cada fuente de forma independiente, empezando por la información padronal almacenada en la BDLPA.

Debido a la naturaleza longitudinal de la BDLPA (se van almacenando todos los cambios a lo largo del tiempo para una misma persona) se pueden tener varias versiones de un mismo dato para el individuo en cuestión. Por ejemplo, una persona se empadrona en un municipio andaluz con una fecha de nacimiento incorrecta y unos meses después se produce una modificación de ese dato, por lo que en la BDLPA se tienen almacenadas dos fechas de nacimiento para ese individuo. En estos casos se trataría de encontrar una versión del dato en BDLPA que sea compatible con el que aparece en la otra fuente y de esta forma resolver el error o aviso encontrado.

En otros casos el origen del error está en un mal enlace identificativo de la persona en la BDLPA. Esto ocurre en situaciones en las que dos personas con datos similares se identifican de manera errónea y se concluye que son la misma. En estos casos, al tratar de conciliar los datos con los provenientes de otra fuente puede que se estén mezclando los datos de dos personas distintas y, por ello, sea necesario deshacer el enlace realizado optando por otra persona como candidata con datos compatibles con los de la fuente a integrar.

En los casos en los que se ha dado esta situación, se ha aprovechado este bucle de retroalimentación para depurar la propia información almacenada en la BDLPA modificando los enlaces de las personas implicadas, aunque cabe reseñar que en ningún caso se modifica la información en las bases de datos originales, sino que el dato se modifica en el fichero de trabajo generado para el propósito del presente proyecto.

3.1.2 Problemas derivados de la metodología y fechas de extracción de la información

Otra fuente de problemas ha venido originada por la presencia de inconsistencias aparentes derivadas de la estancamiento de la propia dinámica de los procesos administrativos y los correspondientes calendarios de consolidación de la información en los registros administrativos temáticos. Un ejemplo de esto sería la presencia de casos en el fichero integrado inicial en los que en un mismo individuo concurren una situación de alta en la Seguridad Social y una situación de parado registrado en la fecha de referencia.

La definición de parado registrado viene dada en la Orden del Ministerio de Trabajo de 11 de marzo de 1985, por la que se establecen criterios estadísticos para la medición del parado registrado. Esta definición contabiliza como parado registrado al colectivo de personas que tengan demandas pendientes de satisfacer el último día del mes en los servicios públicos de empleo, excepto las que se encuentren en alguno de los denominados colectivos de exclusión: demandantes de otros servicios; demandantes ocupados; otros demandantes excluidos del parado registrado; trabajadores eventuales agrarios beneficiarios del subsidio especial por desempleo; otros demandantes de empleo no ocupados.

La aplicación práctica de la exclusión de colectivos por parte del SEPE⁴ a la hora de determinar los demandantes de empleo que deben ser clasificados como parados registrados cuenta con un proceso consistente en el cruce del fichero de demandantes de empleo con el fichero que contiene a todas las personas en situación de alta laboral en la Seguridad Social en ese momento.

En el caso de los regímenes por cuenta ajena, habida cuenta de que existe un plazo legal para que las empresas comuniquen las bajas a la Seguridad Social, se estima que en promedio puede tardar hasta diez días en incluirse una baja en el fichero de la Seguridad Social.

Por consiguiente, existe un cierto retraso en la incorporación de las bajas en el fichero de personas en alta en la Seguridad Social, por lo que su cruce por parte del SEPE con el fichero de demandantes de empleo puede repercutir en un incremento indebido del colectivo de demandantes excluidos del parado registrado. Esto es así porque la inscripción de una demanda de empleo en una fecha posterior a la de baja efectiva en la situación de afiliación a la Seguridad Social puede ser inconsistente (por desincronización) con la información aún no actualizada a fin de mes del fichero de Seguridad Social empleado para el cruce por el SEPE. Para limitar el efecto de esta demora en las cifras de parado registrado, el SEPE aplica un periodo de carencia en el cruce de datos con la Seguridad Social.

La aplicación del periodo de carencia hace que se pase al grupo de demandantes no parados a aquellos que tienen una demanda de empleo anterior en diez días al momento en que se extrae la información de la base de datos de la Seguridad Social.

⁴ La influencia de la implantación del SISPE en España. Luis Toharia Cortés (Universidad de Alcalá) y Miguel Ángel Malo Ocaña (Universidad de Salamanca). 2005

Así, si una persona tiene una demanda registrada en algún servicio público de empleo el 25 del mes y aparece como en situación de alta en la Seguridad Social en el cruce del último día del mes, esa persona no es excluida del cómputo del paro registrado a pesar de figurar como de alta en la Seguridad Social, ya que está dentro de ese periodo de diez días.

Por lo tanto, cuando en el marco de este proyecto se integran los datos procedentes del fichero de demandantes de empleo correspondientes al último día laborable de diciembre de 2015 con el fichero de situaciones de alta laboral en la Seguridad Social para esa misma fecha, es posible encontrar casos en los que se simultanea la situación de paro registrado y afiliado en alta laboral a la Seguridad Social como consecuencia de este efecto de desincronización en los respectivos calendarios de consolidación del dato administrativo y de la aplicación de la regla del periodo de carencia.

3.2 Enriqueciendo la información

En el proceso de integración de las distintas fuentes se habían dejado aparcadas algunas decisiones relativas a la inclusión de variables que pudieran enriquecer los análisis futuros a realizar a partir del fichero generado. Esto es debido, en algunos casos, a la baja calidad en algunas variables concretas que requerirían de un análisis más detallado de las fuentes disponibles y, en otros casos, por tratarse de variables que, si bien no se encontraban en los ficheros temáticos, sí pudieran ser obtenidas de otras fuentes.

Ejemplo de estos casos serían las variables **nivel de estudios**, el **grado de urbanización** del ámbito territorial del individuo y la inclusión de la **perspectiva de los hogares** en el análisis. A continuación se describen las decisiones tomadas para hacer efectiva dicha inclusión en el diseño de registro.

3.2.1 Nivel de estudios

En la primera fase del proyecto, cuando se estaba debatiendo qué variables deberían estar en el marco poblacional, se tomó como referencia el conjunto de variables sociales nucleares de caracterización de los individuos entre las que se encontraba el nivel de estudios.

La importancia de esta variable en el análisis socioeconómico de las personas está fuera de toda discusión, pero recuperarla de las fuentes disponibles no era una cuestión trivial y se presentaban ciertas dificultades que había que solventar si se quería contar de manera efectiva con la información de esa variable. A continuación se describe el procedimiento empleado para recuperar la información del nivel estudios de las personas con ayuda de la información almacenada en distintas fuentes de origen administrativo.

En principio, esta información aparece en la hoja padronal empleada por los ayuntamientos en la recogida de las inscripciones de las personas residentes, pero cuenta con muchas deficiencias que, si bien se han ido subsanado con el tiempo, hacen que no pueda usarse directamente la información padronal del nivel de estudios.

En general, el nivel de estudios padronal nunca ha sido tenido en cuenta por el INE para su explotación estadística, al menos desde que entró en vigor el actual sistema de variaciones padronales en 1996. Este hecho ha implicado que no se le haya prestado mucha atención a la calidad de esta información que, aunque se ha constatado cierta mejoría en los últimos años (desde 2009) debida sobre todo a la incorporación de datos procedentes del Ministerio competente en materia de educación, presenta algunas deficiencias que se reseñan a continuación.

La clasificación actual se recoge en la siguiente tabla:

Código	Descripción
10	NO SABE LEER NI ESCRIBIR
11	No sabe leer ni escribir.
20	TITULACIÓN INFERIOR AL GRADO DE ESCOLARIDAD
21	Sin estudios.
22	Enseñanza primaria incompleta, 5 cursos de EGB o equivalente o Certificado de Escolaridad.
30	GRADUADO ESCOLAR O EQUIVALENTE
31	Bachiller elemental, EGB completa, graduado escolar. Primaria completa. ESO.
32	Formación profesional de primer grado, FP de grado medio, oficialía industrial.
40	BACHILLER, FP 2º GRADO O TÍTULOS EQUIVALENTES O SUPERIORES
41	Formación profesional de segundo grado, FP de Grado superior, maestría industrial.
42	Bachiller Superior, BUP. Bachiller LOGSE.
43	Otros titulados medios (aux. clínica, artes y oficios, programador...)
44	Diplomado en escuelas universitarias (Empresariales, Profesorado de EGB, ATS y similares)
45	Arquitecto o ingeniero técnico.
46	Licenciado universitario, arquitecto o ingeniero superior. Grado Universitario.
47	Titulados de estudios superiores no universitarios.
48	Doctorado y estudios de postgrado o especialización para licenciados. Máster universitario.

Tabla 3. Códigos padronales de nivel de estudios

Junto con la lógica adaptación de los códigos tras la aparición de las titulaciones recogidas en la LOGSE (ESO, Grado y Máster), hay que mencionar además que existe un código "00" que originalmente indicaba "No aplicable por ser menor de 10 años". Este código fue modificado por el

Consejo de Empadronamiento en 2013 pasando a indicar “No aplicable por ser menor de 16 años”, es decir, se amplió 6 años más la edad de las personas para las que no se tenía la obligación de determinar el nivel de estudios poseído.

En cuanto al nivel de detalle de la información también se plantea un problema. La clasificación tiene en realidad dos esquemas: la agregada a un dígito (10, 20, 30, 40) y la no agregada, a dos dígitos. El problema para el uso de esta información en Andalucía es que algunos ayuntamientos con un volumen importante de población (Córdoba, Jerez, Sevilla...) utilizan la codificación a un dígito (o bien la envían a un solo dígito) y esto la hace inservible, ya que, por ejemplo, en la actualidad la categoría 40⁵ incluye a demasiada población y aporta muy poca información.

Para cuantificar todos estos problemas se presenta la siguiente tabla que contiene la información padronal para la población objeto de estudio de 16 o más años y la clasificación que se obtendría del nivel de estudios a 1 de enero de 2016 según año de nacimiento.

Códigos	<1930	1930-1939	1940-1949	1950-1959	1960-1969	1970-1979	1980-1989	1990-1999
00/NA	1.424	3.884	5.111	6.212	8.722	9.320	8.057	26.586
11	16.485	41.909	28.768	17.948	14.230	11.846	32.910	275.293
20	50.322	148.712	201.864	198.749	161.763	121.388	246.728	203.677
21	42.258	120.775	137.501	121.406	100.063	72.775	166.339	218.438
22	8.061	26.704	49.697	62.353	69.879	72.377	84.453	55.133
30	8.863	37.129	94.143	158.027	272.463	278.603	131.140	23.753
31	5.868	25.188	68.635	117.099	205.459	252.952	197.927	71.104
32	158	1.017	3.751	11.120	29.983	42.558	33.033	6.742
40	7.137	25.822	66.785	138.147	224.343	228.329	55.569	7.369
41	432	2.095	6.211	13.849	33.303	47.425	22.122	2.475
42	1.056	4.241	14.364	35.417	61.857	93.252	51.089	17.596
43	152	664	1.748	3.229	5.393	7.689	6.658	951
44	762	3.163	9.976	21.078	34.391	53.846	38.224	2.551
45	84	456	1.856	2.560	4.846	9.527	5.721	229
46	505	2.243	7.405	18.665	40.085	70.656	40.902	2.754
47	89	314	647	945	1.427	2.625	3.360	415
48	44	232	824	2.141	4.161	7.265	6.028	273

Tabla 4. Distribución de la población residente de 16 años o más en los distintos códigos padronales de educación.

Esta tabulación ofrece unos resultados muy decepcionantes, ya que el 44,7% de la población carece de una codificación detallada. Por ejemplo, entre los nacidos en la década de los 50 esta cifra alcanza

⁵Ver http://www.ine.es/metodologia/t20/t2030306_niveduc.pdf

el 53,2%. Además, hay una parte de la población (1%) que por edad debería tener un nivel de estudios asignado y no lo tiene o tiene un código fuera del rango de valores admisibles. En la generación de nacidos en los años 90 esta cifra asciende hasta el 2,9%.

Todos estos problemas conducen a dos posibles soluciones: o se abandona el uso de esta variable, que por otra parte forma parte de las variables nucleares imprescindibles en todo estudio social, o bien se emplea la información recogida en el resto de fuentes implicadas en el proyecto teniendo en cuenta las posibles incoherencias que pueden existir entre la misma información (nivel de estudios) recogida por fuentes administrativas distintas.

En este caso se ha optado por la segunda asumiendo esta premisa: las cifras descritas anteriormente pueden ser corregidas y mejoradas, teniendo siempre presente que el nivel de estudios no se pierde con el tiempo, es decir, siempre puede ir a más pero nunca a menos.

A continuación se exponen las actuaciones que se han desarrollado para poder aprovechar la información sobre el nivel de estudios que tiene cada una de las fuentes implicadas en este proyecto con el fin de que presente un ejemplo de la problemática de la integración de las distintas fuentes de origen administrativo.

Padrón

Aunque ya se han indicado los problemas de la información padronal, paradójicamente esta misma información puede ser útil para ayudar a mejorarla. En los casos en los que una persona ha residido en varios municipios de Andalucía se pueden ver los cambios producidos en el nivel de estudios de cada variación padronal. Así pues, se puede determinar el nivel de estudios más alto que se ha registrado en las inscripciones padronales para cada persona.

Censos de población y viviendas

De los censos de 1991, 2001 y 2011 se puede obtener información sobre el nivel de estudios. La información se obtiene por combinación de los campos *nivel de estudios* y *tipo de estudios* y se utiliza siempre que el individuo no tenga la marca de "imputado", ya que lo que interesa es la información efectivamente comunicada por el individuo. Esta vía presenta limitaciones debido a que se está trabajando con información anticuada y el censo de 2011 no aporta información individual al ser muestral, pero al menos para la población de más de 40 años que ya residía en Andalucía en 2001 es una buena forma de realizar correcciones.

Demandantes de empleo

El fichero de demandantes de empleo es especialmente útil ya que dispone de un código de título y de una fecha de obtención del título que los demandantes facilitan cuando se inscriben o actualizan su demanda de empleo, facilitando una información que, trabajada de forma longitudinal, es muy valiosa. El problema que presenta es la cobertura poblacional, ya que no se trata de un registro de

inscripción obligatoria, no dispone de información de los inactivos, ni de aquellas personas que estaban trabajando y habían dejado de demandar empleo. Por otro lado, dado el alto paro juvenil en España en general y en Andalucía en particular en los últimos años, aporta una gran cobertura al colectivo que los censos dejaban con más dudas.

Además, cabe esperar que cuando una persona termine su periodo de estudios y se inscriba en este registro para iniciar su búsqueda de empleo, este registro irá recogiendo la titulación de gran parte de las personas que se incorporen al mercado laboral.

Como complemento a este código del nivel de estudios, se dispone de otro campo relativo a la codificación del nivel de estudios según una adaptación de la Clasificación Nacional de Educación (CNED2000). Es especialmente útil para distinguir los niveles de estudios más bajos que no precisan de un título (analfabeto, sabe leer y escribir, fue al menos 5 años a la escuela) o cuando el título se obtuvo en el extranjero y por lo tanto no se corresponde con un código a nivel nacional.

Contratos y colocaciones registradas

En algunos de los contratos registrados consta una titulación, con la misma codificación que en el fichero de demandantes de empleo, pero sin una fecha de referencia de obtención del título.

Fichero de escolares matriculados. Enseñanza no universitaria

Este fichero es especialmente útil para asignar un nivel de estudios a los menores de 16, ya que hay obligación legal de estar escolarizado entre los 4 y los 15 años. El fichero permite saber en qué curso está matriculada una persona, lo cual implica que ha superado el nivel de estudios previo. Sirve por tanto para mejorar el nivel de estudios asignado a la población más joven que no cursa estudios universitarios y no se ha incorporado al mercado laboral. Como ejemplo, permite distinguir a aquellos menores con el código de estudios 22, que indica que han cursado "5 cursos de EGB o equivalente".

Boletines estadísticos de parto de MNP

En MNP se solicita el nivel de estudios del padre y de la madre. Evidentemente, este fichero no puede ser tan fiable como otros que necesitan una acreditación documental. Por este motivo, sólo se ha utilizado esta información para algunos casos concretos:

- La persona no tenía ningún nivel de estudios asignado.
- La persona tenía un nivel genérico asignado (un dígito) y el nivel reportado por la fuente es compatible.
- La persona reporta un nivel superior al que tenía asignado y además esta información se repite en más de un boletín y es única.

MUFACE

Este fichero de personas al servicio de la Administración proporciona información, entre otros, sobre profesorado y técnicos del Estado quienes poseen titulación universitaria.

Los campos fundamentales de este fichero para el objetivo de asignación de nivel educativo son el grupo y el cuerpo. Por ejemplo, se sabe que aquellas personas pertenecientes al grupo A1 cuyo cuerpo es “Catedráticos de universidad” deben estar en posesión de un título de doctorado.

Afiliados a la Seguridad Social

Es muy poca la información que puede aprovecharse de este fichero para este objetivo. No obstante, hay ciertas combinaciones de grupo de cotización con relaciones laborales especiales que sí son de utilidad. Ha servido para reasignar casos de licenciados y doctores.

Ficheros de enseñanza universitaria

Se dispone de la información de matriculados y egresados de las nueve universidades públicas andaluzas desde el año 2010. Mediante el tipo de estudios cursados, la titulación, los años de matriculación y acceso y la forma de acceso se puede recuperar una información muy útil para corregir el nivel de estudios.

Tras todo este proceso de búsqueda, se pasa de un 44% de información sin codificación detallada a tan sólo un 7,6%. La codificación con nivel 40 pasa de afectar a un 10,9% a un 2,3%. Esta cifra es bastante más asumible y da pie a imputar los valores no detallados a través de una diferenciación por edad y nacionalidad, aunque el proceso se puede hacer mucho más complejo.

3.2.2 Grado de urbanización⁶

El Grado de urbanización es una clasificación que busca caracterizar la intensidad del asentamiento de la población residente en un territorio usando las celdas de la malla estadística⁷ de 1 km². La geolocalización de los domicilios de residencia de las personas contenidas en la BDLPA permite asignar a éstas el grado de urbanización de la correspondiente celda de residencia.

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 EUROSTAT para la asignación de las tipologías territoriales en las celdas de la malla estadística de 1 km² en tres tipos⁸:

⁶Para un mayor detalle en este punto, se recomienda ver el póster “Grado de urbanización. Las ciudades de Andalucía” presentado por el IECA en estas mismas JECAS.

⁷La malla estadística es un constructo generado siguiendo las indicaciones derivadas de las experiencias llevadas a cabo por el proyecto Geostat (ESSnet project Geostat) del Foro Europeo para la Geografía y la Estadística (EFGS), y que se emplea con la idea de tener un mismo sistema de referencia espacial para la totalidad de Europa en el que la información sea comparable conforme a la iniciativa INSPIRE.

⁸Artículo 4 ter, punto 2 del Reglamento (UE) 2017/2391.

- **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.

3.2.3 Aproximación heurística a la identificación de viviendas/hogares

Para aumentar la potencia analítica de la información integrada a nivel individual, 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. Toda esta información se va a emplear para derivar un constructo o entidad equiparable a la vivienda/hogar que quede determinado por una clave identificativa única. Como resultado, todo individuo dispondría de una clave de vivienda/hogar y aquellos individuos que compartan dicha clave podrían ser considerados miembros de la misma vivienda.

En teoría debería existir una relación 1 a 1 entre el hueco y la hoja padronal, de forma que cada hueco tuviera una única hoja padronal y que ésta no pudiera estar asociada a ningún otro hueco, de lo que se derivaría la identificación unívoca del constructo vivienda. En la práctica este escenario ideal no se presenta ya que además de la relación 1:1 aparecen otras relaciones de cardinalidad variable entre los huecos y las hojas padronales (1:N, N:1 y N:N).

Los principales motivos para que no exista una relación 1:1 entre hueco y hoja padronal serían:

- **Cambios administrativos**, en los que los ayuntamientos modifican la denominación de las vías, realizan alteraciones en el seccionado o no asignan dirección a zonas de nueva construcción.
- **Establecimientos colectivos**, en los que existen hojas padronales con un número elevado de personas.
- **Menores viviendo solos o con otros menores.** Existencia de personas viviendo solas en una hoja padronal. En principio no sería un error, pero se debe comprobar la existencia de **menores** en esta condición ya que puede deberse a una situación anómala.

El estudio realizado sobre estas casuísticas y el procedimiento llevado a cabo para amortiguar en parte su huella en la creación de una clave única de vivienda/hogar, ha llevado a crear una serie de comprobaciones que se describen a continuación.

Se inicia el procedimiento con la determinación de aquellos casos en los que la relación 1:1 fuera segura. En este primer paso se crearían las claves de viviendas, obteniendo que aproximadamente el 80,9% de los identificadores personales quedan asignados a una vivienda segura, en tanto que en el caso de huecos y hojas padronales la asignación es del 86,4% y del 79,9%, respectivamente.

Para trabajar con el resto de situaciones, en las que la relación no es 1:1, se recurre a distintas estrategias con el fin de reconducir todos los casos de cardinalidades distintas a la 1:1 al objetivo de una correspondencia unívoca entre hoja padronal y hueco.

En el caso de los **cambios administrativos** se recurre a la información aportada en campos auxiliares como la pseudovía que, si bien no se explota masivamente al ser un literal, puede ser de utilidad en circunstancias como las de los diseminados. Ejemplos de las actuaciones llevadas a cabo en este sentido serían:

- **Asignación de viviendas sin dirección o con hoja padronal nula.** Se pueden arreglar en las situaciones en las que todos los huecos van acompañados por una misma descripción de pseudovía (en caso de haberla). Esta circunstancia solo permitiría la identificación de viviendas separadas a través de la hoja padronal, de haberla, o mediante el hueco, en el caso contrario. Ejemplos característicos en los que se presentan estas situaciones son los diseminados y los complejos residenciales.
- **Asignación de viviendas con modificaciones en distrito y sección.** Entre las relaciones de cardinalidad 1:N entre huecos y hojas padronales, se presenta un subconjunto en el que si se ignoran los atributos distrito y sección (DDSSS) de la hoja padronal, la cardinalidad pasa a ser 1:1. De esta forma, el caso se transforma en uno de asignación de vivienda segura. Es plausible que la causa de estos casos se deba a que la inscripción padronal de los residentes en la vivienda haya ocurrido en distintos momentos y que en un tiempo intermedio se haya actualizado el distrito y sección en que se ubica la vivienda, sin que esta actualización se haya trasladado por parte del ayuntamiento a los habitantes previamente empadronados.

Otra de las situaciones anómalas es en la que existen huecos u hojas padronales con un número muy elevado de individuos (**Establecimientos colectivos**). Esto presenta un problema, ya que no en todos los casos detectados está clara la identificación como residencia colectiva del hueco u hoja.

Para poder comprobar este aspecto, se han buscado manualmente huecos con un número elevado de personas a fecha de 01/11/2011 (para comparar con datos censales) y el calificador de local colectivo incluido en el Padrón. Para dicha revisión se han empleado fuentes externas (ej. Directorio

de Centros de Mayores de la Junta de Andalucía) y la información de la pseudovía aportada para cada hueco.

Al ser una revisión muy costosa, se ha limitado a los valores más extremos (tanto por arriba como por abajo) dejando el resto sin revisar, sólo con la información que proporciona el Padrón, si la hay. A la vista de los resultados, comparando con la información procedente del Censo de 2011, queda constancia de un subregistro de colectivos que no se alcanza a identificar al no existir fuentes auxiliares.

Para el caso de los **menores que se encuentran solos** en una hoja padronal o hueco ha sido de especial utilidad la tabla de *relaciones familiares* construida a partir de los datos de la BDLPA. Esta tabla contiene la información individual de relaciones de paternidad y maternidad de los residentes en Andalucía.

Así, en los casos en los que se detecta un menor viviendo solo o con otros menores (hueco u hoja) se procede a buscar a sus padres, y a partir de ellos se asigna a los menores el hueco y la hoja padronal donde residen (siempre que no sea en colectivo). Esta asignación se condiciona a que el menor comparta al menos uno de los dos campos con los progenitores, dando prioridad a la madre frente al padre.

En caso de que las condiciones anteriores no se cumplan, se repite el proceso buscando a los abuelos, tanto maternos como paternos. Si tampoco comparte hueco u hoja padronal con ninguno de los cuatro abuelos, el menor queda marcado como “menor viviendo solo” para su tratamiento posterior.

Una vez resueltos estos problemas sigue quedando un **resto** que no cumple con la relación 1:1 que, como ya no queda dentro de los explicados anteriormente, hay que seguir reduciendo. No obstante, hay que tener en cuenta que no necesariamente todos los casos que no respeten esa cardinalidad se deben a errores, ya que en muchos de ellos se trata de viviendas en las que realmente hay varios hogares.

Para este resto de casos se emplean tablas auxiliares como la de *relaciones familiares*, descrita anteriormente, y la de *relaciones matrimoniales* obtenidas de la BDLPA. Esta última contiene los datos personales identificativos correspondientes al cónyuge1 y al cónyuge2.

Estas tablas se usan para aquellas viviendas con relación 1:N entre hueco y hoja padronal que han quedado sin resolver: se busca a los miembros de los correspondientes hogares en dichas tablas y, en caso de encontrarse relación entre ellos, se fusionan los respectivos hogares incluyendo junto a los miembros relacionados los restantes miembros de dichos hogares.

Por último quedaría el colectivo de **personas imputadas**, no tratado anteriormente al no disponer de direcciones y/o hoja de empadronamiento. Para su tratamiento se separan en dos grupos: personas mayores y menores de edad.

Para los mayores de edad que no tienen ningún tipo de información acerca de la descripción postal, se crea una vivienda y hogar ficticio por cada individuo. Si se dispone de información del edificio, se agrupan en la misma vivienda y hogar los imputados que residan en dicho edificio.

Para los menores de edad imputados se lanza una búsqueda de la madre y, en caso de no encontrarla, del padre. Si se localiza algún progenitor, el menor queda asignado con éste y hereda su información territorial, incluido el municipio. Si no se consigue localizar a los progenitores, entonces se recurre a la tabla de relaciones familiares para hacer la asignación con alguno de los abuelos. De este colectivo queda un resto de 11.366 menores sin vivienda/hogar a los que por el momento no se ha podido ubicar con ningún familiar.

Como resultado de la ejecución de todo este procedimiento se consigue una asignación segura de personas en viviendas con relación 1:1 (una hoja, un hueco) para el 88,4% de la población, así como para el 91,8% de los huecos y el 89,7% de las hojas padronales.

Hay que señalar que este procedimiento está en continua revisión por lo que las cifras podrían variar en el futuro debido a la mejora en los procedimientos o a la incorporación de más información auxiliar que ayude en la creación del constructo vivienda/hogar.

4 Resultados esperados

El proyecto se encuentra actualmente en una fase avanzada de gestación, aunque aún no se han iniciado los trabajos tendentes a la generación de productos de difusión. No obstante, se puede avanzar que el propósito es que, una vez constituido el fichero marco poblacional integrado, se elaborarán al menos dos productos:

- Una muestra de microdatos constituida por un 10% de los individuos contenidos en el marco poblacional, junto con la identificación de la vivienda/hogar (la muestra contendrá todos los integrantes de cada vivienda seleccionada).
- Una explotación estadística del marco poblacional orientada a la tabulación de tipologías de hogares en función de las variables integradas.

ANEXO

Errores:

1. La persona cobra una pensión de viudedad y no ha cumplido los 14 años, edad legal para poder casarse.
2. La persona tiene una pensión por incapacidad antes de alcanzar los 16 años, edad mínima para poder trabajar y por tanto para percibir una pensión de este tipo.
3. La persona se encuentra como demandante en SAE, tiene 65 años o más pero su código de exclusión es mayor que 1032 (Jubilación o >65).
4. La persona tiene más de una afiliación en el régimen (0161) del Sistema Especial Agrario o en el régimen (0825) Trabajadores del Mar Cuenta Propia.
5. La persona tiene más de una afiliación en el régimen de autónomos (0521).
6. La persona se encuentra como demandante de empleo pero tiene menos de 16 años.
7. La persona se encuentra en SAE como demandante, tiene más de 24 años y ha trabajado antes (CNAE<>'00') pero está codificada con causa de exclusión 1083, que indica que no se pueden dar ambas condiciones de manera simultánea.
8. La persona se encuentra en SAE como demandante, tiene menos de 55 años pero su causa de exclusión es 1135, que indica que es demandante prejubilado con 55 años o más.
9. La persona se encuentra en SAE como demandante extranjero comunitario o extracomunitario y la nacionalidad indicada por BDLPA es española.
10. La persona se encuentra en SAE como demandante extranjero comunitario y la nacionalidad de BDLPA lo clasifica como extracomunitario
11. La persona se encuentra en SAE como demandante, pero la causa de exclusión de paro indica que se trata de un ciudadano NO comunitario y la nacionalidad de BDLPA lo considera comunitario.
12. La persona está afiliada a la seguridad social, tiene menos de 16 años y no consta en el régimen de artistas (0112).
13. La persona tiene una edad inferior en 16 años y aparece en MUFACE.
14. La persona pertenece al grupo A1 en MUFACE (en teoría exige un grado), tiene menos de 22 años y NO ES profesor de música o arte (que son grados superiores en los que esta edad no es una restricción).
15. La persona pertenece al grupo A2 en MUFACE (en teoría exige una diplomatura) y tiene menos de 21 años.
16. La persona es menor de edad y pertenece a los cuerpos de seguridad del Estado (MUFACE).

Avisos:

1. La persona cobra una pensión de viudedad y tiene entre 14 y 21 años, edades sospechosas de ser erróneas ya que es poco habitual que una persona se case con esas edades y además quede viuda. Se añade la condición de que el año de nacimiento de BDLPA y Pensiones sea diferente, ya que de lo contrario una fuente estaría confirmando la otra.

2. La persona tiene una pensión por incapacidad antes de alcanzar los 25 años (límite arbitrario). Esta edad tan baja es un dato anómalo pero no necesariamente falso. Se añade la condición de que además el año de nacimiento sea distinto entre BDLPA y Pensiones, ya que si fuera coincidente una fuente conformaría a la otra.
3. La persona tiene una pensión de jubilación antes de cumplir los 40 años (límite arbitrario) y el año de nacimiento entre BDLPA y Pensiones es discrepante.
4. La persona aparece en Seguridad Social y Pensiones, consta que tiene una pensión por jubilación y está afiliada en el régimen general sin que conste un coeficiente de parcialidad.
5. La persona aparece como pensionista y como demandante de empleo, tiene menos de 65 años y ninguna de sus prestaciones es de tipo 1 (incapacidad).
6. La persona aparece con contrato fijo-discontinuo en alta como afiliada a la Seguridad Social y consta como demandante de empleo (SAE) con causa de exclusión 1043 (fijo-discontinuo)
7. El individuo aparece en SAE y Pensiones, es menor de 65 años y ninguna de sus pensiones es de jubilación. Sin embargo, en SAE su código de exclusión del paro registrado es el de "1032 Jubilado", no el de pensionista.
8. La persona consta en Pensiones con una prestación de tipo 'J' (jubilación procedente de incapacidad), pero la edad es inferior a 65 años y la indicada en Pensiones también es inferior a dicha edad.
9. La persona continúa afiliada a la Seguridad Social con más de 70 años (arbitrario) y la fecha de nacimiento aportada por Seguridad Social da una edad inferior a ese límite.
10. La persona se encuentra en SAE y Seguridad Social, con un código de exclusión 1032 (Jubilación), pero tiene menos de 65 años y no está presente en el fichero de Pensiones. (Puede tratarse de excluidos por pensiones no contributivas).
11. La persona se encuentra en SAE y Seguridad Social, con un código de exclusión 1033 (Pensionista), pero no está presente en el fichero de Pensiones.
12. La persona se encuentra en SAE (pero no en Seguridad Social), con un código de exclusión 1033 (Pensionista) pero no está presente en el fichero de Pensiones.
13. La persona aparece en SAE como código de exclusión 1044-TEAS, pero no consta en el fichero de afiliados a la Seguridad Social.
14. La persona se encuentra inscrita en SAE y Seguridad Social, tiene causa de exclusión 1044-TEAS y ninguno de sus regímenes de Seguridad Social es 0161 o 0163
15. La persona pertenece al grupo A1 en MUFACE, tiene menos de 22 años y ES profesor de música o arte (que son grados superiores en los que esta edad no es una restricción) y la edad en ambas fuentes no coincide.
16. La persona está en edad de jubilación, está inscrita en MUFACE, pero se encuentra en situación de excedencia o suspensión.
17. La persona está inscrita en MUFACE considerada como funcionaria en activo con una edad superior a 70 años.
18. La persona está inscrita en MUFACE considerada como funcionaria menor de 40 años y está jubilada y la edad no coincide en ambas fuentes.

19. La persona tiene más de 100 años, se encuentra en el registro de MUFACE y el año de nacimiento discrepa entre ambas fuentes.
20. La persona tiene más de 100 años.
21. La persona es demandante ocupado agrícola según SAE, pero no aparece en afiliados a la Seguridad Social.
22. La persona es demandante de empleo excluido del paro registrado por estar de alta en régimen general o de autónomos, pero no aparece en el fichero de afiliaciones a la Seguridad Social.
23. El nivel de estudios de la persona es incompatible con la edad.