



Islas Canarias
Del 15 al 19 de noviembre de 2021



Junta de Andalucía
Consejería de Transformación Económica,
Industria, Conocimiento y Universidades
Instituto de Estadística y Cartografía de Andalucía

SILBE (Sistema Integrado Longitudinal de Base Empresarial)

Martín Manzanera Díaz

Instituto de Estadística y Cartografía de Andalucía
martin.manzanera@juntadeandalucia.es

Marina Enguídanos Weyler

Instituto de Estadística y Cartografía de Andalucía
marina.enguidanos@juntadeandalucia.es

Rubén Noguerras Sánchez

Instituto de Estadística y Cartografía de Andalucía
ruben.noguerras@juntadeandalucia.es

1. Introducción

El objeto del Sistema Integrado Longitudinal de Base Empresarial, en adelante SILBE, es crear un sistema para incorporar toda la información relevante relativa a empresas procedente de los registros económicos, tributarios, administrativos y estadísticos que se reciben o generan en el Instituto de Estadística y Cartografía de Andalucía (IECA), conformando un sistema integrado con los datos de territorio y de población.

Servirá de repositorio único de la información de empresas y debe permitir una relación biunívoca de las empresas entre los ficheros de distinto origen a través de un identificador único. Permitirá conocer las empresas activas y de baja a una fecha concreta, basándose en las jerarquías entre ficheros y entre variables que se definan, así como realizar extracciones de esos cortes temporales. Asimismo, permitirá calcular las magnitudes relativas a demografía empresarial obteniéndose una visión longitudinal de la información.

2. Proceso de desarrollo de SILBE

Para desarrollar el proyecto se está utilizando el GSBPM (Generic Statistical Business Process Model), estándar internacional desarrollado por UNECE para definir los procesos y subprocesos del modelo de producción estadística. Este estándar está concebido para dar cabida a todo tipo de proyectos estadísticos.

Dicho estándar comprende todas las fases de un proyecto, incluida la difusión. En el caso del Sistema Integrado Longitudinal de Base Empresarial (SILBE), se pondrá el foco en las primeras fases del GSBPM que son las que comprenden los procesos de diseño del Sistema. Se mencionan dichas fases, sin entrar a desarrollar para este documento, los subapartados que las componen:

Fase de especificación de necesidades:

- 1.1. Identificación de necesidades
- 1.2. Consulta y confirmación de necesidades
- 1.3. Definición de objetivos
- 1.4. Identificación de conceptos
- 1.5. Comprobación de la disponibilidad de datos
- 1.6. Elaboración del plan general y documentación

Fase de diseño:

- 2.1. Diseño conceptual
- 2.2. Diseño de las descripciones de las variables
- 2.3. Diseño de la recolección
- 2.4. Determinación del marco muestral y tipo de muestreo
- 2.5. Diseño del procesamiento y análisis
- 2.6. Diseño de los sistemas de producción y de los flujos de trabajo

Fase de construcción

- 3.1. Construcción de instrumentos de recolección
- 3.2. Construcción o mejora de componentes
- 3.3. Construcción o mejora de componentes de difusión
- 3.4. Configuración de flujos de trabajo
- 3.5. Pruebas al sistema de producción
- 3.6. Prueba piloto del proceso estadístico

3.7 Finalización del sistema de producción

Se han identificado todas las fuentes de datos internas y externas existentes en el Servicio de Estadísticas Económicas, su diseño de registro, así como las posibles relaciones entre ellas. Además, se han identificado las fuentes de datos de otros Servicios del Instituto de Estadística y Cartografía de Andalucía (IECA), del Sistema Estadístico o de otros organismos que puedan participar en la actualización de las actividades estadísticas del Servicio de Estadísticas Económicas o en la creación de nuevos productos de análisis.

Las fuentes de información identificadas se normalizarán y se incorporarán a las actualizaciones de las diferentes actividades estadísticas, de tal forma que permita una automatización del procesamiento y análisis de datos.

3. Descripción de SILBE

Mediante el proyecto SILBE se está creando un Sistema de Información que permita aglutinar, normalizar, relacionar y estructurar las diversas fuentes de datos, así como establecer relaciones entre los mismos, creando claves únicas que faciliten la identificación y relación de las unidades en las distintas fuentes del Sistema. Se establece la relación de las unidades del Sistema con el territorio (Callejero Digital Unificado de Andalucía CDAU) y con la población (Base de Datos Longitudinal de Población de Andalucía BDLPA).

La unidad fundamental del Sistema es la empresa, definida como la combinación más pequeña de unidades jurídicas que produce bienes o servicios y que disfruta de una cierta autonomía de decisión principalmente a la hora de emplear los recursos de los que dispone (Reglamento de la Unión Europea (696/93). Hay otras unidades relacionadas con la misma (Establecimientos, Empresa Estadística, Grupos de Empresas, Cuentas de cotización, ...) que deberán recogerse en el Sistema, guardando relación con dichas unidades legales.

El Sistema permitirá a su vez obtener el Directorio de empresas con actividad en Andalucía a una fecha concreta, así como la generación de diversos subproductos necesarios para la difusión de la información a través de la web del IECA o para respuestas concretas a los usuarios.

SILBE será a su vez un gran repositorio de información preparado para ser interconectado con otros proyectos del IECA, de forma que sirva de soporte a los mismos.

El Sistema también permitirá acceder a la información relativa a la actividad económica de los distintos registros de manera rápida y eficiente. Se clasificará la información, aplicando criterios de versionado de valores y de jerarquización de prioridades al incorporar la información. Se establecerán prioridades entre las distintas fuentes para establecer una jerarquía entre las mismas y los campos que actualizan. La priorización de fuentes se podrá fijar a nivel de variables, de forma que no todos los campos de una fuente tengan la misma prioridad frente a otras.

En concreto el trabajo consiste en la creación de un entorno tecnológico que permita incorporar y actualizar automáticamente las diversas fuentes de información,

normalizando, creando identificadores únicos y relacionando las unidades, de forma que se automatice y sistematice toda la información económica y administrativa actualmente disponible que se considere relevante. A su vez permitirá el realizar extracciones del Sistema para obtener información económica de sectores, municipios, distritos y otras delimitaciones geográficas.

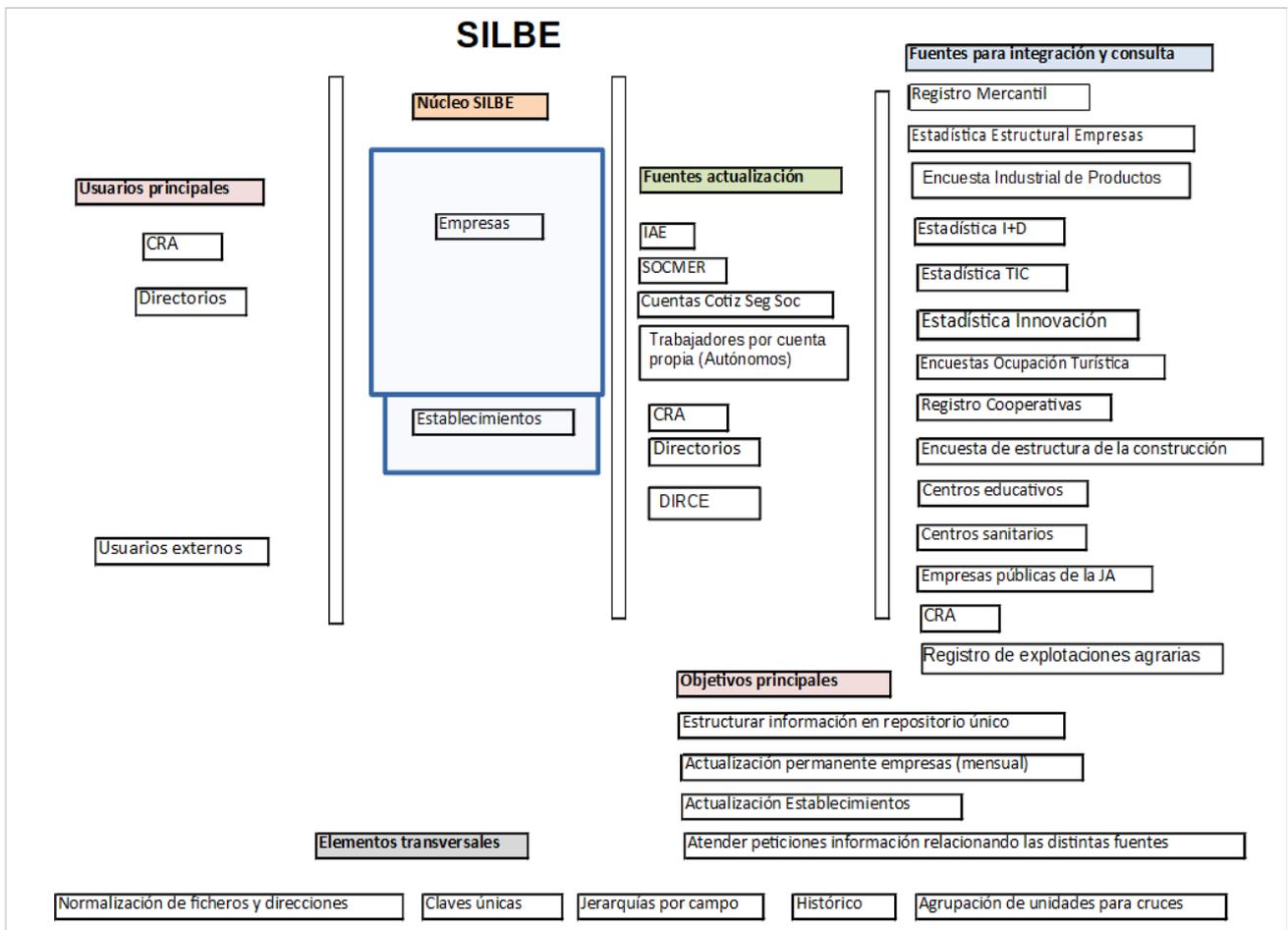
El diseño de los procesos contempla la componente histórica. Proporcionará soporte para los cálculos de la demografía empresarial considerando así el aspecto longitudinal de la información.

El entorno tecnológico del proyecto es:

Pentaho Community 9.1 o superior
Base de datos Postgres 12

El esquema general del proyecto se muestra a continuación:

Sistemas externos conectados: **CDAU** **BDLPA**



4. Objetivos de SILBE

Los objetivos y las razones que impulsan la consecución de SILBE son:

a) Racionalizar la gestión y almacenamiento de los datos de empresas disponiendo de un repositorio único que permita relacionar las empresas en las distintas fuentes que se reciben o generan en el IECA.

Hasta ahora existen numerosos repositorios diferentes para los datos de empresas. Falta conexión entre los mismos y en algunos casos están duplicados debido a los distintos perfiles de sus usuarios, que pertenecen a diferentes proyectos y no tienen acceso a los mismos repositorios, aunque la necesidad de acceso a esa información pueda ser común. Con SILBE se pretende mejorar este aspecto del almacenamiento y gestión de la información.

b) Normalizar las fuentes que se incorporen al Sistema y crear identificadores únicos.

Los ficheros relativos a empresas que se reciben o generan en el IECA no están normalizados en su gran mayoría, lo que dificulta su tratamiento y su cruce con otras fuentes. Por ello es necesario que toda la información que se incorpore al Sistema esté normalizada conforme a unos parámetros homogéneos.

El Sistema generará identificadores únicos que permitan la identificación y relación de las unidades en las distintas fuentes del Sistema.

c) Relacionar el Sistema de empresas con los sistemas de territorio y población.

En el IECA existen dos infraestructuras de información, una de territorio y otra de población, con las que el Sistema de empresas estará interconectado. El objetivo es, de un lado, poder conocer y explotar la información de las empresas desde la perspectiva de la población y, de otro lado, el incorporar toda la información posible del territorio al Sistema de empresas de forma que se pueda potenciar el análisis del sistema empresarial desde el punto de vista del territorio.

d) Permitir la actualización del Directorio de empresas con actividad económica en Andalucía y de la Demografía empresarial con un menor consumo de recursos y en menor plazo de tiempo.

La actualización del Directorio de Empresas y Establecimientos consume gran cantidad de tiempo y recursos dada su complejidad, por lo que se plantea que SILBE incorpore y automatice parte de los procesos del Directorio de forma que permita obtener un Directorio de Empresas reduciendo tiempos y recursos.

Para ello será necesario que SILBE gestione altas y bajas de empresas en el Sistema y relación de fuentes y campos en cuanto a dichas altas, bajas y modificaciones. Una vez obtenido el Directorio de empresas, a partir de dicho Directorio y recurriendo a fuentes adicionales, se obtendría el Directorio de Establecimientos.

e) Facilitar las operaciones relacionadas con el Sistema de Cuentas, en cuanto a la obtención de muestras de empresas y obtención de macromagnitudes de las empresas incorporadas en el Sistema, entre otras.

El Sistema de Cuentas encuesta con periodicidad anual una muestra de empresas de distintos sectores económicos y se trata que a través de SILBE se pueda obtener dicha

muestra, conocer las empresas que están siendo encuestadas por otras operaciones estadísticas, ya sean del IECA o del INE.

El Sistema de Cuentas incorpora a sus cálculos macromagnitudes económicas procedentes de los Registros Mercantiles. SILBE incorporará la información de empresas procedentes de los Registros Mercantiles (SABI) y facilitará dicho intercambio de información con el Sistema de Cuentas.

5. Tareas en curso del Proyecto

Actualmente están en curso varias tareas del proyecto, que se están realizando en paralelo. Al ser un proyecto de envergadura, la ejecución de las tareas se va realizando de forma secuencial.

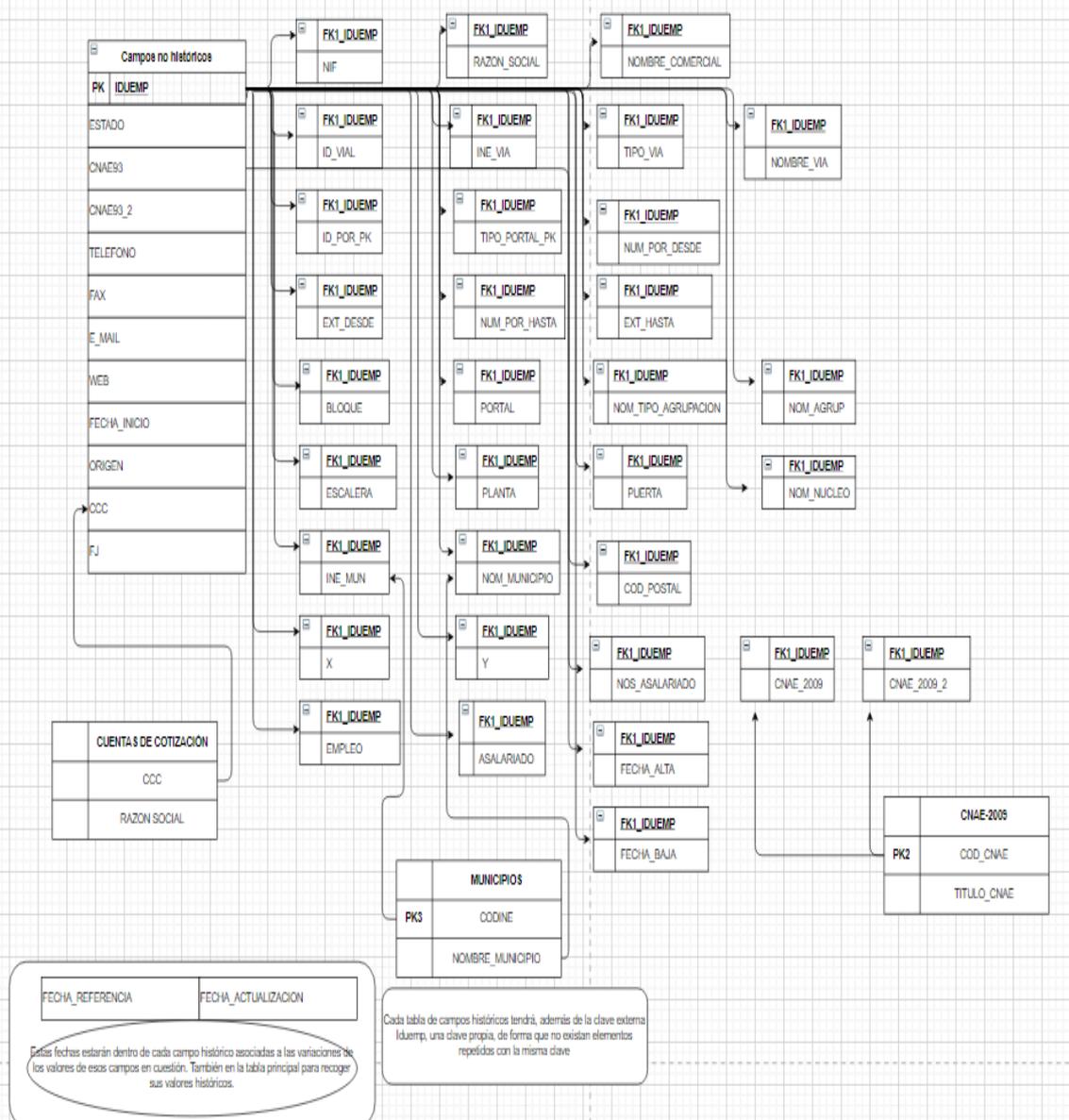
5.1 Creación de estructura Empresas

La creación de la estructura en la que se están volcando las distintas variables de las unidades empresas, disponen de una estructura de campos históricos y de otra de campos no históricos. Los campos históricos recogen todas las modificaciones que se vayan produciendo en dichos campos, con sus distintas fechas de actualización y de referencia. La actualización de los mismos seguirá una jerarquía, que será independiente para cada campo, en función de la importancia de las fuentes en cada caso concreto. Se recogerán las fechas de modificación en cada uno de los campos históricos, de forma que cuando se hagan extracciones del Sistema, de cada campo se seleccione el dato correspondiente a la fecha seleccionada o al intervalo de fechas correspondientes.

En dicha estructura, inicialmente se volcará el repositorio de empresas que actualmente está almacenado en tablas de Oracle, suponiendo más de 11.600.000 registros. Uno de los procesos a llevar a cabo inicialmente es la asignación del Identificador único de empresa (IDUEMP) a cada una de las empresas del repositorio. A lo largo de los años las empresas cambian de NIF, forma jurídica, ...sin que la empresa deje de ser la misma unidad jurídica, por lo que, aunque en el repositorio de Oracle se ha mantenido este seguimiento, no siempre ha sido posible hacerlo con los identificadores existentes desde 1998, por lo que se incidirá en crear y reorganizar los identificadores únicos de empresas.

Actualmente la estructura de empresas tiene una estructura del tipo que se muestra a continuación:

SILBE EMPRESAS 



emp_eins	
ABC ein	varchar(4)
123 seq_empresas	int8

emp_empleos	
123 empleo	int4
123 seq_empresas	int8

emp_entidades	
ABC entidad	varchar(100)
123 seq_empresas	int8

emp_escaleras	
ABC escalera	varchar(2)
123 seq_empresas	int8

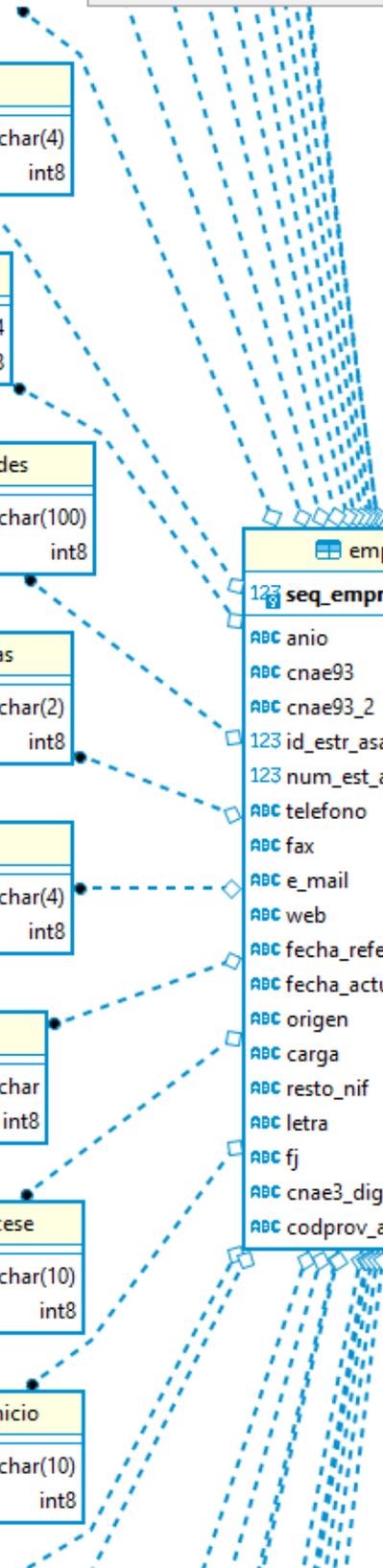
emp_esns	
ABC esn	varchar(4)
123 seq_empresas	int8

emp_estados	
ABC estado	varchar
123 seq_empresas	int8

emp_fechas_cese	
ABC fecha_cese	varchar(10)
123 seq_empresas	int8

emp_fechas_inicio	
ABC fecha_inicio	varchar(10)
123 seq_empresas	int8

empresas20172018h	
123 seq_empresas	int8
ABC anio	varchar(4)
ABC cnae93	varchar(5)
ABC cnae93_2	varchar(5)
123 id_estr_asal	int4
123 num_est_andalucia	int4
ABC telefono	varchar(9)
ABC fax	varchar(9)
ABC e_mail	varchar(100)
ABC web	varchar(100)
ABC fecha_referencia	varchar(10)
ABC fecha_actualizacion	varchar(10)
ABC origen	varchar(2)
ABC carga	varchar(4)
ABC resto_nif	varchar(8)
ABC letra	varchar(1)
ABC fj	varchar(15)
ABC cnae3_dig	varchar(3)
ABC codprov_anda	varchar(2)



5.2 Transformación ficheros Cuentas de Cotización. Conexión ficheros Cuenta propia y Cuenta ajena.

Los ficheros de Cuentas de Cotización se deben transformar para conseguir la unidad empresa, que permita mes a mes ir actualizando la información de las empresas.

Los ficheros de Cuentas de Cotización contienen las cuentas de cotización reales en alta o en baja mensual en las Delegaciones de la Tesorería de la Seguridad Social de empresas que tienen Cuentas de Cotización (CCC) en las Delegaciones de Andalucía, de aquellas empresas que tienen sus CCC en Delegaciones de fuera de Andalucía pero su CCC principal es de Andalucía, así como aquellas CCC de Delegaciones de fuera de Andalucía pero que tienen trabajadores que residen en Andalucía (ej: empresas fronterizas en otras CCAA). Cada cuenta de cotización recoge los trabajadores asociados a la misma. Una empresa puede tener una o más cuentas de cotización asociadas a su NIF. La periodicidad del fichero es mensual desde mayo de 2020, antes de esta fecha se recibía de forma trimestral.

A la hora de agrupar todas las Cuentas de Cotización pertenecientes a un mismo NIF en un solo registro, es necesario realizar una serie de acciones sobre los ficheros de Cuentas de Cotización. Se mencionan a continuación, a modo de muestra, los criterios de agrupación de algunos campos de dicho fichero:

Cuenta de cotización principal. La cuenta de cotización principal se toma del campo “CCC1” o “CCC2”, según corresponda. Si la empresa tiene una única cuenta de cotización, la cuenta de cotización principal será aquella que aparezca en el campo “CCC2”, siempre que dicho valor sea distinto de 0; si el campo “CCC2” es igual a 0 se tomará como principal la cuenta que aparece en “CCC1”. Por otro lado, si la empresa tiene más de una cuenta de cotización, la cuenta de cotización principal es el valor que aparece en el campo “CCC2” distinto de 0.

Actividad principal de la empresa. Se agrupa por División (CNAE a 2 dígitos) y se selecciona aquella División que tenga mayor empleo asociado. Dentro de la División con mayor empleo, se toma aquella clase (CNAE a 4 dígitos) que tenga mayor número de trabajadores. En caso de igualdad de empleo se tomará el menor código de CNAE (el numéricamente inferior).

Actividad secundaria de la empresa. Se toma aquella clase (CNAE a 4 dígitos) con mayor empleo distinta de la actividad principal. En caso de igualdad de empleo se tomará el menor código de CNAE (el numéricamente inferior).

Para obtener la información procedente de los ficheros de Seguridad Social correspondiente a personas físicas, tanto en Cuenta propia y Cuenta ajena, se recurrirá a una conexión con la Base de Datos Longitudinal de Población de Andalucía (BDLPA). Esta conexión, permitirá la actualización de la información relativa a personas físicas sin Cuentas de Cotización asociadas, así como relacionar la información de SILBE con BDLPA.

5.3 Sociedades Mercantiles

La información de Sociedades Mercantiles se obtiene mensualmente, se procesa y carga en WPS (World Programming System), para su explotación en la operación correspondiente de Sociedades Mercantiles. SILBE consultará directamente de las tablas cargadas en WPS, ya que tienen un trabajo de codificación incorporado en CNAE, forma jurídica, etc., que se realiza mensualmente para la publicación de la actividad “Estadística sobre Sociedades Mercantiles de Andalucía”. Los ficheros que se cargan mensualmente en WPS y que sirven a SILBE son:

Constituciones: En este fichero se recogen las constituciones de empresas, que en principio será la primera aparición de la empresa en SILBE. Las principales variables que se recogen son el NIF, la razón social, la forma jurídica, la provincia, el municipio, la CNAE, el capital social, el sexo de los nombramientos en el acta fundacional y la dirección postal.

Extinciones: En este fichero se recoge el NIF y la Razón Social de la empresa, además de su forma jurídica y la fecha de inscripción de la extinción. Se utilizará mensualmente para cruzar con SILBE, de forma que se den de baja las empresas formalmente extinguidas.

Transformaciones: Recoge las transformaciones de forma jurídica de las empresas. Se utilizará para actualizar la Razón Social y el NIF de las empresas recogidas en SILBE (ambos cambian en las transformaciones). El inconveniente es que en la tabla de transformaciones se registra el NIF nuevo, pero no el antiguo, por lo que será necesario localizar la empresa original en SILBE a través del NIF a 8 posiciones o de cadenas de la Razón Social.

Después de estudiar los ficheros de Disoluciones, Fusiones y Escisiones de la operación de Sociedades Mercantiles, se resuelve que la información que contienen se recoge de una forma u otra en los 3 ficheros anteriormente mencionados.

5.4 Registro Mercantil

De las empresas que presentan sus Cuentas Anuales en los Registros Mercantiles de Andalucía, se hacen varias descargas anuales de la información. Son descargas incrementales, con lo que realmente la última descarga es la que aporta toda la información necesaria para el proyecto de Contabilidad Regional. Se descargan en torno a 100.000 Cuentas Anuales de empresas cada año.

Las cuentas anuales se descargan en sus distintos formatos, abreviado, normal, mixto y pyme. Posteriormente dicha información se volcará en SILBE en una tabla que contiene una estructura común a todos los modelos. Se proporciona información de Balance, Resultados, Datos identificativos y Empresa matriz.

Los procesos de carga en SILBE se automatizarán como un subproceso del Sistema de Información, así como la descarga de la información.

5.5 Servicio web CDAU y Variantero

Dentro de un proyecto como SILBE, la parte de normalización de viales cobra gran relevancia y sirve, entre otros aspectos, de nexo de unión con la vertiente territorial. Esta tarea se va a realizar a través del Servicio web del Callejero Digital de Andalucía Unificado (CDAU). En caso de que no se pueda codificar la dirección se lanzará contra un Variantero de direcciones que se ha ido generando a lo largo de los años de codificación de viales en el IECA.

Por tanto la normalización se lanzará contra el Servicio web de forma que en primer lugar se codifique con CDAU (hueco, portal, vía), posteriormente con el Variantero cruzando con variantes de denominación de las vías (si no se ha conseguido ninguna coincidencia con CDAU) y en último lugar se generaría una tabla de incidencias con lo no codificado que se analizará para decidir si tienen que alimentar o no el Variantero.

SILBE ha servido para que el Servicio web de CDAU se plantee como algo transversal dentro del IECA y como un servicio que se ponga a disposición de usuarios externos.

Todas estas tareas de normalización de viales también conducen a mejorar y completar las direcciones postales de los establecimientos ubicados en el mapa de geolocalización del Directorio de establecimientos de más de 50 empleados.

6. Tareas pendientes de realización

Anteriormente se ha descrito el proyecto y las tareas que se están realizando actualmente. Tal y como se ha señalado, al ser un proyecto a medio plazo, se irán abordando el resto de tareas de forma secuencial. Las principales tareas que están pendientes de abordar son:

- Seguir definiendo la carga y actualización de las distintas fuentes de actualización y consulta que quedan pendientes de afrontar. En estas tareas van implícitas aquellas tareas consideradas como transversales en el proyecto: agrupación de unidades para cruces, claves únicas, normalización de ficheros y direcciones, ...
- Implementar para cada campo considerado como histórico la lógica de la jerarquía de actualización.
- Definir la conexión y procedimientos de consulta con la Base de Datos Longitudinal de Población de Andalucía.
- Concretar las conexiones con otros proyectos (ej: Contabilidad Regional de Andalucía) para automatizar en lo posible los intercambios de información.
- Crear consultas genéricas que faciliten afrontar las peticiones habituales de información de empresas con un menor coste de recursos y en un menor plazo de tiempo.

Además de las tareas anteriormente mencionadas, se espera que vayan surgiendo tareas adicionales derivadas de la implementación de las mismas.