

**Sistema de Identificación para unidades
estadísticas complejas**

**Metodología e implementación en el
Directorio Central de Empresas (DIRCE)**

Esteban Barbado Miguel
Pedro García Segador
Pilar Montero Robles
Valentín Llorente García
Mirian Hernandez Valencia

Departamento de Metodología y Desarrollo
de la Producción estadística

Unidad DIRCE

El Instituto Nacional de Estadística no se identifica necesariamente con las opiniones
formuladas por los autores en este informe

Marzo 2021

Sistema de Identificación para unidades estadísticas complejas

Metodología e implementación en el Directorio Central de Empresas (DIRCE)

En los últimos años, la arquitectura del Directorio Central de Empresas (DIRCE) ha incorporado a los grupos empresariales y las empresas como nuevas unidades estadísticas que son consistentes con el resto de las que ya estaban incluidas en el DIRCE como unidades legales y locales.

A nivel global, los grupos de empresas no suelen estar dotados de personalidad jurídica. Tampoco lo son las empresas compuestas por agrupaciones de unidades legales siguiendo métodos puramente estadísticos. Esto implica la no disponibilidad de un identificador legal.

Esta situación limita los análisis de la vida de estas unidades a lo largo del tiempo y afectan especialmente a la producción de Indicadores de Demografía Empresarial. Es necesario desarrollar nuevos procesos orientados al diseño de un Sistema de Identificación robusto para unidades estadísticas complejas.

En este documento se describe la construcción de este identificador y la gestión dinámica a lo largo del tiempo de identificadores para grupos empresariales y empresas operando en su seno.

Palabras clave: Unidades estadísticas, grupos empresariales, empresas, identificador, seguimiento a lo largo del tiempo

Identification system for complex statistical units

Methodology and implementation in the National Business Register (DIRCE)

In recent years, the architecture of the National Business Register (DIRCE) has incorporated enterprise groups and enterprises as new statistical units that are consistent with the rest of those that were already included in the DIRCE as legal and local units.

On a global level, enterprise groups and enterprises (groups of legal units formed following purely statistical methods) do not usually have legal personality. This implies the unavailability of a legal identifier.

This situation increases the difficulties in the analysis of the life of these units over time and especially affects the production of Business Demographic Indicators. It is necessary to develop new processes aimed at the design of a robust Identification System for complex statistical units.

This document describes the construction of this identifier and the dynamic management over time of identifiers for enterprise groups and enterprises operating within them.

Key words: Statistical units, enterprise groups, enterprises, identifier, follow up over time

Sistema de Identificación para unidades estadísticas complejas

**Metodología e implementación en
el Directorio Central de Empresas
(DIRCE)**

Contexto general

En los últimos años, la arquitectura del Directorio Central de Empresas (DIRCE) ha experimentado una profunda reforma, motivada por la incorporación de nuevas unidades estadísticas y su necesaria consistencia con el resto de niveles de información.

El modelo de trabajo adoptado para conformar los *grupos empresariales* y las *empresas complejas* que operan en su seno, puede considerarse implementado¹, al garantizarse procesos estables, totalmente integrados en el ciclo de producción y outputs útiles para los usuarios.

A nivel global, los grupos de empresas no suelen estar dotados de personalidad jurídica; esto implica la no disponibilidad de un identificador legal que permita el uso de información de forma compartida entre diferentes Organizaciones y países.

Las técnicas de *profiling* se ejecutan con base a referencias temporales determinadas y bajo una perspectiva anual. El seguimiento sistemático de las unidades delineadas con el paso del tiempo no es una cuestión suficientemente tratada desde el punto de vista metodológico.

En consecuencia, las limitaciones para abordar análisis dinámicos son claras y afectan especialmente a la producción de Indicadores de Demografía Empresarial. Es necesario desarrollar nuevos procesos orientados al diseño de un Sistema de Identificación robusto para unidades estadísticas complejas.

El primer capítulo de este documento describe el método desarrollado. Se proporcionan los fundamentos en los que se basa la generación y gestión dinámica de identificadores para grupos empresariales y empresas operando en su seno.

En el segundo capítulo se presenta la operativa vinculada a las variables, su implementación en el DIRCE y los principales resultados obtenidos tras sucesivos ciclos de producción.

¹ Delineación de empresas complejas en el INE. Metodología del algoritmo de Profiling automático.

Capítulo 1

Producción y gestión dinámica de identificadores

1. Base conceptual

La disponibilidad de un Sistema de Identificación es una actividad nuclear en la gestión de un registro estadístico de empresas. La propia arquitectura del registro, con múltiples niveles de información interconectados y consistentes entre sí, hace extremadamente necesario tener un procedimiento sólido y bien documentado.

Cuando las unidades estadísticas son simples, tienden a asimilarse a las unidades legales que habitualmente aparecen en los registros administrativos. En estos casos, se puede adoptar sin problema el identificador administrativo a efectos estadísticos. La propia dinámica de gestión de estos identificadores externos puede ser incorporada al registro estadístico en los ciclos de producción.

El problema aparece con las unidades estadísticas complejas, conformadas mediante combinaciones de unidades legales o partes de estas, en función de diferentes criterios: *control* para los grupos empresariales y *elementos de profiling* para las empresas complejas. Con el paso del tiempo, las unidades legales cambian sustancialmente, afectando a la composición o *cluster* preestablecido en un año base.

Un Sistema de Identificación robusto debe satisfacer la siguiente propiedad básica: el identificador debe persistir en el registro mientras la unidad estadística continúe viva. En otras palabras, el sistema debe ofrecer la respuesta a la siguiente cuestión: observadas dos unidades estadísticas en dos puntos del tiempo, ¿Cuándo podemos afirmar sin ambigüedad que son la misma o distintas?

Es esencial definir criterios de *continuidad estadística*, que permitan gestionar de forma inequívoca la asignación de identificadores a las unidades desde una perspectiva temporal. En esta fase, hay que considerar adicionalmente la relación jerárquica existente entre los dos tipos de unidades: grupos y empresas complejas. Este hecho es fundamental a la hora de adoptar el formato adecuado para los identificadores.

Los manuales internacionales sobre registros estadísticos de empresas asimilan los criterios de continuidad estadística a la continuidad de los factores de producción. Con esta base de partida, hay ya algunas experiencias desarrolladas y bien fundamentadas, particularmente por el INSEE francés. La mayor parte de sus criterios han sido ligeramente perfilados e incorporados a nuestra metodología.

Definir aproximaciones operativas a la continuidad de los factores de producción, significa buscar conexiones razonables con la información disponible, tratando de evitar criterios redundantes o que potencialmente puedan dificultar el algoritmo de gestión del identificador. En este sentido, se han tomado como elementos de análisis: el conjunto de las diferentes *unidades legales* que conforman cada unidad estadística, su *jerarquía* y el *empleo* registrado para cada una de ellas. Con base a todos estos principios, nuestra metodología de trabajo se ha estructurado en las siguientes fases:

- Definición de reglas de continuidad para grupos empresariales y empresas complejas

- Adopción de formatos para los identificadores. En lo sucesivo, se denominarán ID_GROUP e ID_ENT
- Identificación de tipos de sucesos demográficos asociados a la gestión de identificadores, con vistas a su pertinente utilización en *Demografía Empresarial*
- Implementación del proceso; generación de un primer conjunto de identificadores para un año base t del DIRCE.
- Aplicación de las condiciones de continuidad y monitorización del Sistema de Identificación en sucesivos ciclos de producción
- Análisis de resultados, robustez del procedimiento y sostenibilidad en el tiempo

2. Reglas de continuidad para grupos

Consideremos dos grupos de empresas observados en dos años consecutivos, t y t+1:

- Si tienen la misma matriz (unidad situada en el TOP de la cadena de control), diremos que el grupo continúa si se verifica

$$\frac{\text{Empleo comun}}{\text{Empleo total}_{t+1}} \geq 0,5 \qquad \frac{\text{Empleo comun}}{\text{Empleo total}_t} \geq 0,5$$

Siendo:

Empleo común: Asalariados registrados en las unidades legales comunes en los dos años

Empleo total $_{t+1}$: Asalariados del grupo en t+1

Empleo total $_t$: Asalariados del grupo en t

- Si la matriz es diferente, el grupo continúa si se verifica

$$\frac{\text{Empleo comun}}{\text{Empleo total}_{t+1}} \geq 0,7 \qquad \frac{\text{Empleo comun}}{\text{Empleo total}_t} \geq 0,7$$

Obviamente, en ambos casos debe detectarse la presencia común de alguna unidad legal en los dos años de observación. Estas reglas se procesan de forma jerárquica, chequeándose en el orden presentado. Esto es, si no se detecta continuidad por el primer criterio, se pasa a evaluar el segundo. El cambio de matriz hace en principio menos verosímil la continuidad de un determinado grupo y por esta razón se ha elevado la cota de persistencia del factor empleo.

Las reglas no están bien definidas cuando $\text{Empleo total}_t = 0$ ó $\text{Empleo total}_{t+1} = 0$. Aunque no es un caso común e involucra a unidades muy poco significativas, es necesario resolver estas indeterminaciones para el correcto proceso del algoritmo.

Tomemos un par de grupos (G1, G2) donde la primera componente se refiere a una observación de t y la segunda, de t+1. Definimos la propiedad: *par de enlace único*.

(G1, G2) es de enlace único si entre los grupos de t+1, sólo G2 tiene unidades legales comunes con G1. Entonces:

- Si $Empleo\ total_t = 0$ para un grupo G1 perteneciente a un par de enlace único entonces, se asume continuidad con G2
- Si $Empleo\ total_t = 0$ para un grupo G1 que no forma parte de un par de enlace único, se asume pérdida de continuidad para G1
- Si $Empleo\ total_{t+1} = 0$ para un grupo G2 perteneciente a un par de enlace único entonces se asume continuidad con G1
- Si $Empleo\ total_{t+1} = 0$ para un grupo G2 que no forma parte de un par de enlace único, entonces se considera nuevo grupo.

Una vez aplicadas las condiciones de continuidad, el tratamiento de los flujos se establece de la siguiente manera:

- Los grupos activos de t+1 que no pasen las reglas, se consideran *entradas* en el DIRCE y llevan asignado un nuevo identificador con un status de actividad específico.
- Los grupos activos de t que no pasen las reglas, se consideran *salidas* del DIRCE y son grabados en t+1 con un status de actividad específico

3. Reglas de continuidad para empresas

El conjunto de criterios para empresas se ha definido de forma similar, pero considerando además que la existencia de una empresa compleja es inherente a la del grupo en el que opera. En este sentido, si el grupo (el todo) deja de existir, sus empresas (las partes) también. El conjunto de reglas queda formulado de la siguiente manera:

- Todas las empresas pertenecientes a entradas de grupos en t+1 serán también consideradas como entradas en ese año; un nuevo identificador será asignado.
- Todas las empresas pertenecientes a salidas de grupos activos en t, serán también consideradas como salidas de empresas; de forma similar al apartado anterior, la información se replica en t+1.
- Las empresas operando en grupos que continúan, son analizadas con base a la continuidad del empleo en unidades legales comunes. Si se verifica:

$$\frac{Empleo\ comun}{Empleo\ total_{t+1}} \geq 0,5 \qquad \frac{Empleo\ comun}{Empleo\ total_t} \geq 0,5$$

Entonces la empresa persiste y por tanto, mantiene su identificador. Esta propiedad se debe comprobar para cada par de empresas que tengan unidades legales comunes. En caso de discontinuidad, se genera un identificador nuevo

De forma similar al caso de grupos, la regla no está bien definida cuando $Empleo\ total_t = 0$ ó $Empleo\ total_{t+1} = 0$. Entonces se vuelve a aplicar la propiedad de enlace único con los mismos criterios.

Los tratamientos de flujos y su caracterización, tanto para grupos como para empresas, se detalla en el apartado 5.

4. Formatos adoptados para los Identificadores

Las directrices internacionales recomiendan evitar la presencia de cualquier *rasgo* de la unidad estadística en el identificador. En efecto, una determinada característica puede cambiar sin afectar a la componente existencial de la unidad. Este movimiento generaría evidentes problemas en la gestión del identificador si dicha característica está insertada en la variable.

En términos operativos, este es precisamente el propósito de las reglas de continuidad; examinar cambios que no tienen trascendencia en la identidad de las unidades y mantener el identificador. Partiendo de esta base:

- La estructura del identificador de grupo (ID_GROUP) se ha conformado a partir de una letra inicial, 5 dígitos numéricos y un carácter de control final: **L00000A**. La primera letra y los 5 dígitos se generan de forma correlativa. El carácter alfabético de control se genera aplicando la regla del módulo 23, de forma idéntica a los NIF de persona física.
- El identificador de empresa (ID_ENT) contiene el respectivo del grupo en el que opera, añadiendo un número correlativo de 3 posiciones: **L00000A000**.

Si un grupo llegara a sobrepasar las 1.000 empresas, caso altamente improbable, el Sistema de Identificación prevé la introducción de letras combinadas con números, creando un orden que llega a magnitudes prácticamente imposibles de alcanzar ($000 < 001 < \dots < 998 < 999 < 00A < 00B < \dots < 99Z < \dots < A99 < A9A < \dots < A9Z < AZZ < B00 < B01 < \dots < BZZ < \dots < ZZZ$). Este procedimiento permite identificar hasta $(26 + 10)^3 = 46656$ empresas de un mismo grupo con el paso del tiempo.

El control y gestión de estos identificadores se apoya en la creación de repositorios separados para cada tipo de unidad estadística, aunque obviamente conectados. De hecho forman una parte nuclear del propio sistema DIRCE. En cada ciclo de producción, ambos repositorios son actualizados, en línea con los marcos finales de unidades estadísticas y conteniendo el histórico de cada código generado. Ninguno de estos identificadores será reciclado, esto es, si un grupo/empresa sale del DIRCE, en el futuro no existirá ninguna otra unidad con ese identificador.

5. Tipologías de sucesos y conexiones con la Demografía Empresarial

El interés por los indicadores sobre dinamismo empresarial ha crecido sustancialmente en los últimos años. Son una base esencial para el seguimiento de aspectos como la innovación, la competitividad, el desarrollo económico y la creación de empleo en las economías nacionales. Esta información se usa en múltiples ámbitos, desde políticas sobre planificación económica, hasta estudios de investigación, análisis de mercados, etc.

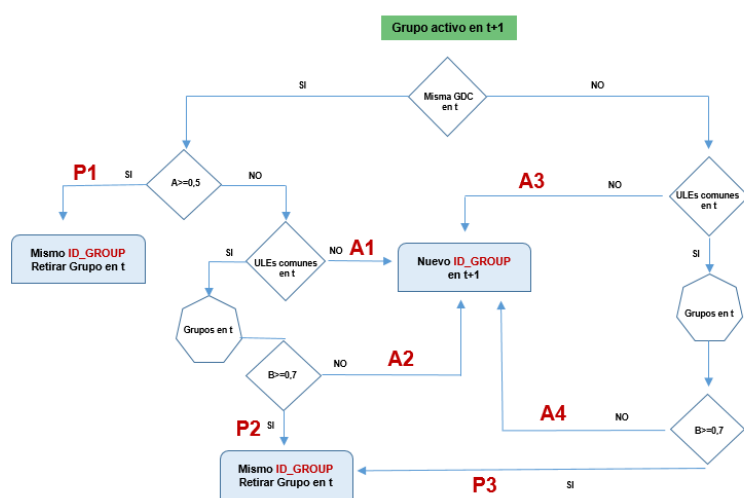
Los registros de empresas para usos estadísticos juegan un papel central en la producción de estos indicadores. Su capacidad informativa contribuye a reducir significativamente los costes e incrementar la eficiencia en la producción de datos en este ámbito. En este apartado se caracterizan determinados movimientos asociados a unidades complejas, con incidencia directa en la persistencia o discontinuidad de los identificadores. El propósito es facilitar los análisis dinámicos del DIRCE, buscando grados de adecuación razonable a los sucesos de

nacimiento / muerte / supervivencia de empresas descritos en las directrices metodológicas internacionales sobre Demografía Empresarial².

La aplicación de las condiciones de continuidad en dos puntos del tiempo permite inequívocamente decidir cuándo una unidad es la misma o no. En caso de discontinuidad, las nuevas unidades pueden o no tener una cierta *huella* en el año precedente. Desde el punto de vista estadístico, esta propiedad es clave pues significa que parte de los factores de producción ya preexistían o que por el contrario, son totalmente nuevos. Lo mismo se puede argumentar para las unidades activas en un año determinado y que pierden su identidad al año siguiente. Todos estos aspectos son detectados durante en el funcionamiento del Sistema de Identificación, generándose una tipología de sucesos que se describen en los siguientes esquemas:

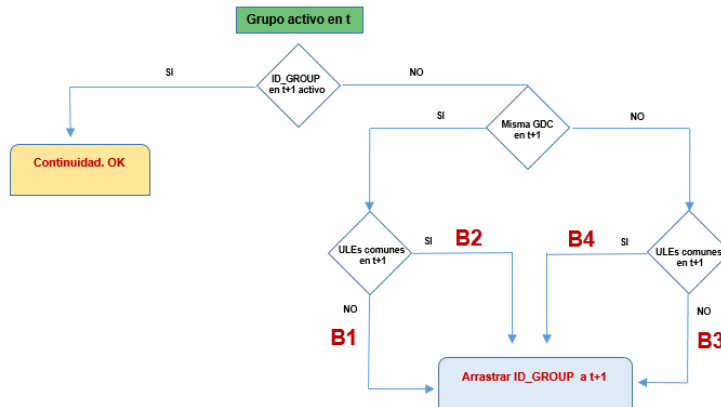
Grupos de empresas

Asignación ID_GROUP en t+1; Entradas y Permanencias



² Eurostat-OECD Manual on Business Demography Statistics // Guidelines on the use of SBR for Business Demography and Entrepreneurship Statistics

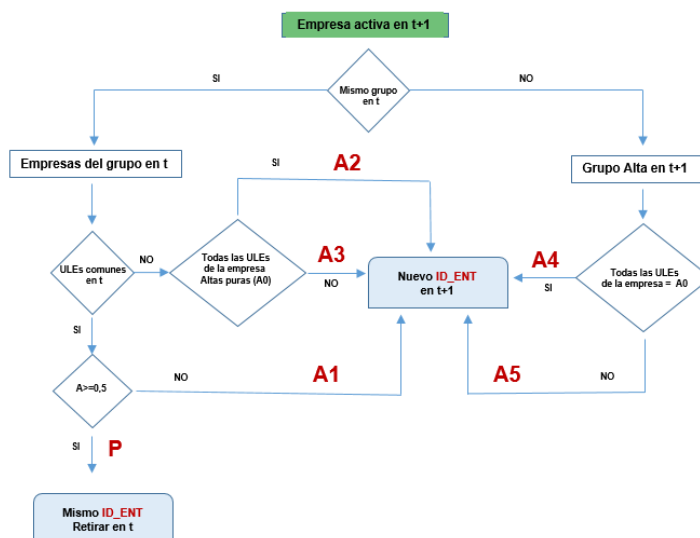
Asignación ID_GROUP en t+1; Bajas



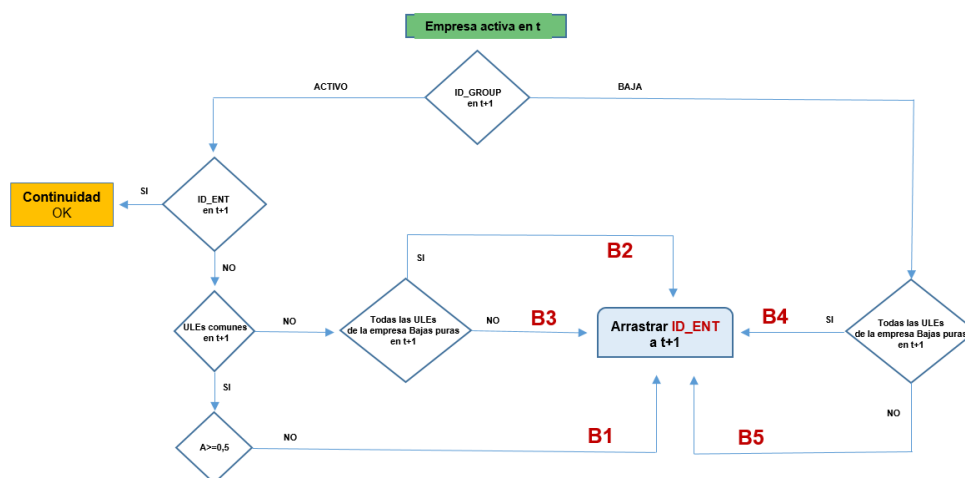
SUCESO	Descripción
Continuidad	
P1	Grupos de t+1 con la misma GDC en t y superan el umbral (>0.5)
P2	Grupos de t+1 con diferente GDC en t y superan el umbral (>0.7). Hay otro grupo con la misma GCD en t pero por debajo del umbral (>0,5)
P3	Grupos de t+1 con diferente GDC en t y superan el umbral (>0.7). No hay otro grupo con la misma GCD en t
Entradas	
A1	Grupos de t+1 con la misma GDC y sin filiales comunes en t
A2	Grupos de t+1 con la misma GDC y filiales comunes en t
A3	Grupos de t+1 y sin ninguna unidad legal detectada en t
A4	Grupos de t+1 con GDC no detectada pero con filiales comunes en t
Salidas	
B1	Grupos de t con la misma GDC detectada en t+1 pero sin filiales detectadas en t+1
B2	Grupos de t con GDC y alguna filial detectada en t+1
B3	Grupos de t sin ninguna unidad legal detectada en t+1
B4	Grupos de t sin GDC detectada en t+1 pero alguna filial detectada en t+1

Empresas

Asignación ID_ENT en t+1; Entradas y Permanencias



Asignación ID_ENT en t+1; Bajas



SUCESO	Descripción
Continuidad	
P	Empresas de t+1 en grupos que sobreviven y superan el umbral (>0.5)
Entradas	
A1	Empresas de t+1 en grupos que sobreviven, con filiales comunes con alguna empresa del grupo en t, pero por debajo del umbral ($\alpha > 0.5$)
A2	Empresas de t+1 en grupos que sobreviven pero sin filiales detectadas en el grupo en t. Todas sus unidades legales son entradas puras en t+1
A3	Empresas de t+1 en grupos que sobreviven pero sin filiales detectadas en el grupo en t. Algunas unidades legales no son entradas puras en t+1
A4	Empresas de t+1 en grupos nuevos de t+1. Todas sus unidades legales son entradas puras en t+1
A5	Empresas de t+1 en grupos nuevos de t+1. Algunas unidades legales no son entradas puras en t+1
Salidas	
B1	Empresas de t en grupos que sobreviven, con filiales comunes con alguna empresa del grupo en t+1, pero por debajo del umbral ($\alpha > 0.5$)
B2	Empresas de t en grupos que sobreviven, sin filiales comunes. Todas sus unidades legales son bajas puras en t+1
B3	Empresas de t en grupos que sobreviven, sin filiales comunes. alguna de sus unidades legales no es baja pura en t+1
B4	Empresas de t en grupos que salen en t+1. Todas sus unidades legales son bajas puras en t+1
B5	Empresas de t en grupos que salen en t+1. alguna de sus unidades legales no es baja pura en t+1

Esta tipología de sucesos enriquece la información disponible en el DIRCE y facilita su uso para estudios sobre dinamismo empresarial. Las metodologías disponibles en este campo inciden en elementos que afecten a la propia existencia de la unidad estadística o a los vínculos existentes entre ellas. Operativamente, los impactos se miden en términos de cambios en (combinaciones de) factores de producción, lo cual conecta de forma directa con el Sistema de Identificación implementado. Cuando no se detectan cambios o se alteran ligeramente los factores de producción, estaríamos ante un fenómeno de persistencia o continuidad de la unidad estadística.

Esta idea adquiere especial valor cuando se analizan los grupos de empresas, involucrados casi de forma continua en procesos de reestructuración y generando múltiples cambios de unidades legales. No todos estos flujos van a implicar necesariamente cambios sustanciales en los factores de producción.

La conexión entre estas nuevas categorías de sucesos y los conceptos utilizados en *Demografía Empresarial*, se esboza a continuación:

- **Un nacimiento de empresa** corresponde a la creación de nuevos factores de producción. Esto implica que no hay ninguna otra empresa involucrada en este

evento. Por tanto, entradas debidas a fusiones, escisiones o reestructuraciones internas de un grupo no se consideran nacimientos estadísticos.

Una aproximación razonable al concepto de nacimiento de empresa compleja estaría conformada por la población de unidades clasificadas en A2 y A4. Respecto al resto de entradas, previsiblemente habrá que incorporar algún criterio alternativo, quizás basado en el empleo de sus unidades legales según su clave de actividad o similar, para delimitar el conjunto final de nacimientos.

- La evolución temporal de las cohortes de nacimientos de empresas es otro fenómeno de gran interés. **Una empresa nacida en t sobrevive en $t+k$** si continúa activa en este periodo de observación. La identificación de supervivencias se realiza año a año y por un periodo de cinco años.

Conceptualmente, este suceso implica continuidad de los factores de producción, esto es, o no hay variación o *apenas* hay variación entre los dos puntos de observación. En el caso particular de empresas complejas, el concepto de supervivencia contempla posibles variaciones en las unidades que las conforman o sus características, pero con la condición de que los cambios detectados no sean significativos. Este factor de *no relevancia* lo determinan las reglas de continuidad y la consiguiente persistencia del Identificador en el DIRCE.

- **Las muertes de empresas** corresponden a la disolución de una combinación de factores de producción preexistentes. No hay ninguna otra empresa involucrada en este evento y por tanto no deben contabilizarse las salidas debidas a fusiones, absorciones, escisiones o reestructuraciones internas de un grupo.

Una aproximación razonable al concepto de muerte de empresa compleja estaría conformada por la población de unidades clasificadas en B2 y B4. Respecto al resto de *salidas* y de forma análoga al caso de nacimientos, deberá incorporarse algún criterio alternativo para delimitar el conjunto final de muertes.

- El resto de movimientos detectados por el Sistema de Identificación serán considerados como *entradas* o *salidas* de unidades, pero no asimilables a los conceptos estadísticos de nacimiento o muerte.

Capítulo 2

La operativa en el sistema DIRCE; principales resultados

En este capítulo se aborda la incorporación de las variables **ID_GROUP** e **ID_ENT** al DIRCE y se proporcionan algunos resultados obtenidos tras su monitorización durante varios años. Asociadas a los identificadores, se implementan nuevas variables de gestión referidas al status de actividad y el año de inicio de actividades. Estas características son un requisito explícito de la reglamentación europea vigente y que persistirá con el escenario EBS.

El método utilizado ha sido consecuencia del análisis y toma de decisiones respecto a los siguientes elementos:

- El primer *año de referencia* de aplicación
- El *momento del ciclo* en que se incorpora el Sistema de Identificación y sobre *qué bases de datos*
- La meta información asociada a las nuevas *variables* y las *estructuras* de las bases de datos

Inicialmente es imposible tener en cuenta todas las implicaciones del nuevo procedimiento: interacciones con el EGR, Demografía Empresarial, muestras para los diversos productores, etc. No obstante, es necesario dar el paso en algún momento, con la visión más amplia posible y con la idea de que previsiblemente, en el futuro serán necesarios ajustes.

1. El año de comienzo

Una vez identificadas las necesidades y requisitos para la producción de Demografía Empresarial, se tomó como año base el DIRCE de referencia 2016. El motivo principal es que la Demografía de Empleadores de 2018 necesita disponer de información retrospectiva sobre empleo de las empresas en los dos años precedentes.

Durante el citado ciclo se han creado los repositorios de identificadores para ambos tipos de unidades junto con las fechas de Alta y Baja asociadas a los mismos.

El primer seguimiento temporal y la consiguiente aplicación de los criterios de continuidad se ejecutó en el ciclo *DIRCE 2017*. A partir de este año, la población de grupos y empresas complejas ya puede ser clasificada según los nuevos eventos demográficos.

2. El momento de implementación en el ciclo

Tomar decisiones acertadas sobre *momento* es un factor crítico en cualquier proceso de reforma. En el caso particular del DIRCE, es necesario además que exista una interacción viable con los procesos habituales de producción, ya de por sí complejos.

En el análisis preliminar, se determinó que las posibilidades de éxito se incrementarían si la nueva operativa se incorpora en las partes finales del ciclo y sobre marcos definitivos. Otro tipo de escenario generaría una errática y complejidad añadida difícil de asumir, máxime cuando el tiempo de actualización es ya muy limitado dados los requerimientos de calendario, especialmente para la transmisión de información al *Euro Groups Register* (EGR).

Al final del proceso, toda la información producida debe ser completamente consistente, tanto desde el punto de vista estático como dinámico. Esto implica que, por ejemplo,

cualquier ID_ENT generado en el año t debe tener su correspondiente información en t+1 y ser además coherente con la información del año precedente.

Considerando todo lo anterior, se ha tomado como base de datos de partida el marco final de grupos y empresas complejas del DIRCE, obtenidas tras la aplicación de las técnicas de Profiling.

3. Las variables a implementar

El Sistema de Identificación genera y mantiene el siguiente conjunto de variables:

3.1 Variables para grupos

- **ID_GROUP**.- El identificador, de acuerdo a la metodología establecida
- **STATUS_GROUP**.- Status de actividad del grupo, tras la aplicación de los criterios de continuidad.
 - A1, A2, A3, A4. *Entradas* en el año de referencia, implicando la generación de nuevos identificadores
 - P1, P2, P3. *Permanencias*. Los identificadores persisten para el año de referencia
 - B1, B2, B3, B4, B5. *Salidas* en el año de referencia; los identificadores también persisten
- **AÑO_GROUP**.- Año de detección del grupo en el DIRCE. En función de su STATUS, tendríamos
 - $AÑO_GROUP = t$ si $STATUS_GROUP = A_i$
 - $AÑO_GROUP =$ Valor del año anterior, en cualquier otro caso.

Excepcionalmente para el año base, estas variables se inicializan con información contenida en características ya disponibles.

3.2 Variables para empresas

- **ID_ENT**.- El identificador, de acuerdo a la metodología establecida
- **STATUS_ENT**.- Status de actividad de la empresa, tras la aplicación de los criterios de continuidad.
 - A1, A2, A3, A4, A5. *Entradas* en el año de referencia, implicando la generación de nuevos identificadores
 - P. *Permanencias*. Los identificadores persisten para el año de referencia
 - B1, B2, B3, B4, B5. *Salidas* en el año de referencia; los identificadores también persisten
- **AÑO_ENT**.- Año de detección de la empresa en el DIRCE. En función de su STATUS, tendríamos
 - $AÑO_ENT = t$ si $STATUS_ENT = A_i$
 - $AÑO_ENT =$ Valor del año anterior en cualquier otro caso.

Excepcionalmente para el año base, estas variables se inicializan con información similar ya contenida en características disponibles.

4. La estructura de los datos y su monitorización temporal

La solución adoptada ha consistido en replicar la información disponible en la base de datos de partida, pero incorporando las nuevas variables. Al menos para los primeros ciclos, se han preservado las variables originales, aunque esta base informativa irá perdiendo relevancia, hasta su eliminación definitiva en el momento oportuno. La nueva estructura fue generada por primera vez para el año base, creándose adicionalmente los repositorios de identificadores. La verificación de consistencia completa entre todos los ficheros, cerró los trabajos para este primer ciclo.

Con esta primera referencia, el sistema generó los identificadores para el 2017, aplicando las reglas de continuidad. En el mismo proceso, se actualizaron los repositorios y se aplicó de nuevo el análisis de consistencia. Esta misma dinámica se replica en sucesivos ciclos de producción del DIRCE. Veamos a continuación los principales agregados producidos y su evolución con el paso del tiempo.

Resultados año base: 2016

Tomando de partida los resultados del Profiling obtenidos para ese año, hay identificadas **92.177 unidades legales activas en clusters** de empresas operando en grupos.

Es necesario considerar de antemano que para grupos multinacionales, la información disponible en el DIRCE es la parte truncada nacional. Esto es, si una multinacional desarrolla actividades en España únicamente a través de una filial, esta unidad se considera y contabiliza como *grupo*.

El sistema ha generado por primera vez identificadores para **45.517 empresas complejas** operando en **31.286 grupos**.

La población de unidades estadísticas complejas se estructura de acuerdo a la información de las siguientes tablas, considerando como variable el *estrato de unidades* de rango inferior.

Estrato ULEs	GRUPOS	Estrato ENTs	GRUPOS	Estrato ULEs	EMPRESAS
Total	31.286	Total	31.286	Total	45.517
1 ULE	6.635	1 ENT	24.820	1 ULE	21.626
2 ULEs	15.175	2 ENTs	3.395	2 ULEs	16.116
3 a 5 ULEs	7.119	3 a 5 ENTs	2.492	3 a 5 ULEs	6.468
6 a 9 ULEs	1.477	6 a 9 ENTs	449	6 a 9 ULEs	867
10 a 19 ULEs	621	10 a 19 ENTs	117	10 a 19 ULEs	308
20 y más ULEs	259	20 y más ENTs	13	20 y más ULEs	132

Seguimiento en sucesivos periodos

Las siguientes tablas dan una estructura general de las poblaciones tipo Stock correspondientes a los distintos tipos de unidades, clasificadas según los nuevos eventos demográficos. Todos ellos han sido identificados y monitorizados en los siguientes ciclos de producción del DIRCE, desde 2017 hasta la última actualización disponible, de referencia 2019. El stock de unidades legales se ha distribuido en función de los eventos de sus empresas correspondientes.

Año 2017

Suceso	GRUPOS	Suceso	EMPRESAS	Suceso	ULES
Total	37.296	Total	56.117	Total	113.327
Entradas	6.010 (16.1%)	Entradas	10.600 (18.9%)	Entradas	21.389 (18.9%)
A1	67	A1	2.126	A1	5.260
A2	679	A2	272	A2	330
A3	4.783	A3	879	A3	1.053
A4	481	A4	964	A4	1.235
Salidas	4.410 (11.8%)	A5	6.359	A5	13.511
B1	63	Salidas	7.983 (14.2%)	Salidas	16.463 (14.5%)
B2	683	B1	1.656	B1	4.309
B3	3.057	B2	246	B2	257
B4	607	B3	713	B3	882
Permanencias	26.876 (72.1%)	B4	465	B4	631
P1	25.622	B5	4.903	B5	10.384
P2	78	Permanencias	37.534 (66.9%)	Permanencias	75.475 (66.6%)
P3	1.176	P	37.534	P	75.475

Año 2018

Suceso	GRUPOS	Suceso	EMPRESAS	Suceso	ULES
Total	42.458	Total	62.664	Total	124.163
Entradas	9.572 (22,5%)	Entradas	14.530 (23,2%)	Entradas	25.933 (20.9%)
A1	94	A1	2.020	A1	4.636
A2	708	A2	291	A2	323
A3	8.179	A3	1088	A3	1.248
A4	591	A4	1268	A4	1.682
Salidas	4.347 (10.3%)	A5	9.863	A5	18.044
B1	90	Salidas	8.545 (13.6%)	Salidas	17.431 (14.0%)
B2	743	B1	2.104	B1	4.455
B3	2.830	B2	238	B2	250
B4	684	B3	702	B3	834
Permanencias	28.539 (67,2%)	B4	465	B4	576
P1	27.050	B5	5.036	B5	11.316
P2	107	Permanencias	39.589 (63.2%)	Permanencias	80.799 (65.1%)
P3	1.382	P	39.589	P	80.799

Año 2019

Suceso	GRUPOS	Suceso	EMPRESAS	Suceso	ULES
Total	47.038	Total	68.735	Total	138.120
Entradas	8.927 (19.0%)	Entradas	14.616 (21.3%)	Entradas	28.616 (20.7%)
A1	108	A1	2.419	A1	5.640
A2	812	A2	356	A2	437
A3	7.427	A3	1.275	A3	1.479
A4	580	A4	1.402	A4	1.797
Salidas	5.542 (11.8%)	A5	9.164	A5	19.263
B1	100	Salidas	9.541 (13.9%)	Salidas	17.835 (12.9%)
B2	839	B1	2.189	B1	4.711
B3	3.934	B2	283	B2	298
B4	669	B3	551	B3	747
Permanencias	32.569 (69.2%)	B4	652	B4	936
P1	31.035	B5	5.866	B5	11.143
P2	113	Permanencias	44.578 (64.8%)	Permanencias	91.669 (66.4%)
P3	1.421	P	44.578	P	91.669

El **86%** de los grupos inicialmente identificados en el año base, persisten activos en el año siguiente. La tasa de supervivencia presenta buen grado de estabilidad para los sucesivos periodos de observación, siendo de un **86,8%** (supervivencias a 1 año en 2018) y de un **85,5%** (supervivencias a 1 año en 2019).

Con relación a los movimientos de entradas y salidas, la tasa neta de crecimiento de grupos en 2017 es del **4,3%**. Este indicador se calcula dividiendo el saldo neto (entradas – salidas) entre el stock poblacional. Crece significativamente para 2018, situándose en un **12,3%** y un **7,2%** para el año de referencia 2019. El incremento experimentado en 2018 obedece a una mayor presencia de filiales y grupos en nuestra fuente nuclear (D&B) utilizada en el mantenimiento del sistema de grupos del DIRCE. No es por tanto, directamente atribuible al proceso de gestión de identificadores.

Las tasas de supervivencia para empresas complejas son similares a las de grupos, aunque algo más bajas. El **82,5%** de las empresas activas del año base sobreviven en 2017. Para los dos años sucesivos los indicadores fueron **82,2%** y **82,4%** respectivamente. Es lógico que las supervivencias para empresas sean algo inferiores, dada la jerarquía existente entre ambas unidades y las correspondientes reglas de continuidad definidas.

Con relación al stock de filiales, el DIRCE refleja incrementos significativos con el paso del tiempo. Esta circunstancia es el resultado de mejoras de cobertura en los perímetros legales de los grupos, y en consecuencia, mejor adecuación a la información publicada por los propios grupos en sus informes anuales.

Las siguientes tablas ofrecen información demográfica resumida para cada unidad estadística, incorporando el *estrato de asalariados como indicador de tamaño*.

Año y Suceso	Estrato de asalariados del GRUPO				
	Total	Menos de 9	De 10 a 249	De 250 a 999	De 1.000 ó más
2017	37.296	16.593	17.628	2.227	848
Entradas	6.010	3.407	2.396	166	41
Salidas	4.410	2.520	1.699	151	40
Permanencias	26.876	10.666	13.533	1.910	767
2018	42.458	19.482	19.660	2.380	936
Entradas	9.572	5.598	3.627	262	85
Salidas	4.347	2.433	1.715	156	43
Permanencias	28.539	11.451	14.318	1.962	808
2019	47.038	22.570	20.986	2.496	986
Entradas	8.927	5.731	2.953	187	56
Salidas	5.542	3.218	2.063	203	58
Permanencias	32.569	13.621	15.970	2.106	872

Año y suceso	Estrato de asalariados de la EMPRESA				
	Total	Menos de 9	De 10 a 249	De 250 a 999	De 1.000 ó más
2017	56.117	26.615	26.008	2.610	884
Entradas	10.600	6.135	4.110	277	78
Salidas	7.983	4.531	3.141	241	70
Permanencias	37.534	15.949	18.757	2.092	736
2018	62.664	30.345	28.524	2.834	961
Entradas	14.530	8.527	5.493	393	117
Salidas	8.545	4.927	3.275	262	81
Permanencias	39.589	16.891	19.756	2.179	763
2019	68.735	34.689	30.092	2.970	984
Entradas	14.616	9.524	4.691	316	85
Salidas	9.541	5.729	3.449	288	75
Permanencias	44.578	19.436	21.952	2.366	824

Cuando se analizan los sucesos demográficos en combinación con los indicadores de tamaño, la principal consecuencia es que las unidades con mayor empleo tienen mayor probabilidad de supervivencia. Esta es una propiedad ya detectada en experiencias anteriores y plenamente adecuada a lo que sucede en el mundo económico (es extremadamente raro que una gran multinacional muera o por el contrario surja de la *nada*). En efecto, las unidades más relevantes movilizan cantidades importantes de recursos y es menos probable que todos ellos se disuelvan al mismo tiempo.

En complemento a lo anterior, la representatividad de unidades pequeñas es sensiblemente mayor en los flujos poblacionales de entradas/salidas anuales que en la subpoblación de permanencias, con independencia del año y tipo de unidad estadística considerada.

Por último, se señalan algunos enlaces de interés con análisis detallados sobre dinamismo empresarial, desde diferentes ópticas metodológicas:

- https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736161927&menu=ultiDatos&idp=1254735576550
- <https://ec.europa.eu/eurostat/web/structural-business-statistics/business-demography>
- <https://www.oecd.org/sdd/business-stats/structuralanddemographicbusinessstatisticssdbsoecd.htm>

Validaciones finales y Conclusiones

Durante estos cuatro años, el análisis macro ha sido complementado con evaluaciones micro. Estos procesos fueron más intensos en el periodo de arranque, dado que estaba en juego la definición completa de las reglas de continuidad y las cotas de empleo asociadas. La validación de las cotas mediante fundamentos empíricos chequeados a partir de casos reales, fue una prioridad en el desarrollo de este trabajo.

En la práctica, esta evaluación se hizo seleccionando muestras de grupos sujetos a hechos relevantes acaecidos en diferentes periodos de observación (compras, absorciones, reestructuraciones internas o cambios de modelos de negocio). Estos grupos tuvieron cambios significativos en sus perímetros de consolidación, tanto en el número de filiales como en su estructura operativa interna. Para la inmensa mayoría, la consecuencia fue un cambio de identificador estadístico en el DIRCE (algunos ejemplos fueron COMSA, VITALDENT, TECNA, COBEGA, TECNOCOM, GENERAL MOTORS ESPAÑA,...). En sentido opuesto, la estabilidad del identificador fue verificada año a año para los grupos españoles que conformaban el IBEX 35.

La monitorización de los resultados producidos durante este periodo de cuatro años, permite asegurar un comportamiento robusto y estable para el nuevo Sistema de Identificación creado. Ambas propiedades garantizan su viabilidad y sostenibilidad dentro del ciclo de producción DIRCE, al menos a medio plazo.

Con el paso de los años y en la medida que los identificadores se vayan incorporando al resto de cadenas de producción, tanto a nivel nacional como en los diversos proyectos globales que continuamente surgen sobre Multinacionales, es probable que sean necesarios algunos pequeños ajustes o mejoras.