

El uso de datos administrativos para la investigación

El caso de Dinamarca: un ejemplo de
buenas prácticas

Autoras:

Pablo Brugarolas
Lucía Gorjón
Sara de la Rica

Informe
2020/3



Resumen

Este informe ISEAK ilustra el potencial de la integración de datos administrativos para la evaluación de políticas públicas y, por lo tanto, la mejora de la toma de decisiones en el ámbito económico y social. El documento presenta la Base de Datos Integrada para la Investigación del Mercado Laboral (IDA)—una herramienta del sistema danés de acceso a microdatos administrativos—como un caso de buenas prácticas. El documento repasa la estructura de la base de datos y explica el proceso de unificación de IDA con fuentes de datos externas. El buen diseño de su estructura y la facilidad con que es posible armonizar su contenido con el proveniente de fuentes de datos externas hacen de IDA un modelo a imitar para avanzar en la integración de microdatos para el conocimiento empírico y la evaluación.

Agradecimientos

Este informe forma parte de una colaboración con la Diputación Foral de Gipuzkoa. La Fundación ISEAK agradece al Departamento de Promoción Económica, Medio Rural y Equilibrio Territorial de la Diputación Foral de Gipuzkoa su apoyo durante la realización del presente estudio.

**Gipuzkoako
Foru Aldundia**
Ekonomia Sustapeneko,
Landa Ingurune eta
Lurralde Orekako Departamentua



Diputación Foral
de Gipuzkoa
Departamento de Promoción
Económica, Medio Rural
y Equilibrio Territorial

iseak



Esta obra tiene una licencia Creative Commons License. Se permite la reproducción total o parcial, la distribución, la comunicación pública de la obra y la creación de obras derivadas, siempre que no sea con fines comerciales, y siempre que se reconozca la autoría de la obra original.

Índice

1 Introducción	4
2 El sistema danés de microdatos.....	5
2.1 Las reglas de acceso a los microdatos	5
2.2 Datos disponibles	7
3 IDA como ejemplo de buenas prácticas.....	8
3.1 El proceso de cruce de microdatos en IDA	10
3.2 Descripción de las variables background de cada fichero de IDA.....	14
3.3 La unificación de IDA con bases de datos externas	15
4 Conclusiones	18
Referencias	20

1 Introducción

El análisis enfocado a mejorar la toma de decisiones en el ámbito económico y social ha experimentado en los últimos años un creciente interés por el acceso a los microdatos de individuos, hogares y empresas almacenados en ficheros administrativos. Este interés nace, por un lado, de su mayor disponibilidad gracias a las mejoras tecnológicas y los avances estadísticos en el procesado de datos. Al mismo tiempo, en los últimos años ha aumentado la percepción de que un diagnóstico adecuado requiere prestar atención a las decisiones y características de los agentes individuales (Crato y Paruolo, 2019). En esta línea, el diagnóstico sobre la efectividad de diferentes políticas o programas requiere conocer el perfil exhaustivo de los colectivos que participan en ellas, ya sean estos colectivos empresas innovadoras o personas en situación de pobreza y exclusión social. Avanzar en la mejora de las decisiones en el ámbito público requiere por tanto de un mejor aprovechamiento de los datos administrativos de su población.

Para ilustrar el potencial de la integración de los microdatos gracias a la combinación de fuentes administrativas, presentamos en este documento la **Base de datos Integrada para la Investigación del Mercado Laboral (IDA)** como un caso de buenas prácticas. IDA es una base de datos creada por el servicio estadístico de Dinamarca que combina la información de diferentes fuentes administrativas, registrando las características de todas las personas y empresas del país. Cuenta con una estructura muy efectiva que le permite enriquecer su información con la proveniente de otras fuentes administrativas externas, como los ingresos y el patrimonio o las trayectorias educativas de toda la población danesa. Su riqueza la ha convertido en una de las bases de datos más frecuentemente requeridas por los investigadores de las áreas de economía y políticas públicas y un referente del potencial del uso de microdatos para mejorar las decisiones que afectan al bienestar de la sociedad.

La estructura del documento es la siguiente. La sección 2 del documento introduce el sistema danés de microdatos. Se describen en primer lugar las reglas que rigen el acceso a los microdatos en Dinamarca. A continuación se ofrece una panorámica de las distintas fuentes de datos administrativas que el servicio estadístico de Dinamarca pone a disposición del investigador y se sitúa IDA dentro del mismo. La sección 3 se centra exclusivamente en IDA, describiendo en primer lugar el proceso de unificación de datos en la base de datos. Se detalla posteriormente la información disponible en IDA. Al final

de la sección, se explica el proceso de unificación de IDA con fuentes de datos externas y se ofrece una serie de ejemplos en esta línea. La sección 4 recoge las conclusiones de este documento.

2 El sistema danés de microdatos

Para facilitar el análisis de datos utilizando registros administrativos, Statistics Denmark ofrece a los investigadores el acceso a microdatos anonimizados. Debido a su sensibilidad, el acceso a los microdatos requiere de un marco regulatorio que permita a los investigadores acceder a los mismos de forma eficiente a la vez que se garantiza la la confidencialidad de las personas y empresas que componen los registros.

Esta sección presenta en primer lugar las normas y el marco organizativo bajo el cual los investigadores pueden acceder a los microdatos de Statistics Denmark. Al final de la sección se ofrece una panorámica de las distintas fuentes de datos administrativas que Statistics Denmark pone a disposición del investigador.

2.1 Las reglas de acceso a los microdatos

A continuación se describen las reglas que Statistics Denmark establece en el acceso a los microdatos. La tabla al final de esta sección ofrece un resumen de este marco.

Statistics Denmark sólo concede acceso a los **microdatos anonimizados**, es decir, bases de datos en las que se han eliminado las variables que permiten identificar a la unidad de análisis—como el nombre, el número de identificación o la dirección de los individuos. En determinadas ocasiones, los datos también se agrupan o segmentan para garantizar la confidencialidad. Con estas medidas se busca que el análisis riguroso que permite la utilización de microdatos no menoscabe la confidencialidad de las personas que viven en Dinamarca.

El acceso a los microdatos de Statistics Denmark requiere el cumplimiento de una serie de requisitos. En primer lugar, el acceso se concede únicamente a investigadores y a equipos de análisis. Éstos han de demostrar que no se han creado de forma ad-hoc para acceder a los microdatos. De esta forma, Statistics Denmark se asegura que las sanciones en caso de violación de las reglas de acceso se hacen efectivas sobre la

institución pertinente en caso de que así corresponda. La concesión de acceso requiere además la aprobación por parte de Statistics Denmark de la credibilidad del solicitante como receptor de los datos. Para las solicitudes de investigadores y equipos de análisis privados, se tienen en cuenta requisitos adicionales como el tipo de empresa, y la trayectoria investigadora o las cualificaciones de los empleados. Este proceso examinador puede incluir peticiones formales a los clientes donde demuestren el cumplimiento de un requisito concreto.

En relación a los microdatos concedidos, los investigadores y equipos de análisis obtienen acceso a los mismos bajo el **principio *need to know* o necesidad de conocimiento**. Siguiendo este principio, el investigador sólo accede a los datos necesarios para llevar a cabo su propósito investigador. Los solicitantes deben por tanto documentar una relación razonable entre los datos solicitados y la descripción de su proyecto. Asimismo, en caso de que la investigación no requiera datos poblacionales, Statistics Denmark pone a disposición del investigador una muestra representativa. De forma general, las personas autorizadas tienen no obstante acceso a todos las bases de datos existentes—siguiendo siempre el principio de need-to-know.

Tras concederse acceso al solicitante, la persona investigadora firma un acuerdo con Statistics Denmark. En el acuerdo, la persona investigadora accede desde su puesto de trabajo a un **servidor remoto** localizado en las oficinas de Statistics Denmark desde el que se compromete a desarrollar todo su trabajo de análisis. El servidor está separado del resto de la red de Statistics Denmark y únicamente contiene los microdatos administrativos necesarios para el proyecto. También se encuentran instalados a disposición del investigador diferentes programas de análisis de datos, como SAS, SPSS, STATA, GAUSS y R. Estos programas se actualizan de forma frecuente a sus nuevas versiones.

Cualquier intento de identificar a personas o empresas o extraer parte de los microdatos se considera una violación grave del contrato entre el investigador y Statistics Denmark. Sí se permite la **extracción de datos agregados**, que evitan la identificación de personas o empresas. Los resultados agregados se almacenan en un fichero a elección del investigador y se envían a su e-mail cada cinco minutos. Esto permite al investigador tener copias de seguridad de todo el proceso de extracción de resultados. La unidad de

Tabla 1 : Resumen de las reglas de acceso a microdatos de Statistics Denmark

Requisitos de acceso	<ul style="list-style-type: none"> → Solo se concede acceso a investigadores y equipos de análisis. → Los equipos han de demostrar que no se han creado de forma ad-hoc para la extracción de datos. → Requisitos adicionales de experiencia y formación para empresas privadas. → Deben recibir la aprobación de SD de su credibilidad como receptor de datos.
Microdatos concedidos	<ul style="list-style-type: none"> → Los datos se conceden bajo el principio “need to know“ → Muestra en el caso de que el estudio no requiera datos poblacionales
Desarrollo de la investigación	<ul style="list-style-type: none"> → Acceso a los microdatos a través de un servidor remoto → Muestra en el caso
Extracción de datos	<ul style="list-style-type: none"> → No se permite la extracción de microdatos → Se permite la extracción de datos agregados, que envían al correo del investigador

Servicios de Investigación de Statistics Denmark utiliza estos correos para comprobar que la extracción sigue los estándares fijados.

2.2 Datos disponibles

Statistics Denmark cuenta con hasta 250 bases de datos de registro de diferentes agencias de gobierno, ofreciendo datos de gran calidad a nivel poblacional en campos de investigación como el mercado laboral, la sociología, el medioambiente, la salud o la economía de la empresa.

La ventaja del sistema danés es que los **identificadores anonimizados** (identificador personal, dirección, número de identificación de la empresa, número de título de la propiedad) permiten el cruce de los datos de los distintos registros administrativos de la forma necesaria para responder a la investigación.

Para reducir los costes de las bases de datos y ayudar al investigador con el cruce de los datos y otras cuestiones técnicas, Statistics Denmark cuenta ya con las llamadas *bases de datos de investigación*. Las bases de datos de investigación son un producto intermedio proveniente de combinar diferentes fuentes administrativas en un área muy solicitada por los investigadores. Una de estas bases de datos de investigación es la

Base de datos Integrada para la Investigación del Mercado Laboral (IDA). En la siguiente sección exponemos IDA como un ejemplo de buenas prácticas en el cruce de datos administrativos.

3 IDA como ejemplo de buenas prácticas

Esta sección presenta a IDA como un ejemplo de buenas prácticas en la combinación de diferentes fuentes de registros administrativos. IDA, como se ha adelantado, es una de las bases de datos danesas más frecuentemente requeridas por los investigadores de las áreas de economía y políticas públicas. Su desarrollo por parte de Statistics Denmark nació precisamente para facilitar a los investigadores la tarea de identificar correctamente a las empresas a lo largo del tiempo, ahorrando a éstos recursos y tiempo.

La discusión comienza con el proceso de cruce de microdatos en IDA, explicando cómo se armonizan los datos de empresas y personas. Posteriormente se ilustra el proceso de enriquecimiento de la información en IDA con otros registros almacenados por Statistics Denmark.

Antes de comenzar a comentar el proceso de cruce de los datos de registro en IDA, comentamos brevemente algunos elementos claves de su diseño.

La información en IDA se agrupa en tres niveles: el nivel establecimiento, el nivel trabajador y el nivel persona. Combinados, IDA ofrece información poblacional de personas de la población activa¹ y empresas en Dinamarca desde 1980. Esta base de datos se ha actualizado anualmente desde entonces. De esta forma, investigadores de diversas disciplinas tienen acceso a datos longitudinales y universales que conectan individuos con establecimientos y empresas.

Parte del éxito de IDA como ejemplo de buenas prácticas en la combinación de microdatos procede de su nacimiento fruto de la colaboración investigadores-sector público. El equipo de trabajo encargado de la creación de IDA estaba formado por personal de Statistics Denmark e investigadores de diversas universidades de Dinamarca (en concreto, de las universidades de Aalborg y la Copenhagen Business School).

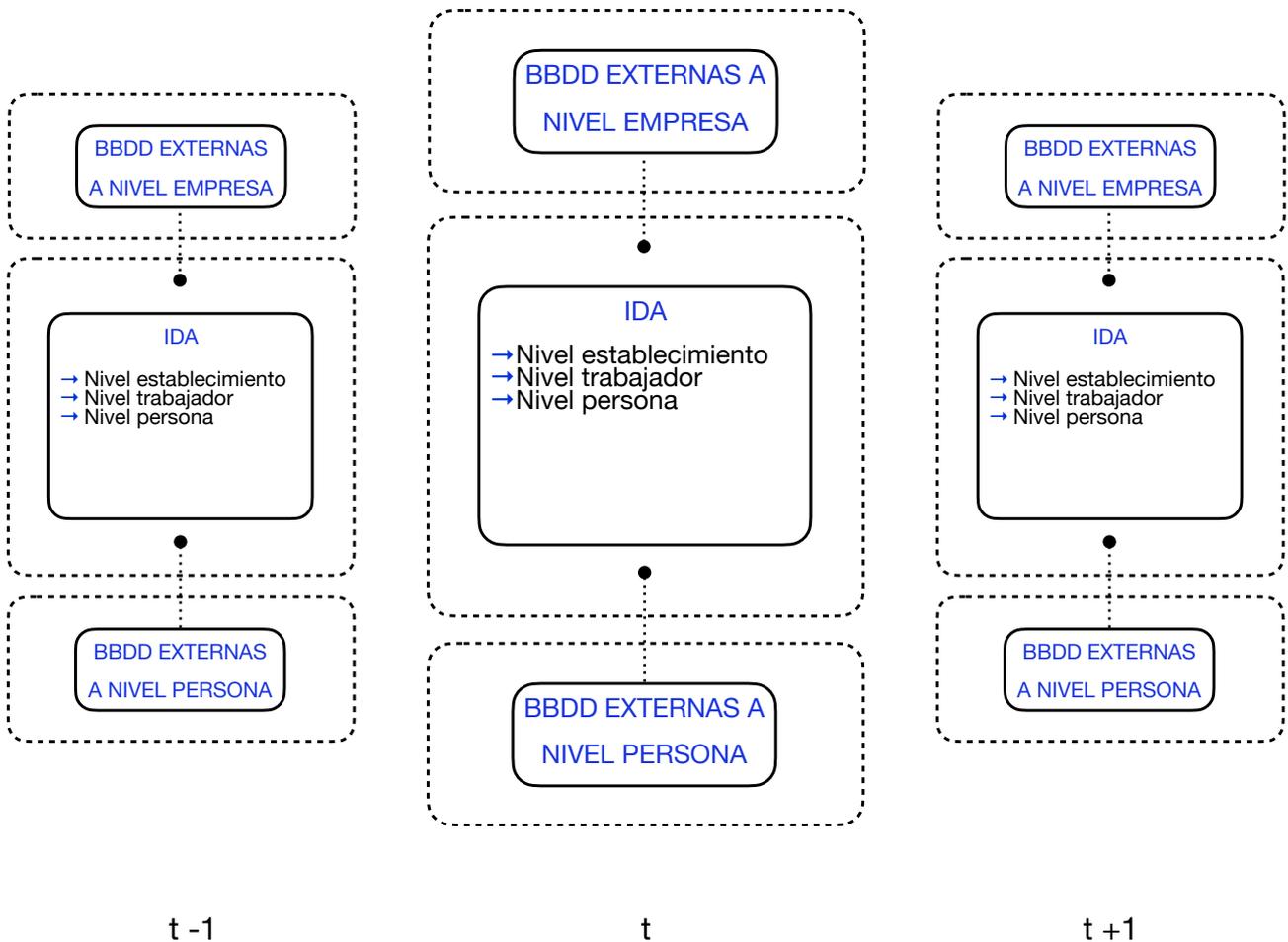
¹ El nivel persona, como se discutirá más adelante, sólo contiene información de la población activa danesa.

Durante un periodo de tres años, entre 1988 y 1990, este equipo creó la primera versión de IDA. La participación de los investigadores en el diseño se reconoce como uno de los elementos clave en su construcción, que permitió, entre otras cosas, la inclusión de registros administrativos previos a la fecha del inicio del proyecto.

Antes de proceder a explicar el proceso de unificación de datos en IDA, presentamos brevemente su estructura, que se muestra en la Figura 1. Como se aprecia en la figura, la información en IDA se integra en dos dimensiones, vertical y horizontal. Verticalmente, IDA combina información de ficheros administrativos de tres niveles para cualquier año de la base: el nivel personal, el nivel trabajador y el nivel establecimiento. Así mismo, es posible enriquecer la información existente en IDA con la disponible en otras bases de datos externas almacenadas por Statistics Denmark. Por otra parte, la dimensión horizontal de IDA presenta su perspectiva longitudinal. Esto permite observar cambios en el mercado de trabajo desde el inicio de la base (1980 en el caso de IDA, cubriendo más de 25 años). En la figura se muestran así la información del año anterior y posterior a un determinado año base.

El resto de la sección detalla el proceso de cruce de datos administrativos en IDA de la información a nivel personal, trabajador y establecimiento. Se pospone la discusión sobre el contenido de cada nivel a la siguiente sección.

Figura 1 : Estructura simplificada de IDA



3.1 El proceso de cruce de microdatos en IDA

Una de las virtudes de IDA es su eficacia en combinar diferentes fuentes de registros administrativos. En esta sección se detalla dicho proceso.

El proceso de unificación de ficheros administrativos comienza partiendo del **nivel establecimiento**. Por establecimiento se entiende a todo local comercial que realiza sus propias actividades económicas y está establecido en una localización geográfica determinada. De este modo, una empresa puede contar con múltiples establecimientos.

La información de los establecimientos se clasifican en dos grupos: identificadores y variables background. Los identificadores permiten determinar de forma única la

información de cada empresa, codificada en las variables background. Posponemos hasta la siguiente sección la discusión de las variables background del registro de establecimientos y nos centramos ahora en los identificadores, pues son las variables que permitirán el proceso de cruce de datos informativos.

En IDA, cada establecimiento cuenta con dos identificadores. El primero de ellos es el *identificador de establecimiento*. El identificador de establecimiento es un número único de 10 dígitos. Los cuatro primeros dígitos hacen referencia al año en que el establecimiento aparece por primera vez en la base de datos, seguido de un número aleatorio de seis dígitos. El segundo identificador es el *identificador empresa*, que sirve para poder determinar qué establecimientos corresponden a la misma empresa.

El **nivel trabajador** es el segundo nivel de información en IDA. El nivel trabajador proporciona información sobre las relaciones laborales de toda la población activa danesa. Contiene información sobre la ocupación principal del trabajador, entendida como aquella que genera su principal fuente de ingresos—aunque es posible determinar la segunda ocupación de cada trabajador. De forma similar a la información del nivel establecimiento, las variables del nivel trabajador agrupan los identificadores y las variables background. De nuevo, posponemos la discusión sobre los detalles sobre las variables background hasta la siguiente sección.

En el registro de los trabajadores de IDA contiene tres identificadores. El primer identificador es el *número de identificación del trabajador*. El número de identificación del trabajador es un número aleatorio conectado al número de la seguridad social de la persona. Este identificador se mantiene fijo a lo largo del tiempo, lo que permite seguir los patrones de la movilidad laboral del trabajador e identificar su historial de empleo. Los dos identificadores restantes son los identificadores de establecimiento y empresa, descritos previamente al comienzo de la sección.

El último de los niveles de información en IDA es el **nivel personal**. El registro de personas está disponibles sólo para la población activa, si bien, como veremos en la sección 5, es posible combinar los microdatos de la población activa con los de la población total. Como ocurre para los registros de establecimientos y trabajadores, las variables del nivel persona se agrupan en identificadores y variables background.

Centramos en esta sección la discusión sobre los detalles de los identificadores personales y el proceso de cruce con el resto de registros

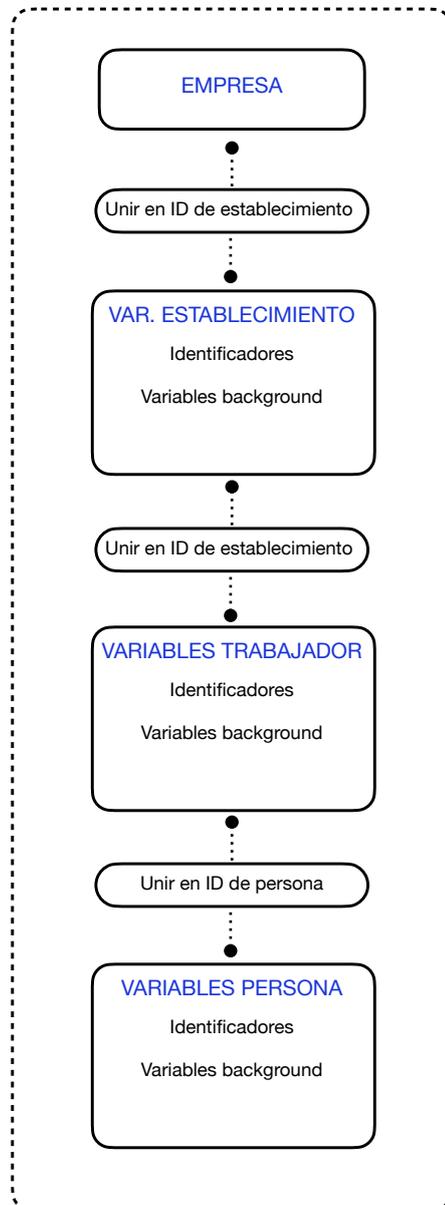
El registro de personas en IDA contiene dos identificadores. El primer identificador es el *número de identificación personal*. Este número permite identificar de forma única a cada persona del registro. El segundo identificador es el número de identificación del trabajador, descrito previamente. Al igual que ocurría con el número de identificación del trabajador, el número de identificación personal se mantiene fijo a lo largo del tiempo.

Utilizando los identificadores de establecimiento y persona es posible cruzar la información del mercado laboral con las características de cada persona activa para cualquier año del registro. Gracias a estos identificadores, es posible saber en qué empresa presta sus servicios cada trabajador. El proceso se ilustra en la Figura 2. Partiendo del identificador del establecimiento, es posible agrupar los establecimientos en empresas. En sentido descendente, el identificador del establecimiento permite también unificar los datos de cada establecimiento con sus trabajadores. Finalmente, la incorporación los datos del fichero de personas se realiza utilizando el identificador de persona.

Por ejemplo, pongamos que el identificador de un determinado establecimiento es el 2000123451. Recordemos, los cuatro primeros dígitos hacen referencia al año de en que el establecimiento aparece por primera vez en la base de datos, y los seis restantes son un número aleatorio. Por simplificar esta ilustración, supongamos que el establecimiento mencionado anteriormente es el único establecimiento de la empresa. Utilizando el identificador del establecimiento, localizamos a los trabajadores que trabajan en el establecimiento 2000123451 en el fichero de trabajadores. A su vez, el número de identificación de cada trabajador nos permitirá completar su perfil con las características recogidas en el fichero persona.

De esta forma, IDA permite combinar la información background de los niveles establecimiento, trabajador y persona para obtener una visión completa del mercado de trabajo que permita realizar diagnósticos y evaluaciones más precisas. En la siguiente sección se describe en mayor profundidad la información background recogida en cada fichero.

Figura 2 : Unificación de los microdatos de IDA



3.2 Descripción de las variables background de cada fichero de IDA

Una vez detallado el proceso de unificación de datos en IDA, se procede a explicar la información que cada uno de los niveles de la base de datos contiene. Se muestra así toda la información que IDA pone a disposición de los investigadores de forma accesible y limpia. La información background existente de cada uno de los niveles de IDA también se puede completar con información proveniente de otros registros de Statistics Denmark, como se detalla en la siguiente sección.

El primer nivel de IDA está formado como hemos visto por el **registro de establecimientos**. Las variables de background de este registro proporcionan información detallada sobre las características de cada establecimiento. En concreto, el registro recoge información sobre el sector del establecimiento, su localización geográfica, el tipo de empresa, el número de establecimientos afiliados (utilizando el identificador de empresa), si pertenece al sector público o privado y el año en el que se creó el establecimiento. También se recogen en el fichero una serie de variables sobre el número de empleados del establecimiento. La primera de ellas hace referencia al número de empleados activos del establecimiento en un año determinado. También se encuentra disponible el número de personas que no tienen esta ocupación como su fuente primaria de ingresos. El resto de variables sobre el número de empleados proporcionan información sobre el número de empleados activos en el establecimiento durante el año anterior. La primera de estas variables identifica el número de empleados que trabajaron a tiempo completo el año anterior, la última recoge el registro de todos los empleados que trabajaron en ese establecimiento durante ese año.

Las variables de background del **registro de trabajadores** proporcionan información detallada sobre las relaciones empresario-trabajador. La primera de estas variables es el año en que el trabajador comienza su relación laboral en ese establecimiento. La segunda variable distingue si la persona es un trabajador o empresario. Otras dos variables describen si el trabajador trabaja a tiempo completo o a tiempo parcial y el número de horas de trabajo que realiza cada semana. **También se registra el salario hora del trabajador**. La última de estas variables indica la distancia en kilómetros entre la residencia del trabajador y el establecimiento.

En su **registro de personas** que forman parte de la población activa, IDA cuenta con numerosas variables background, que pueden agruparse en las varias categorías. En primer lugar se encuentran la información demográfica, que incluye el género, la edad, la residencia, la ciudadanía y el país de origen. Se identifica también si la persona es danés o inmigrantes de primera o segunda generación. Por otro lado, en los microdatos de personas también se registra la trayectoria laboral (el total de años de experiencia y los años en desempleo). Las variables del registro de personas de IDA también incluyen información sobre el perfil educativo. La más utilizadas son el nivel de educación alcanzado y la educación de cada persona en un punto concreto del tiempo. También existen a disposición del investigador microdatos sobre el nivel educativo (por ejemplo, educación secundaria, grado universitario, máster, etc), la rama educativa (ciencias sociales, humanidades, ingeniería, etc) o el área concreta dentro de estas ramas (economía, sociología, psicología, etc). IDA proporciona otra información educativa como el año de finalización de los estudios y el municipio en que la persona está completando sus estudios. Finalmente, el fichero de personas recoge varias variables de ingresos (los ingresos disponibles, ingresos brutos y salario neto).

3.3 La unificación de IDA con bases de datos externas

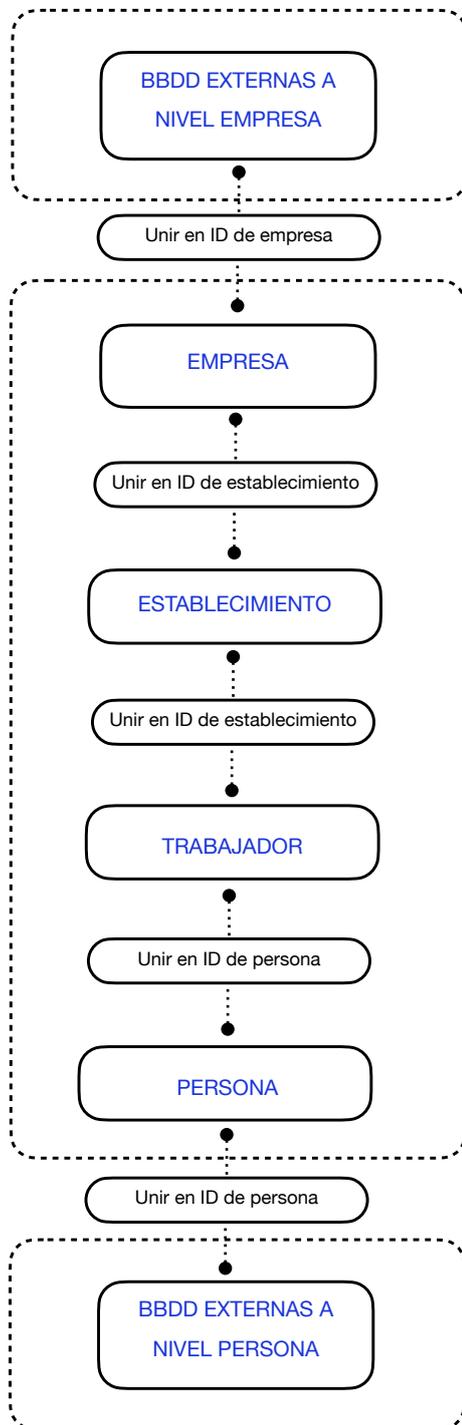
Otra fortaleza de IDA, más allá de su naturaleza de registro y su diseño longitudinal, es la posibilidad de unificarla con las gran variedad de registros administrativos ya almacenadas en Statistics Denmark, encuestas realizadas en colaboración con Statistics Denmark o bases de datos de otras organizaciones.

Tres ejemplos notables de este proceso son la unificación de IDA con los registros médicos (Dahl, 2009), la unificación de IDA con una encuesta de emprendedores daneses y asalariados para identificar las características de los emprendedores en Dinamarca (Dahl et al., 2009) o la combinación de los datos de IDA con los registros de ingresos, patrimonio y demográficos para estudiar el impacto de tener un hijo en los ingresos y la participación laboral de las mujeres (Kleven et al., 2019).

Esta unificación se realiza utilizando el número de identificación pertinente para incorporar la base de datos externa. Como se muestra en la Figura 3, el enriquecimiento de IDA con bases de datos externas a nivel persona utiliza el identificador de persona.

Para las bases de datos externas a nivel empresa, se utiliza el identificador de empresas. Detallamos a continuación ejemplos notables de unificación con fuentes externas.

Figura 3 : Unificación de los microdatos de IDA con bases de datos externas



Como se ha comentado previamente, Statistics Denmark cuenta con hasta 250 bases de datos de registro en campos de investigación como el mercado laboral, la sociología, el medioambiente, la salud o la economía de la empresa. A continuación se presentan ejemplos de unificación de bases de datos externas con IDA. Este proceso permite aumentar en gran medida la riqueza de los microdatos ya existentes. Se explican en primer lugar el contenido de los registros de población, ingresos y estudiantes como ejemplos notables de enriquecimiento de los microdatos vía número de identificación personal. Finalmente, comentamos el contenido la encuesta de formación continua a los trabajadores con un ejemplo de unificación de bases externas de empresas utilizando el identificador de establecimiento.

La primera base de datos externa que permite completar la información ya disponible en IDA es el **registro de población**. El registro de población permite estudiar el tamaño y la composición de toda la población en Dinamarca en base a sus características demográficas (género, edad, estado civil, municipalidad...). Permite por tanto extender la información demográfica ya recogida en IDA para la población activa al conjunto de la población en Dinamarca.

Statistics Denmark también almacena el **registro de ingresos** de Dinamarca. Dicho registro contiene las estadísticas de ingresos y pago de impuestos de toda la población de Dinamarca. El registro cuenta con los microdatos anuales desde 1980 a nivel persona y familia. En concreto, contiene información sobre los ingresos primarios, las transferencias, las propiedades y los impuestos de la población. Existe así información a nivel registro a disposición de los investigadores que contiene la información poblacional de los salarios, las ayudas por desempleo, las becas al estudio y otras ayudas finalistas, las pensiones sociales y otras ayudas sociales. Además de recoger el nivel de ingresos, el fichero administrativo de ingresos pone también el foco en su distribución. Se incluyen así indicadores como el decil correspondiente de ingresos, el riesgo de pobreza o el coeficiente de Gini, medidos en términos de ingreso equivalente disponible.

Los microdatos de IDA se pueden también enriquecer con su unificación con el **registro de estudiantes**. Creado durante el inicio de los años setenta, el registro de estudiantes es un registro longitudinal que permite seguir el desarrollo educativo de los estudiantes en el sistema de educación formal. Permite también detectar la inscripción y la graduación de los estudiantes en los diferentes niveles educativos, así como sus

cualificaciones, la institución donde desarrollan su formación o los cambios en sus trayectorias. Se actualiza anualmente y utiliza los microdatos que las instituciones educativas ponen al servicio de Statistics Denmark. Recoge información desde la etapa preescolar hasta el doctorado (desde la educación secundaria para los ejercicios anteriores a 2006).

Concluimos esta sección con un ejemplo notable de unificación de bases externas de empresas utilizando el identificador de establecimiento, **la encuesta de formación continua a los trabajadores (CVTS)**. Dicha encuesta permite estudiar qué empresas apuestan por la formación continua y cómo afecta a sus diferentes resultados—su competitividad, por ejemplo. La encuesta recoge tanto los empleados de cada establecimiento que participan en actividades de formación y como el gasto de cada empresa en formación continua a sus trabajadores. Se detallan a su vez el contenido de los programas y su intensidad, prestando especial atención a los programas enfocados a mejorar las competencias digitales de los trabajadores.

4 Conclusiones

La unificación de los microdatos provenientes de diferentes registros administrativos emerge como una gran oportunidad para avanzar en la toma de las decisiones desde la óptica de las políticas basadas en la evidencia. De la riqueza de estos registros depende que las evaluaciones sobre la efectividad de los programas y políticas existentes sean lo más rigurosas posibles.

Para ilustrar el potencial de la integración de los microdatos, presentamos en este documento la Base de Datos Integrada para la Investigación del Mercado Laboral (IDA) como un caso de buenas prácticas. Como se ha expuesto, IDA se ofrece como una herramienta muy completa para analizar el mercado laboral. A nivel establecimiento, se recoge información sobre el año de creación del establecimiento, el número de empleados, el sector, su localización geográfica, el tipo de empresa y el número de establecimientos afiliados. A nivel trabajador, IDA registra el año en que el trabajador comienza su relación laboral en ese establecimiento, el número de horas de trabajo semanales, su salario hora e, incluso, la distancia entre la residencia del trabajador y el establecimiento. En su registro de personas, IDA documenta el género, la edad, la

residencia, la ciudadanía, el país de origen, el nivel de educativo y los ingresos de la persona.

Finalmente, en este documento también se ha detallado como IDA permite la integración con la información administrativa de los 250 registros administrativos externos que Statistics Denmark almacena. Como ejemplos notables, se ha explicado la riqueza del registro de población—que extiende la información demográfica ya recogida en IDA al conjunto de la población en Dinamarca—, el registro de ingresos—que contiene información sobre los ingresos primarios, las transferencias, las propiedades y los impuestos de la población—, el registro de estudiantes—que permite extender la información educativa ya recogida en IDA al conjunto de la población en Dinamarca incluyendo el historial académico y la información concreta de cada ciudadano desde la etapa preescolar hasta el doctorado—, la encuesta de formación continua a los trabajadores (CVTS)—que recoge, entre otras cuestiones, el contenido de programas de formación enfocados a mejorar las competencias digitales de los trabajadores - y, finalmente, registros relacionados con la salud y el bienestar de las personas.

Referencias

Crato, N. and Paruolo, P., 2019. Data-driven policy impact evaluation: How Access to microdata is transforming policy design (p. 346). Springer Nature.

Dahl, M.S., 2009. The Cancer of Organizational Change: Organizational Change and Employee Stress. In *2009 Annual Meeting of the Academy of Management*.

Dahl, M.S., Jensen, P.G. and Nielsen, K., 2009. Jagten på fremtidens nye vækstvirksomheder: Hovedrapport. DJØF forlag.

Kleven, H., Landais, C. and Søgaard, J.E., 2019. Children and gender inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, 11(4), pp.181-209.

