

LANZAMIENTO DEL LABORATORIO DE DATOS DEL BANCO DE ESPAÑA (BELab) EN LA CENTRAL DE BALANCES

El Banco de España puso en marcha, el pasado 22 de julio de 2019, un proyecto piloto de laboratorio de datos (BELab) para facilitar el acceso de la comunidad investigadora a microdatos de alta calidad recopilados por la institución. Al igual que otros laboratorios de entidades del Eurosistema, BELab está adherido a la red INEXDA¹ y en él se aplican los protocolos de acceso seguro a datos sensibles.

La primera fase de BELab ha empezado con el acceso a los microdatos de las empresas no financieras recogidos en la Central de Balances. La difusión de estos datos se ampara en el convenio firmado el pasado 11 de julio entre el Banco de España y el Colegio de Registradores de la Propiedad y Mercantiles de España, que autoriza el uso controlado de esa información en un entorno seguro (mediante «laboratorio *in situ*» o *data room*) en las instalaciones del Banco de España, en la calle Alcalá, 522. Traducido en números, significa que los investigadores tienen la posibilidad de utilizar más de 130 indicadores, para cada empresa y año, derivados de las cuentas anuales y otra información relevante (empleo, rama de actividad productiva, tamaño, etc.), referidos a un período de 23 años (1995-2018), de más de dos millones de empresas, alcanzando más de 16 millones de cuentas anuales. Para un conocimiento más detallado de la información disponible, se puede acceder a la [Guía de usuario de los microdatos](#) de la Central de Balances (MCB). Además, con el fin de facilitar el diseño del código y reducir el tiempo de estancia en el laboratorio *in situ*, se pone a disposición del investigador un fichero de muestra

de MCB con datos ficticios de todas las variables disponibles.

En esta fase piloto solo se puede acceder a los datos a través del modelo de laboratorio más seguro de los posibles, el denominado «laboratorio *in situ*», lo que conlleva el desplazamiento físico del investigador a una sala de los locales del Banco de España (en su sede de Madrid, en la calle Alcalá, 522) con ordenadores independientes y aislados de la red del Banco y con medidas de seguridad para evitar la difusión externa de datos confidenciales. Existen otros procedimientos utilizados en los laboratorios de datos europeos, principalmente dos, que se aplican según el nivel de confidencialidad de la información: a) la ejecución subrogada, en la que el investigador facilita al personal del laboratorio el programa que desea utilizar, para su ejecución por dicho personal, y b) el acceso remoto, en el que el investigador utiliza un *software* específico desde su puesto en la universidad, que le permite ejecutar por sí mismo los programas, pero no puede descargarse los resultados sin el concurso del personal del laboratorio (véanse cuadros 1 y 2). La vocación del BELab del Banco de España es ampliar las formas de acceso a sus bases de datos siguiendo el ejemplo de los miembros de INEXDA, antes referida.

El cuadro 2 informa de las ventajas y los inconvenientes de los distintos modelos de acceso.

La guía de funcionamiento de BELab está publicada en su sitio web, en el siguiente enlace: [«Guía de funcionamiento](#)

Cuadro 1
CARACTERÍSTICAS DE LOS DIFERENTES MODELOS DE ACCESO EN LOS LABORATORIOS DE DATOS

Nivel de confidencialidad	Tipo de acceso	Tipo de visualización de los datos	Ejecución de los programas sobre datos reales
Bajo	<i>Data room</i> , ejecución subrogada y acceso remoto	Datos reales anonimizados	Por el investigador
Medio	<i>Data room</i> , ejecución subrogada y acceso remoto	Datos perturbados anonimizados	Por personal del laboratorio de datos
Alto	<i>Data room</i> y ejecución subrogada	Datos fuertemente perturbados (muestra) y anonimizados	Por personal del laboratorio de datos

FUENTE: Banco de España, a partir de la información suministrada por BPLim (laboratorio de datos del Banco de Portugal).

1 INEXDA es el acrónimo de International Network for Exchanging Experiences on Statistical Handling of Granular Data.

LANZAMIENTO DEL LABORATORIO DE DATOS DEL BANCO DE ESPAÑA (BELab) EN LA CENTRAL DE BALANCES (cont.)

de BELab», y proporciona información a los investigadores sobre los datos disponibles, las condiciones de acceso y de uso, el *software* habilitado en los puestos de BELab (Stata, R y Python, entre otros), el protocolo para la extracción del *output* y la política de almacenamiento de los datos utilizados en un proyecto.

Los investigadores interesados en acceder a BELab deberán enviar un *curriculum vitae* con la relación de sus actividades de investigación realizadas y el [formulario de solicitud](#). Ambos documentos deben ser enviados a través del [formulario de contacto](#). Una vez recibida la solicitud, el Comité de Evaluación Técnica de Investigación de BELab determinará la aptitud del solicitante para acceder a los microdatos y comunicará su decisión por correo

electrónico a la dirección remitente. Los investigadores cuyo proyecto sea aprobado podrán acceder al laboratorio *in situ* de BELab durante un período que se determinará de común acuerdo.

Una vez se evalúe la experiencia piloto, se analizarán la viabilidad y el interés para los investigadores de acceder mediante otros procedimientos y a otras bases de datos existentes en la institución, según el desarrollo tecnológico y los requerimientos de confidencialidad lo vayan permitiendo. BELab informará de estos progresos en su sitio web.

Los lectores interesados en conocer más detalles del laboratorio de datos del Banco de España pueden acceder al sitio web de [BELab](#).

Cuadro 2

VENTAJAS E INCONVENIENTES DE LOS MODELOS DE ACCESO EN LOS LABORATORIOS DE DATOS

Tipo de acceso	Ventajas	Inconvenientes
Laboratorio <i>in situ</i>	Garantía de que solo la persona autorizada es quien usa los datos	Necesidad de habilitar espacio dentro de la institución y puestos informáticos Dificultad de acceso para investigadores que no residan en la ciudad donde está ubicado el laboratorio
Ejecución subrogada	No es necesario el desplazamiento del investigador	Necesidad de crear bases de datos <i>dummies</i> Necesidad de revisión de códigos de programa ejecutados sobre datos <i>dummies</i> Mayor número de iteraciones, debido a que el diseño del código no se hace sobre datos reales
En remoto	No es necesario el desplazamiento del investigador El investigador actúa directamente sobre datos reales	Mayores requerimientos legales Imposibilidad de difundir las bases de datos más confidenciales

FUENTE: Banco de España.