

BANCO DE INNOVACIÓN EN LAS ADMINISTRACIONES PÚBLICAS

TÍTULO

Universidata, el primer portal de datos abiertos sobre educación superior en España

INFORMACIÓN INICIAL:

PROBLEMA:

Los datos abiertos son esenciales para el desarrollo de la economía digital y la ciencia de datos (campo interdisciplinario que combina estadísticas, programación y conocimientos de dominio para analizar y extraer información de grandes conjuntos de datos). Así pues, son utilizados por investigadores y empresas para la creación de modelos de ciencia de datos y aplicaciones digitales que permiten mejorar la eficiencia y la productividad, así como diseñar modelos predictivos e identificar patrones y tendencias.

Las empresas infomediarias, dedicadas a la recolección, procesamiento, análisis y distribución de información, actúan como intermediarios entre los productores de información y los usuarios finales, proporcionando servicios, desde la recopilación y análisis de datos financieros hasta la gestión de contenidos multimedia y la provisión de servicios de inteligencia de mercado.

Los datos abiertos y la ciencia de datos son fundamentales para la economía digital porque fomentan la innovación, ya que al poner a disposición datos abiertos, se crea un ambiente propicio para el desarrollo de nuevas ideas y soluciones que pueden impulsar la innovación en la economía digital. La ciencia de datos, por su parte, ayuda a analizar y comprender esos datos, lo que puede dar lugar a nuevas oportunidades y modelos de negocio.

La Asociación Multisectorial de Información ([ASEDIE](#)) señala en su 10º informe anual “[Economía del dato en el ámbito infomediario](#)” relativo a 2022 las siguientes conclusiones: identifica un total de 703 empresas, lo cual supone un crecimiento del 58% respecto a 2013; con cerca de 23.000 profesionales, frente a los 10.000 de hace una década; y, finalmente, un volumen de facturación de más de 2.000 millones de euros, cifra que representa un 103% en comparación con 2013.

Asistimos a un movimiento global consolidado de apertura de datos que arrancó hace más de quince años cuando un grupo de defensores del gobierno abierto se reunió para desarrollar una serie de [principios para](#)

[la apertura de los datos gubernamentales](#). El movimiento de datos abiertos en las Administraciones públicas, una herramienta de transparencia, al que más tarde se unirían todo tipo de organizaciones públicas y privadas, ha revalorizado el *big data* tanto a nivel económico como social.

En el contexto europeo, España se encuentra en una [posición avanzada](#), tal como revelan los [informes de madurez de datos abiertos de la UE](#) (*Open Data Maturity Report*), publicados desde 2015 en el [Portal Europeo de Datos Abiertos](#) de la Comisión Europea. El estudio evalúa el desarrollo de los países europeos en el ámbito de los datos abiertos y ofrece una visión general de las buenas prácticas aplicadas en Europa. El [último informe publicado](#), de diciembre de 2022, evidencia que España está muy adelantada, manteniéndose en el grupo de los *trend-setters* (creadores de tendencias o grupo de prescriptores), con una puntuación del 92%, trece puntos por encima de la media de la EU27. Francia ocupa el primer puesto.

En España la plataforma gubernamental datos.gob.es es la comunidad para administraciones, empresas y ciudadanos interesados en la apertura de la información pública y el desarrollo de servicios avanzados. Esta plataforma, cuyo cometido es promocionar la cultura de los datos abiertos y desarrollar el sector infomediario, pone a disposición sus datos para que puedan ser reutilizados, haciéndolos accesibles, de acuerdo con el [marco legislativo de España](#) (descrito en el apartado “Descripción de la política o programa” de esta buena práctica).

Otras plataformas gubernamentales destacadas son data.gov.uk del Reino Unido; open.canada.ca de Canadá, así como la Oficina del Censo de Estados Unidos ([United States Census Bureau](#)) y el Centro Nacional de Estadísticas de Educación de EE.UU ([The National Center for Education Statistics](#)).

Pese a los avances experimentados en el sector y el movimiento global de apertura de datos abiertos, todavía se observan resistencias a la transparencia para poder acceder a la información, como evidenciaba ASEDIE en su informe de la [edición 2021](#), así como insuficiente implantación de la digitalización y una deficiente calidad de los datos.

SOLUCIÓN GLOBAL:

La ciencia de datos en el sector educativo universitario tiene el potencial de mejorar significativamente la calidad educativa y la eficiencia operativa de las universidades. Puede utilizarse para analizar el rendimiento académico de los estudiantes u optimizar la programación académica, pues las universidades

pueden identificar patrones en la selección de cursos por parte de los estudiantes y ajustar la oferta académica para satisfacer la demanda, etc.

La plataforma nacional (datos.gob.es) sobre el sector de educación incluye cerca de 1000 datasets sobre alumnos matriculados según su Comunidad Autónoma, modalidad de universidad, ámbito de estudio, sexo, nacionalidad, etc., publicados por el Ministerio de Universidades.

Por su parte, en 2020 surge [Universidata](#), una iniciativa impulsada por seis universidades para fomentar los datos abiertos en el sector de la educación superior en España, facilitando su uso, explotación y puesta en valor por parte tanto de los publicadores (las universidades) como de los consumidores (el sector infomediario y la sociedad en general).

Se trata del primer portal de datos abiertos en el ámbito universitario en España. Es un portal centralizado con un acceso único, donde cada universidad puede ofrecer los datos que elabora y gestiona. Esto facilita y potencia la capacidad de análisis sobre los datos publicados y su reutilización. Asimismo, permite trabajar de manera armonizada y conjunta para aumentar el impacto de la publicación de los datos.

Esta solución surge para facilitar al máximo el trabajo y minimizar el coste a las universidades, descargándolas de gran parte de las tareas técnicas y administrativas asociadas al lanzamiento y mantenimiento de una iniciativa de datos abiertos. Asimismo, aumenta la confianza en las instituciones educativas, mejora la calidad de la educación –al permitir identificar puntos fuertes y débiles– y permite la promoción de la innovación mediante la creación de nuevas herramientas y servicios que mejoren la educación. También facilita la toma de decisiones y la colaboración entre empresas, organizaciones y ciudadanos, lo que puede generar sinergias y nuevas oportunidades de negocio.

Universidata ofrece 70 conjuntos de datos que incluyen 754 ficheros, los cuales contienen, en total, varios millones de datos abiertos por parte de las seis universidades participantes hasta el momento (Universidad Rey Juan Carlos, Universidad Complutense de Madrid, Universidad Autónoma de Madrid, Universidad Carlos III de Madrid, Universidad de Valladolid, Universidad de Huelva).

Foto 1. Portal web de Universidata



Fuente: UniversiDATA

COSTE APROXIMADO:

El coste del lanzamiento fue asumido íntegramente por la propia empresa Dimetrical a partir de 2019. Una vez desarrollado el proyecto, los usuarios de la plataforma Universidata aportan 12.000 € al año por el coste de la suscripción, importe que incluye todos los servicios asociados, descritos en el apartado “Descripción de la política o programa” de esta buena práctica.

TERRITORIO:

Nivel nacional. Ámbito universitario.

PÚBLICO DESTINATARIO:

- Para los **reutilizadores** facilita el tratamiento de los datos al ofrecer los contenidos de manera homogénea y perfectamente documentada disponibles en formatos estándar y fácilmente usables mediante un API gratuito y de libre acceso.
- Para la **sociedad en general** ofrece acceso directo a datos verificados sobre las principales áreas de actividad de las universidades y proporciona canales de participación para contribuir a que la

voz de la comunidad universitaria se haga oír y contribuya decisivamente en la mejora de la calidad y alcance de este proyecto.

- Para las **universidades** permite optimizar la inversión realizada por la comunidad universitaria en análisis, codificación y estandarización y, al mismo tiempo, permite publicar los sets de datos en estándares internacionales en materiales tan sensibles como presupuestos públicos o contratación del sector público.

ENTIDAD QUE LA HA LLEVADO A CABO:

Este proyecto se ha llevado a cabo, por un lado, por la empresa DIMETRICAL, empresa tecnológica especializada en investigación y desarrollo de proyectos, productos y servicios de procesamiento y análisis avanzado de datos para instituciones de educación superior, y las Universidades Rey Juan Carlos, Complutense de Madrid y Autónoma de Madrid. Desde su nacimiento, otras tres Universidades públicas se han incorporado al proyecto –la Universidad Carlos III de Madrid, la Universidad de Valladolid y la Universidad de Huelva– y otras están actualmente en proceso de incorporación.

DESCRIPCIÓN DE LA POLÍTICA O PROGRAMA:

Se trata de un proyecto colaborativo público-privado, a través del cual las universidades participantes abren conjuntos de datos –los conocidos como *datasets*– a toda la sociedad, de manera que los mismos se pueden compartir, ser reutilizados por toda la comunidad universitaria, cualquier ciudadano, así como por todo el sector empresarial infomediario.

El componente más visible de Universidata es su **punto de acceso único**: un portal centralizado y compartido por todas las universidades participantes, mediante el cual se ponen a disposición de la sociedad los diversos conjuntos de datos abiertos, que conforman el denominado “**núcleo común**”: una colección estándar de conjuntos de datos que todas las universidades participantes han acordado abrir de forma coordinada y armonizada, esto es, usando los mismos formatos, semántica (definiciones), granularidad (nivel de detalle), calendario de publicación, etc.

Existen otros componentes internos no visibles para los usuarios, pero no por ello menos importantes. Entre estos destacan los procesos de generación automática de conjuntos de datos partiendo de ficheros internos de las universidades, los procesos de enriquecimiento de datos (por ejemplo, geolocalización de los centros de estudios o los municipios referenciados en los datos) o los procesos de anonimización.

Las principales **ventajas** para las universidades que participan en el proyecto son:

- Facilita el cumplimiento de los requerimientos planteados tanto por la [Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno](#), como la [Ley 10/2019, de 10 de abril, de Transparencia y de Participación de la Comunidad de Madrid](#) (para el caso de dicha Comunidad Autónoma) y otras normativas autonómicas similares relativas a la publicación de sus datos en formatos reutilizables, cumpliendo así con las recientes directrices fruto de la modificación de la [Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público](#) en aplicación de la [Directiva \(UE\) 2019/1024 del Parlamento Europeo y del Consejo, de 20 de junio de 2019, relativa a los datos abiertos y la reutilización de la información del sector público](#).
- Provee una solución integral, compartida y estandarizada para la gestión, tratamiento, enriquecimiento, anonimización automatizada y publicación de los conjuntos de datos.
- Es 100% software libre: garantiza la completa disponibilidad de la solución por parte de las universidades al publicarse los datos en la plataforma de código abierto DKAN.
- La definición de contenidos de manera homogénea para todas las universidades participantes en el proyecto aporta un valor añadido al abrir la posibilidad de realizar estudios comparados entre los participantes.
- Permite, de manera automatizada y sin sobreesfuerzos por parte de las universidades, la publicación los conjuntos de datos bajo estándares internacionales como Open Fiscal Data Package para los presupuestos públicos.
- La federación automática con el portal nacional de datos datos.gob.es aumenta la visibilidad y el impacto de los datos publicados.
- Las aplicaciones analíticas que se irán publicando en el portal permiten el análisis interactivo e inmediato de los datos desde el mismo momento de su publicación.

En cuanto a los servicios incluidos en la suscripción de las universidades, Universidata aporta prácticamente todos los servicios necesarios para el mantenimiento de la plataforma con una intervención mínima por parte del personal de las Universidades:

- Apoyo a la puesta en marcha: aspectos legales, administrativos y técnicos de una iniciativa de apertura de datos.
- Gestión del alojamiento, mantenimiento y evolución del portal.
- Gestión de contenidos asociados: noticias, destacados, etc.
- Generación y gestión de la [documentación](#) asociada a los *datasets* publicados.
- Publicación inicial y actualización periódica de los datasets a partir de ficheros de datos internos:
 - Generación automática en los formatos estándar a partir de los ficheros fuente.
 - Validación y control de calidad de los datos.
 - Anonimización de los datos y generación de informe de gestión de riesgos asociado.
 - Enriquecimiento de los datos: etiquetado de códigos y geolocalización de ubicaciones (países, centros de estudios, localidades...).
- Generación y mantenimiento de un *endpoint* (punto de acceso final) de federación de datos con el portal nacional datos.gob.es.
- Publicación de actualizaciones vía [RRSS](#).
- Elaboración y publicación de análisis y aplicativos sobre los datos publicados.
- Gestión de la interacción con los usuarios (comentarios, peticiones, etc.).
- Eventos de promoción de uso de los datos: [encuentros](#) con usuarios, [Datathon](#), etc.

Por otro lado, el portal Universidata ha sido también diseñado pensando en facilitar el trabajo a los **reutilizadores**. Estas son algunas de las facilidades:

- Contenidos homogéneos y perfectamente documentados: la especificación abierta "[núcleo común](#)" informa hasta el detalle del contenido, granularidad, estructura, semántica, frecuencia de actualización y demás metadatos de cada conjunto de datos publicado.
- Se busca siempre ofrecer los contenidos en los estándares aceptados y los formatos más prácticos posible: dcat, csv, xlsx, Open Fiscal Data Package, etc.
- No sólo se publica la información de forma armonizada y en formatos estándares, sino que se hace siempre al máximo nivel de detalle, atendiendo así una de las principales demandas del sector infomediario.

- Un [API](#) (*Application Programming Interface* o interfaz de programación de aplicaciones), gratuito y de libre acceso, sin necesidad de registro.
- Eventos para promocionar la reutilización de los datos (*datathones*), encuentros con reutilizadores o, etc.
- Ejemplos didácticos que muestran las posibilidades de análisis de los datos publicados, actualmente en la sección “laboratorio” y muy próximamente en forma de aplicaciones analíticas interactivas y de código abierto (actualmente en desarrollo).
- Al ponerse de acuerdo todas las instituciones en el uso de las mismas definiciones de datos, permite al ciudadano en particular y a la sociedad en general, comparar distintos aspectos o hacer estudios en conjunto añadiendo un extra al concepto de rendición de cuentas financieras y no financieras y al carácter ejemplarizante que deben tener las Administraciones públicas.

En cuanto a la evolución, el rumbo del proyecto, los nuevos conjuntos de datos a publicar, los estudios y aplicaciones analíticas que se van desarrollando, etc., son decididos de manera conjunta por todas las instituciones participantes.

Como evolución natural de la sección "laboratorio" de Universidata surgió [Universidata-Lab](#), que pasará de ser una sección estática en el portal actual a constituir un nuevo portal "hermano" y complementario. Es una herramienta de visualización de datos abiertos de la plataforma Universidata centrado en el análisis avanzado de los datasets publicados en Universidata. Precisamente fue esta la iniciativa premiada en el III Desafío Aporta, certamen que ha premiado iniciativas de valorización del dato en la educación digital, impulsado por la Iniciativa Aporta, junto con la Secretaría de Estado de Digitalización e Inteligencia Artificial y Red.es (organismo público adscrito al Ministerio de Asuntos Económicos y Transformación Digital, que impulsa la Agenda Digital en España desarrollando iniciativas y proyectos de digitalización y desarrollo tecnológico en el ámbito de la economía, los servicios públicos, la ciudadanía, las infraestructuras y la internacionalización de empresas).

OBSTÁCULOS SUPERADOS:

Los principales obstáculos enfrentados han sido fundamentalmente dos:

- La complejidad de coordinar y alinear a todos los participantes, elemento crucial en una iniciativa colaborativa como esta.

- El cumplimiento de todas las normativas vigentes aplicables al sector universitario –tanto legales como técnicas– en toda su extensión, especialmente en lo relativo a la protección de datos personales.

Foto 2. Entrega del premio Aporta con varios miembros fundadores



Fuente: UniversiDATA.

IMPACTO:

El proyecto Universidata ha recibido hasta el momento los siguientes reconocimientos:

- I Premio “[III Desafío Aporta](#)” de Red.es. El premio dotado de 4.000 euros ha servido para financiar el I Datathon Universidata (certamen de ideas de estudiantes relacionadas con la ciencia y el tratamiento de datos), concedido en diciembre de 2022. En el evento se plantea a profesores y universitarios que propongan soluciones innovadoras, reutilizando los datos de la plataforma, con el objetivo de generar comunidad alrededor de la iniciativa.
- Selección como uno de los 30 casos de estudio europeos que forman parte del proyecto “[Seguimiento durante 3 años de 30 casos de reutilización sobre el impacto económico, gubernamental, social y ambiental](#)” por parte del observatorio del [Portal Europeo de Datos Abiertos](#).
- Impacto sectorial: incremento sin precedentes, y en muy poco tiempo, del número de universidades que disponen de un portal de datos abiertos federado con el portal nacional de datos (datos.gob.es).
- Mejora en los rankings de transparencia de las universidades participantes.
- Generación de riqueza mediante la reutilización efectiva de los datos por parte de la sociedad en general y del sector infomediario en particular, por ejemplo:

- Ciudadanos que utilizan los datos de matrículas y admisión para asesorar a sus hijos en la elección de estudio.
 - Docentes de dentro y fuera de España que usan los conjuntos de datos como material de prácticas en asignaturas relacionadas con el tratamiento de datos.
 - Investigadores que usan los datos para llevar a cabo estudios académicos.
 - Estudiantes que usan los datos como fuente de sus trabajos de fin de estudios.
 - Empresas de gestión de alojamientos estudiantiles que explotan los conjuntos de datos de movilidad entrante y saliente en sus procesos de inteligencia de negocio.
 - Organismos supervisores que utilizan los datos de personal y presupuestos para llevar a cabo labores de contraste.
 - Otras universidades, que utilizan los datos publicados como información de contexto para compararse.
- Promoción de la reutilización de datos mediante la convocatoria del I Datathon Universidata, con [gran éxito de participación](#).

CALENDARIO DE IMPLANTACIÓN Y REFERENCIA TEMPORAL:

En términos generales, y gracias al alto nivel de automatización, en pocas semanas la universidad que se incorpore al proyecto puede publicar sus primeros conjuntos de datos, y en un par de meses actualizar todos los datos históricos publicados (por convenio, todas las universidades participantes publican datos históricos desde 2017). Los procesos de incorporación al proyecto tienen dos grandes ámbitos:

- **Ámbito organizativo:** en el que se atienden todos los requerimientos legales y organizativos de la iniciativa y se establecen los responsables y canales de comunicación.
- **Ámbito técnico:** en el que se concretan los aspectos relacionados con los formatos de intercambio, los canales de envío, etc.

DOCUMENTACIÓN DE CONSULTA Y APOYO:

- Portal web de Universidata: www.universidata.es
- Informes de madurez de los datos abiertos de la UE: <https://data.europa.eu/es/publications/open-data-maturity>
- España, entre los países que marcan tendencia en materia de datos abiertos: <https://datos.gob.es/es/blog/espana-entre-los-paises-que-marcan-tendencia-en-materia-de-datos-abiertos>

- Informes ASEDIE sobre el valor del Sector Infomediario: <https://www.asedie.es/es/informes-anuales>
- Ganamos el III Desafío Aporta: <https://www.dimetrical.es/es/noticias/ganamos-el-iii-desafio-aporta>
- Informe “A 3-year monitoring of 30 reuse cases to understand the economic, governmental, social and environmental impact of open data”: (UniversiDATA en página 44): https://data.europa.eu/sites/default/files/report/The_Use_Case_Observatory_Volume_I_0.pdf
- Vídeo-resumen de lanzamiento <https://www.youtube.com/watch?v=fJ1OVnZrGt>
- Vídeo del evento de presentación del proyecto: <https://www.youtube.com/watch?v=hhPdys903i4>
- Noticia de lanzamiento en datos.gob.es: <https://datos.gob.es/es/noticia/nace-universidata-el-portal-colaborativo-de-datos-abiertos-especializado-en-el-sector-de-la>