



e-cienciaDatos, EL REPOSITORIO DE DATOS DE INVESTIGACIÓN DEL CONSORCIO MADROÑO

e-cienciaDatos, THE RESEARCH DATA REPOSITORY OF THE MADROÑO CONSORTIUM

Autora:

Teresa Malo de Molina. Universidad Carlos III de Madrid. teresa.malo@uc3m.es.
ORCID 0000-0001-6863-7057

Resumen:

Este artículo describe las actividades del Consorcio Madroño en relación con su compromiso con el Acceso Abierto y con la Ciencia, especialmente en el desarrollo de e-cienciaDatos, su repositorio de datos de investigación. Se describen las características de este repositorio y su puesta en marcha, así como algunos de los proyectos que ha permitido desarrollar como ejemplos de buenas prácticas. Por último, se informa sobre las estadísticas de uso y los planes de futuro.

Abstract:

This article describes the activities of the Madroño Consortium in relation to its commitment to Open Access and to Open Science, especially in the development of e-cienciaDatos, his research data repository. It describes the characteristics of this repository and its implementation, as well as some of the projects that it has allowed to develop as examples of good practices. Finally, it is informed about use statistics and future plans.

Palabras clave: Datos FAIR; repositorio de datos de investigación; Consorcio Madroño

Keywords: FAIR Data; research data repository; Madroño Consortium

Uno de los pilares de la Agenda Europea de la Ciencia Abierta es todo lo relacionado con Datos FAIR, entendiendo por ello que los datos usados en la investigación sean localizables (*Findables*) mediante un identificador único persistente, un DOI;



accesibles (*Accessibles*) usando protocolos de comunicación estándares; Interoperables (*Interoperables*) mediante metadatos debidamente estandarizados; y reutilizables (*Reusables*) porque los datos están perfectamente descritos y cuentan con licencias claras de uso.

La importancia de contar con Datos FAIR es esencial porque permite la reproducibilidad de la ciencia, es decir, comprobar que los resultados de investigación que se han publicado son veraces y, lo que es más importante, que los datos pueden ser reutilizados para dar nuevos resultados o matizar los existentes.

La preocupación por la publicación de los datos de investigación se inició en 2014 cuando la Comisión Europea incorpora un plan piloto para su publicación en el marco de Horizonte 2020 que afectaba exclusivamente a algunas áreas temáticas. En estas áreas, se solicitaba a los investigadores la incorporación de un Plan de Gestión de Datos en la petición del proyecto y se exigía la publicación en abierto de los datos finales de investigación. En el inicio de 2017, este piloto se extendió a todos los proyectos incluidos en Horizonte 2020.

Antecedentes

El Consorcio Madroño es el consorcio de las bibliotecas de las universidades públicas de la Comunidad de Madrid: Universidad de Alcalá (UAH), Universidad Autónoma de Madrid (UAM), Universidad Carlos III de Madrid (UC3M), Universidad Politécnica de Madrid (UPM) y Universidad Rey Juan Carlos (URJC)¹, además de la Universidad Nacional de Educación a Distancia (UNED) cuya biblioteca principal se encuentra en Madrid. El Consorcio Madroño tiene como misión *Ser una red de conocimiento como soporte esencial para la innovación e investigación científica de las instituciones miembro, y proporcionar unos servicios de excelencia a sus usuarios con un uso eficiente de los recursos* y desde su origen su principal objetivo ha sido apoyar la investigación de las universidades miembro.

Es por eso que ha estado comprometido con el movimiento Open Access casi desde su inicio, y ya en el año 2007 puso en marcha [e-Ciencia](#), un agregador de todos los

¹ La Universidad Complutense de Madrid abandonó el Consorcio Madroño en el año 2016.

repositorios institucionales de las universidades miembro al que, con el tiempo, se han ido agregando otros centros de investigación de la Comunidad de Madrid. Actualmente cuenta con más de 490.000 documentos.

Figura 1
Buscador de e-ciencia

En junio del año 2013, el Consorcio Madroño publicó su [Declaración en apoyo al Acceso Abierto a la Información Científica y Académica](#), en la que dice que *comparte los principios del Movimiento de Acceso Abierto, y desea impulsar la adopción de una serie de medidas que faciliten la transición de las universidades hacia el paradigma de "acceso abierto al conocimiento" dentro de sus posibilidades y competencias*. La declaración incluye recomendaciones para los autores, para las universidades y para las administraciones públicas.

Desarrollo del portal PaGoDa

Con la puesta en marcha del plan piloto de datos de Horizonte 2020, el Consorcio Madroño se propone apoyar a sus investigadores en la presentación del Plan de Gestión de Datos que hay que incluir en la petición de proyectos y desarrolla el Portal PaGoDa en donde, además de incluir diferentes recursos, se ofrece una nueva herramienta, [PGDOnline](#), diseñada para ayudar a los investigadores en la creación

del Plan de Gestión de Datos, de acuerdo con los requerimientos de la Comisión Europea. Este PGDOnline es la traducción y adaptación al español de la herramienta DMPOnline, diseñada por el *Digital Curation Centre* de Reino Unido y es, en realidad, un tutorial que intenta responder a los principales temas a cubrir en el Plan de Gestión de Datos según la tabla *FAIR Data Management at a glance: issues to cover in your Horizon 2020 DMP*, incluida en las [Guidelines on FAIR Data Management in Horizon 2020](#). Esta herramienta está desarrollada sobre software libre (MYSQL, Ruby on Rails).

PDGOnline permite crear el Plan inicial de datos que puede ser revisado de manera intermedia y siempre debe tener una revisión final. El plan puede compartirse con diferentes colaboradores en sus diferentes etapas y finalmente permite exportarlo como un documento que es lo que, al fin y al cabo, hay que incluir en la propuesta.

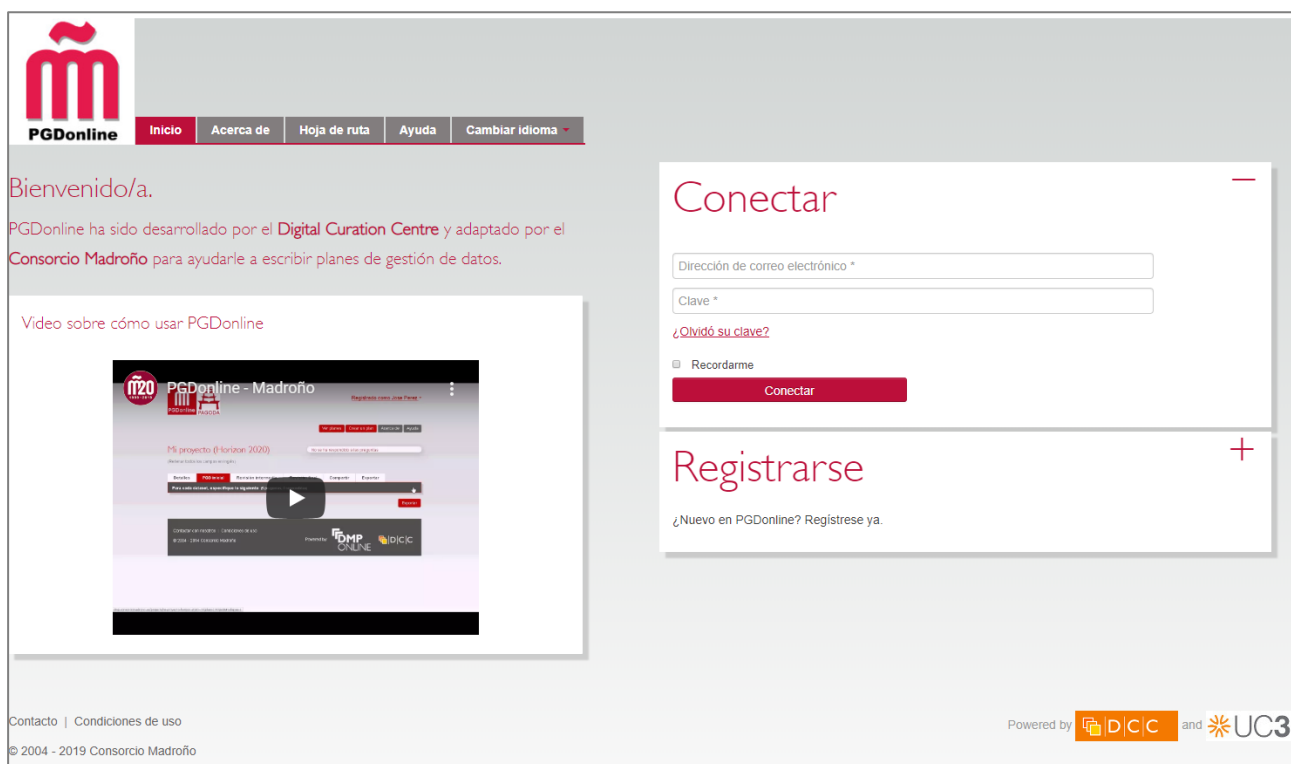


Figura 2
Pantalla de acceso a PGDOnline

Además de esta herramienta, PaGoDa incluía una selección de recursos útiles sobre planificación de gestión de datos, guías útiles sobre gestión de datos de investigación y algunos ejemplos de planes de gestión de datos. PaGoDa tuvo un gran éxito, no



solo entre los investigadores del Consorcio Madroño, sino que se convirtió en referente para las bibliotecas universitarias españolas, ya que para usar PGDOnline no es necesario ser miembro del Consorcio Madroño. basta con crearse una cuenta. Hasta el momento, se han creado 530 Planes de Gestión de Datos con esta herramienta.

Creación y desarrollo de e-cienciaDatos

Después de crear PaGoDa, había que dar respuesta a una nueva necesidad de los investigadores. Una vez que se concluye el proyecto y se empiezan a publicar los resultados, también es necesario publicar los datos. Así pues, Madroño se plantea como paso siguiente la creación de un repositorio de datos que permitiera a los investigadores del Consorcio cumplir con el requerimiento de publicación de Horizonte 2020.

Así, en junio de 2015, se crea un Grupo de Trabajo específico² con uno o dos representantes de cada universidad miembro y coordinado por un Director de Biblioteca y una Vicerrectora, además del Director Técnico del Consorcio, lo que da idea del compromiso del Consorcio con este objetivo.

Este grupo elabora el documento funcional que sienta las bases del proyecto que fue aprobado por el Consejo de Gobierno del Consorcio Madroño en diciembre de 2015.

El objetivo del proyecto era la creación y mantenimiento de un repositorio para albergar los datos de investigación gestionados por los investigadores del Consorcio, con el fin de dar visibilidad a estos datos, potenciar su acceso, permitir su reutilización y garantizar su preservación.

² El Grupo de Trabajo, coordinado por Ianko López, Eva Méndez y Teresa Malo de Molina, está integrado por las siguientes personas: Ainara Cisneros y Rosa Gallego de la UAH, M^a Luisa Pérez Aliende de la UAM, Victoria Rasero, Eva Ortiz y Rebeca Marín de la UC3M, Alicia López Medina de la UNED, Nacho González y Marcos Cuesta de la UPM, Marisa Sansegundo de la URJC y Juan Corrales y Fernando González del Consorcio Madroño.

Figura 3
Pantalla de acceso a e-cienciaDatos

Características de e-cienciaDatos

- e-cienciaDatos se estructura como un sistema centralizado constituido por distintas comunidades que agrupan los datasets o conjuntos de datos de cada una de las universidades, permitiendo así una futura disgregación si así interesara.
- Tras hacer una evaluación de las herramientas existentes, se decide utilizar el software libre del proyecto [Dataverse](#).
- Los datos están en acceso abierto por defecto, excepto si existen razones legales, éticas o de transferencia de tecnología que lo impiden, lo que implica el cumplimiento de criterios de uso, reutilización y redistribución sin restricciones.



- Los datasets se describen conforme a uno o varios estándares de metadatos como Dublin Core y DataCite, para garantizar la compatibilidad con la infraestructura europea de investigación [OpenAIRE](#).
- Por supuesto, se implementa el protocolo OAI-PMH para configurar el repositorio como proveedor de datos, lo que permitirá la recolección de los metadatos por diferentes agregadores de contenido.
- Un requisito indispensable es que el repositorio asigne identificadores de objetos digitales (DOIs) a los datasets y a los ficheros.
- Además, proporciona una cita normalizada para cada dataset que incluye información para que el dataset pueda ser identificado y localizado, incluyendo su DOI.
- Se admite cualquier formato de datos, aunque se recomienda que se sigan formatos estandarizados y lo más abiertos posible dentro del área temática en el que se trabaja y que se detallen instrucciones sobre cómo usar los datos y sobre el software que puede interpretarlos.
- Aunque por necesidades prácticas de almacenamiento hay un límite de 2GB por fichero, no existe ningún límite en el número de ficheros y los administradores de cualquier institución pueden subir ficheros mayores.
- e-cienciaDatos incluye control de versiones lo permite guardar el historial de cambios de versión y poder ver el contenido de versiones anteriores de cada dataset.
- Sólo se aceptan datos finales de la investigación, no datos intermedios manejados a lo largo de la vida del proyecto.
- En la licencia de depósito se hace referencia a la cesión de los derechos de explotación de manera no exclusiva al repositorio.
- Se incluyen licencias libres para los datasets: CC0, CC-BY, CC-BY-SA y Open Data Commons PDDL. Si un investigador necesita excepcionalmente otra licencia o una combinación de las mismas, deberá solicitarlo expresamente a los administradores, También se podrán aplicar licencias específicas para software,

aprobadas por la [Open Source Initiative \(OSI\)](#), GNU General Public License Versions y MIT license entre otras.

- El depósito de los datos en e-cienciaDatos se hace mediante archivo delegado. Son las bibliotecas de cada universidad las que se ocupan de hacerlo, por lo que los investigadores deberán solicitárselo a su biblioteca.
- Por último, cada dataset, deberá ir acompañado de un fichero de texto en formato txt (texto plano) que, bajo el título de *Readme*, explique en detalle la metodología empleada para la recogida o generación de datos y el software para su uso, además de información general, autoría, palabras clave, cobertura geográfica, cobertura temporal y licencia de uso.

A continuación, se presenta un ejemplo de Datasets con todos sus componentes:

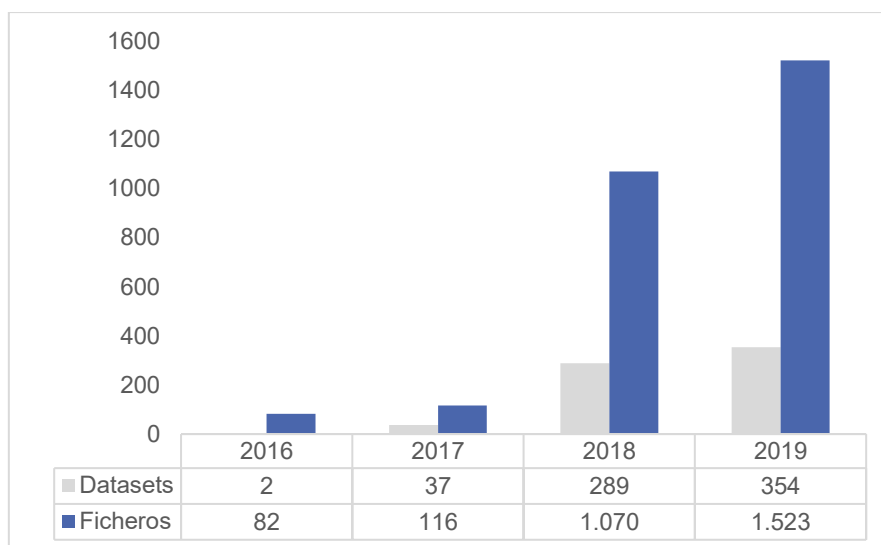
The screenshot shows the dataset page for 'Historia de la Lengua Española de Ramón Menéndez Pidal: Romanización. Latín vulgar y literario. Época visigótica'. The page includes a header with the UAM logo and navigation links, a breadcrumb trail, and a download button for 8 files. The dataset title is 'Historia de la Lengua Española de Ramón Menéndez Pidal: Romanización. Latín vulgar y literario. Época visigótica' (Versión 1.1). The description states that the dataset provides access to manuscript fragments from Ramón Menéndez Pidal's archive. The matter is 'Arts and Humanities' and the key words are 'Filología, Historia de la lengua, Lingüística, Lengua española'. The related publication is 'Repositorio Institucional de la UAM, Biblos-e Archivo'. Below the description, there are tabs for 'Ficheros', 'Metadatos', 'Condiciones', and 'Versiones'. A search bar is present. The file list shows two files: 'readme_RMFA01C002.txt' (4,8 KB) and 'RMFA01C002.tab' (885,0 KB).

Descripción	El presente conjunto de datos describe y proporciona acceso a las papeletas manuscritas por Ramón Menéndez Pidal en el archivo de la Historia de la Lengua Española, correspondientes al archivador 1, cajón 2 en el que figura el título: "Imperio. Áreas germánicas". La organización de contenidos permite diferenciar las siguientes secciones: Elemento púnico y fenicio; colonias griegas; romanización; latín vulgar, general e hispano. Fin de las lenguas indígenas; latín literario. Causas y forma del cultismo. Elemento griego. Elemento germánico; época visigótica (414-711). Causas de la diferenciación dialectal románica y carácter de los principales romances (2018-07)
Materia	Arts and Humanities
Palabra clave	Filología, Historia de la lengua, Lingüística, Lengua española
Related Publication	Repositorio Institucional de la UAM, Biblos-e Archivo

Ficheros	Metadatos	Condiciones	Versiones
Buscar en estos ficheros de datos... <input type="button" value="Buscar"/>			
1 a 2 de 2 Ficheros			
<input type="checkbox"/>	readme_RMFA01C002.txt Texto plano - 4,8 KB - 08-oct-2018 - 6 Descargas MDS: 41465f9a0bba779e1323e66dacc035 Documentación	<input type="button" value="Descargar"/>	<input type="button" value="Descargar"/>
<input type="checkbox"/>	RMFA01C002.tab Datos tabulares - 885,0 KB - 08-oct-2018 - 2 Descargas 8 Variables, 1512 Observaciones - UNF:6:mkE8kk+NBpnnE3Pss7CQ== Datos	<input type="button" value="Descargar"/>	<input type="button" value="Descargar"/>

Figura 4
Ejemplo de Dataset en e-cienciaDatos

En la actualidad, e-cienciaDatos alberga 354 datasets que incluyen 1.523 ficheros, con la siguiente evolución temporal:



Y hasta el momento se han realizado 7.926 descargas.

Además, e-cienciaDatos es un repositorio válido para OpenAIRE y los principales editores científicos desde diciembre de 2017.

Para llevar adelante el proyecto se recibió una ayuda específica de la *Dirección General de Universidades e Investigación de la Comunidad de Madrid* que ha permitido financiar la infraestructura tecnológica y, especialmente, un plan de formación para los bibliotecarios del Consorcio Madroño que se realizó de forma complementaria al lanzamiento de e-cienciaDatos, para que los bibliotecarios implicados en este proyecto puedan responder a las necesidades de los investigadores en la gestión de sus datos.

El curso se celebró los días 29 y 30 de noviembre de 2016 y asistieron cerca de 100 bibliotecarios de las universidades miembro. Todo el material de este curso está disponible en <https://www.consorciomadrono.es/investigam/curso-publicacion-de-datos-cientificos/> e incluye temas como la gestión de datos de investigación, el plan de gestión de datos, el uso de PDGOnline, metadatos para datos, modelos y políticas de datos, aspectos legales y derechos de autor en los datos y conocer y gestionar el repositorio de datos.

Por fin, para aunar todas las herramientas relacionadas con la publicación en abierto, tanto de los resultados como de los datos de investigación, el Consorcio Madroño crea el portal InvestigaM.

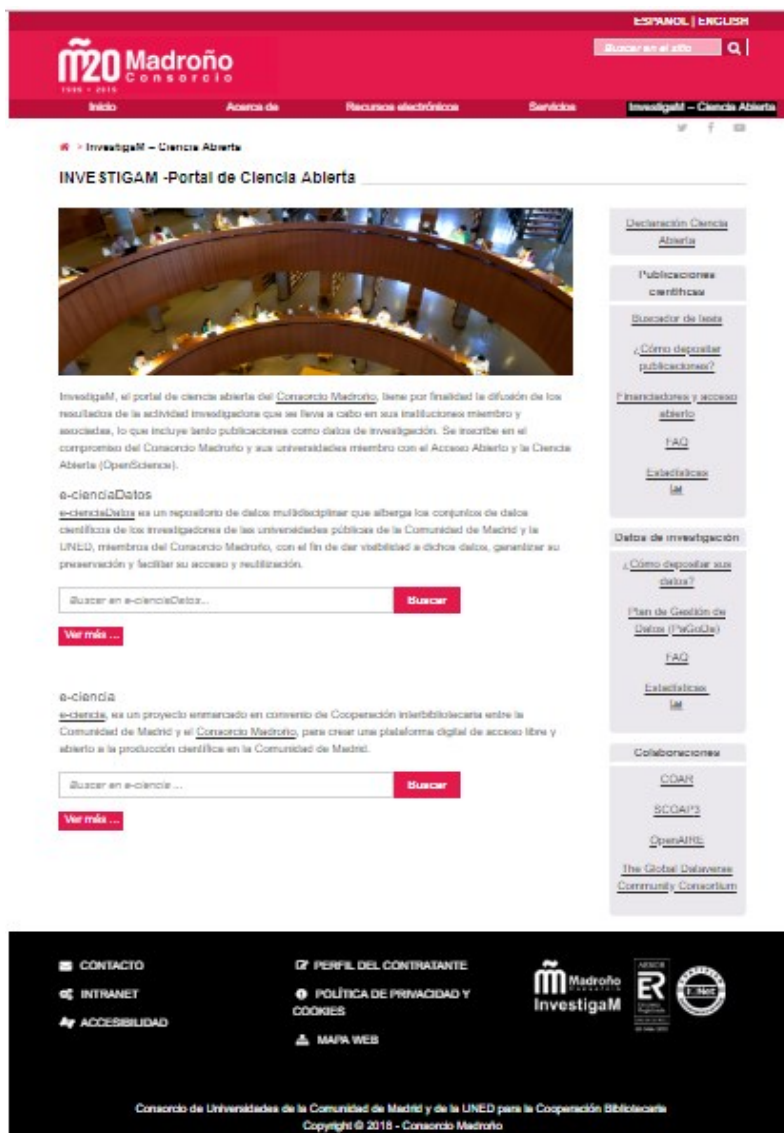


Figura 5
Portada de InvestigaM

En Diciembre de 2017, finalmente, el Consorcio Madroño firma su [Declaración en apoyo a la Ciencia Abierta a la información académica y científica](#), en la que se incluye un modelo de *Política para la Gestión de Datos de Investigación* en la que se definen las responsabilidades de los investigadores y de las universidades miembro respecto a la publicación en abierto de los datos de investigación.

Desde el Consorcio, además, se está intensificando la difusión del repositorio y de todo lo relacionado con la Open Science. Así, en la *Semana Internacional de Acceso Abierto* del año 2018, se realizó una campaña de difusión específica que tuvo mucho eco en base a estas dos infografías:

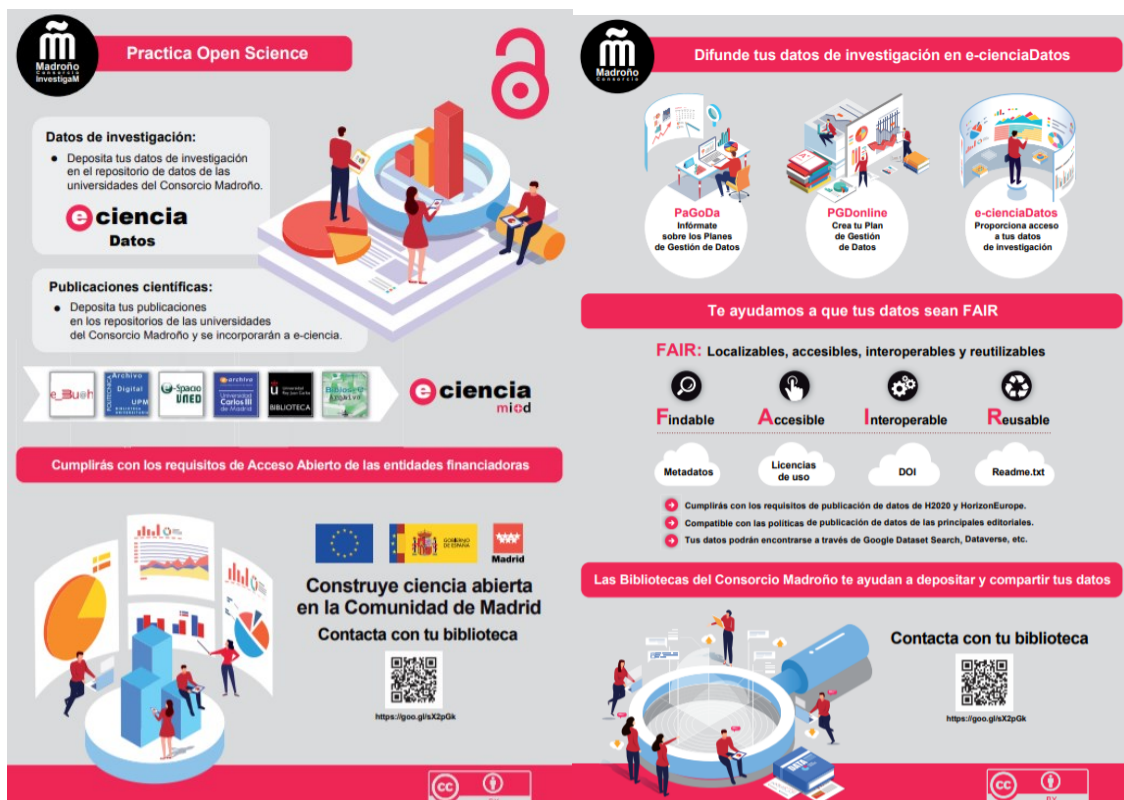


Figura 6
Infografías del Consorcio Madroño sobre Open Science y los Datos FAIR

Además, en 2018 e-cienciaDatos ha obtenido la [mención de honor](#) en la categoría de 'Open Data' de los premios *Open Education Awards for Excellence* que otorga el *Open Education Consortium*,

Buenas prácticas

Como dijimos al principio, el objetivo principal del Consorcio Madroño es apoyar la investigación de las universidades miembro. En este sentido, e-cienciaDatos se ha revelado como una herramienta muy útil para apoyar el desarrollo específico de proyectos relacionados con datos en los que la biblioteca ha ayudado no solo integrando los datos en el repositorio, sino también asesorando en la estructura,

presentación y difusión de estos. Presentamos aquí dos ejemplos de buenas prácticas realizados por dos bibliotecas del Consorcio.

La Biblioteca de la Universidad de Alcalá se está ocupando de incorporar en e-cienciaDatos los archivos que corresponden al corpus oral del Proyecto *Frontera hispano-portuguesa: documentación lingüística y bibliográfica* (FRONTESPO).

The screenshot shows the e-cienciaDatos interface. At the top, there is a navigation bar with the logo 'Madroneo Ciencia Datos' and links for 'Guía de usuario', 'Soporte', 'Español', 'Iniciar sesión', and 'Ver mapa'. Below the navigation bar, the project title 'Frontera España - Portugal: documentación lingüística y bibliográfica (FRONTESPO)' is displayed, along with a link to the library website. The main content area features a search bar with the text 'Buscar en este dataverse...' and a 'Buscar' button. To the left, there are filter options for 'Dataverses (0)', 'Datasets (153)', and 'Ficheros (1.016)'. Below these, there are filters for 'Fecha de publicación' (2018 (93), 2019 (80)), 'Autor: Nombre' (Álvarez Pérez, Xosé Afonso (coord.) (153)), 'Materia' (Artes y humanidades (153)), and 'Palabra clave: Término' (Contrabando (19), Ganado porcino (19), Transformaciones en las últimas décadas (19), Mobiliario (18), Utensilios (18)). The main search results area shows '1 a 10 de 153 Resultados' and a list of three items, each with a document icon, title, date, and a brief description. The first item is 'Informe 1 (Igreja, Cabreiro). Juicio lingüísticos y relación con los españoles' (05-abr-2019) by Álvarez Pérez, Xosé Afonso (coord.), 2019. The second is 'Informe 1 (Igreja, Cabreiro). Hombre: ser físico y social' (05-abr-2019) by Álvarez Pérez, Xosé Afonso (coord.), 2019. The third is 'Informe 1 (Igreja, Cabreiro). Plantas' (05-abr-2019) by Álvarez Pérez, Xosé Afonso (coord.), 2019.

Figura 7
Pantalla de acceso al proyecto FRONTESPO en e-cienciaDatos

El objetivo de este proyecto es la documentación lingüística exhaustiva del área fronteriza entre España y Portugal, tradicionalmente conocida como la *Raya*, y una parte muy importante del mismo ha sido la recolección en formato vídeo del corpus oral en las localidades fronterizas. Se ha trabajado con 64 localidades, distribuidas en 9 zonas de trabajo que se reparten a lo largo de la frontera. Para poder hacer una buena comparación, en cada zona de trabajo hay al menos dos localidades portuguesas y dos localidades españolas, y en cada sitio se entrevista a un mínimo de tres personas de diferente sexo y edad, aunque debido a las características demográficas de la zona, las categorías de edad son únicamente tres: mayores de 75 años, entre 50 y 75 años y menores de 50 años.



Los temas que se abordan en las entrevistas son:

- La vida tradicional (la vida en el campo, animales domésticos, alimentación, etc.).
- La vida en la *Raya* (contrabando, relaciones entre las aldeas, el modo de vida tradicional, etc.).
- Los juicios lingüísticos de la persona entrevistada (comparación con el habla de los pueblos vecinos, lengua usada para comunicarse con el otro lado de la *raya*, etc.).

En la actualidad se han integrado en e-cienciaDatos 182 datasets de este proyecto que incluyen 1.152 ficheros.

The screenshot shows the e-cienciaDatos interface for a dataset. At the top, there is a navigation bar with 'Guía de usuario', 'Soporte', 'Español', and 'Iniciar sesión'. Below this, the breadcrumb trail reads: 'Frontera España - Portugal, documentación lingüística y bibliográfica (FRONTESPO) > Informante 1 (Igreja, Cabreiro) > Juicio lingüísticos y relación con los españoles'. The dataset title is 'Informante 1 (Igreja, Cabreiro). Juicio lingüísticos y relación con los españoles'. The description section contains the following text: 'Juicio lingüístico: ¿Cómo cree que se habla aquí el portugués? El influjo de la escuela. Diferencias con las localidades más próximas. Las diferencias con la ciudad (eran objeto de burla por parte de su lengua). Los diferentes dialectos del portugués. ¿Cómo hablan cuando van a Galicia? Van a Galicia a comprar caramelos. El contrabando. Los gallegos hablan como nosotros, es otra pronunciación y unas palabras diferentes. Cuando un gallego habla, ¿se sabe que es gallego? ¿Ha aumentado el contacto transfronterizo desde Schengen? La influencia de otras lenguas en el cambio léxico. Juicios lingüísticos: como você acha que o português é falado aqui? A influência da escola. Diferenças com as localidades mais próximas. As diferenças com a cidade (os informantes foram ridiculizados pela sua língua). Os diferentes dialetos do português. Como falam quando vão à Galiza? Vão à Galiza para comprar docas. O contrabando. Os galiegos falam como nós, é outra pronúncia e palavras diferentes. Quando um gallego fala, sabe que ele é gallego? O contacto transfronteiriço aumentou desde Schengen? A influência de outras línguas na mudança léxica. Linguistic judgments: How do you think Portuguese is spoken here? The influence of the school. Differences with the nearest localities. The differences with the city (informants were subject to ridicule because of their language). Dialects of Portuguese. How do they talk when they go to Galicia? They went to Galicia to buy candies. The contraband. Galicians speak like us, it's another pronunciation and different words. When a Galician speaks, do you know that he is Galician? Has cross-border contact increased since Schengen? The influence of other languages on lexical change. Metadata includes: 'Materia: Arts and Humanities', 'Palabras clave: Comercio, Contrabando, Relaciones con las comunidades fronterizas, Comportamiento lingüístico con los vecinos del otro país, Lengua de las localidades vecinas del mismo país, Lengua de país vecino, Situación lingüística de la comunidad', 'Related Publication: Atas del IV Congreso Internacional de la SEEPLU, Cartografías del Português: Língua, Literatura, Cultura y Diáléctos en los Espazos Lusófonos, Cáceres: Sociedade Galega de Estudos Portugueses y de la Lusofonía, 2015, p. 55-72 (ISBN 978-84-16993-37-9)', and 'Notas: Web del proyecto: http://www.frontespo.org/es'. At the bottom, there is a search bar and a list of 5 files with download buttons.

Figura 8 Ejemplo de la visualización del registro de un dataset del proyecto FRONTESPO en e-cienciaDatos



Otro ejemplo de buenas prácticas es de la biblioteca de la Universidad Carlos III de Madrid, que ha prestado apoyo al proyecto *Federico-Tena World Trade Historical Database*. Se trata de un proyecto internacional realizado por un investigador de la Universidad de Pisa, Giovanni Federico y un investigador de la UC3M, Antonio Tena, que se han ocupado de recoger series anuales de datos de comercio por países desde 1800 a 1938, que se suman también como series por continente y como series mundiales.

The screenshot displays the website interface for the Federico-Tena World Trade Historical Database. At the top, there is a navigation bar with the logo of the Universidad Carlos III de Madrid and the text 'e-cienciaDatos'. Below the navigation bar, there is a search bar and a list of search results. The search results are organized into a table with columns for the dataset title, date, authors, and a brief description. The results include:

- Federico-Tena World Trade Historical Database : List of trading polities (World countries)** (04-feb-2019). Authors: Federico, Giovanni; Tena Junguito, Antonio. Description: Project developed by Giovanni Federico (Università di Pisa) and Antonio Tena Junguito (Universidad Carlos III de Madrid). This data base outlines changes in world trade from 1800 to 2016. Dataset: List of trading polities (World countries).
- Federico-Tena World Trade Historical Database : World Freights Rates to and from UK** (30-jul-2018). Authors: Federico, Giovanni; Tena Junguito, Antonio. Description: Project developed by Giovanni Federico (Università di Pisa) and Antonio Tena Junguito (Universidad Carlos III de Madrid). This data base outlines changes in world trade from 1800 to 2016. Dataset: World Freights Rates to and from UK.
- Federico-Tena World Trade Historical Database : World Share Primary Products Exports and Imports** (11-jul-2018). Authors: Federico, Giovanni; Tena Junguito, Antonio. Description: Project developed by Giovanni Federico (Università di Pisa) and Antonio Tena Junguito (Universidad Carlos III de Madrid). This data base outlines changes in world trade from 1800 to 2016. Dataset: World Share Primary Products Exports and Imports.
- Federico-Tena World Trade Historical Database : World Trade** (07-may-2018). Authors: Federico, Giovanni; Tena Junguito, Antonio. Description: Project developed by Giovanni Federico (Università di Pisa) and Antonio Tena Junguito (Universidad Carlos III de Madrid). This data base outlines changes in world trade from 1800 to 2016. Dataset: World Trade.

Figura 9

Pantalla de acceso al proyecto *Federico-Tena World Trade Historical Database*

Este proyecto se apoya en el uso por parte de biblioteca de cuatro plataformas que se relacionan entre sí: el repositorio institucional eArchivo, la web corporativa de la UC3M, el repositorio de datos e-cienciaDatos y la geolocalización con Google My Maps. Todo ello permite hacer una presentación de los datos de forma muy completa como puede verse en <https://www.uc3m.es/ss/Satellite/UC3MInstitucional/es/TextoMixta/1371246237481/Federico-Tena World Trade Historical Database>.

Este proyecto incorpora actualmente 161 datasets que incluyen 322 ficheros. Está previsto que los investigadores actualicen las series de datos en octubre de 2020.

The screenshot shows the user interface of the e-cienciaDatos repository. At the top, there is a red header with the Madroño logo and navigation links: 'Guía de usuario', 'Soporte', 'Español', 'Iniciar sesión', and 'Ver mapa'. Below the header, the breadcrumb trail reads: 'Federico-Tena World Trade Historical Database | www.uclm.es/ineciencia_datos'. The main content area features a 'e-cienciaDatos' logo, a breadcrumb trail: 'Repositorio de Datos UCLM > Federico-Tena World Trade Historical Database > Federico-Tena World Trade Historical Database : Spain', and a download button showing '47 Descargas'. A 'Ver mapa' link is also present. Below this, there is a section for the dataset: 'Federico-Tena World Trade Historical Database : Spain' with a 'Versión 1.1' label. A citation box contains the text: 'Federico, Giovanni; Tena Junguito, Antonio, 2018, "Federico-Tena World Trade Historical Database : Spain", <https://doi.org/10.21203/rs.3.rs-16699-cv1>, e-cienciaDatos, V1'. A 'Citar dataset' button is located to the right. Below the citation, there is a 'Revisar los [Requisitos de citas de datos](#)' link. A description box provides details: 'Descripción: Project developed by Giovanni Federico (Università di Pisa) and Antonio Tena Junguito (Universidad Carlos III de Madrid). This data base outlines changes in world trade from 1800 to 2018. Dataset: Spain (2017-18)'; 'Materia: Social Sciences'; 'Palabras clave: World Trade, Economic History, Historical Trade'; 'Related Publication: Federico, G., Tena Junguito, A. (2018). World trade, 1800-1929: a new dataset. EHES Working Papers in Economic History, n. 83'; 'Notas: Web site: http://www.uclm.es/ineciencia_datos Web: MAU'. Below the description, there are tabs for 'Ficheros', 'Metadatos', 'Condiciones', and 'Versiones'. A search bar with the text 'Buscar en estos ficheros de datos...' and a 'Buscar' button is present. At the bottom, there is a section for '1 a 2 de 2 Ficheros' with a table listing two files: 'Spain_1871_1929_FTWTHQ_201718_v01.pdf' (Adobe PDF - 101.2 KB - 14-mar-2018 - 36 Descargas) and 'Spain_1871_1929_FTWTHQ_201718_v01.xlsx' (Microsoft Excel (XLSX) - 27.2 KB - 14-mar-2018 - 21 Descargas). Each file has a 'Descargar' button.

Figura 10

Ejemplo de la visualización del registro de un dataset del proyecto *Federico-Tena World Trade Historical Database* en e-cienciaDatos

Conclusión y planes de futuro

El repositorio de datos del Consorcio Madroño es hoy una realidad consolidada y es el único repositorio de datos existente por el momento en las universidades españolas. Aunque todavía integra pocos contenidos, hay algunos ejemplos de buenas prácticas que auguran un futuro prometedor.



Sin embargo, aún quedan muchas cosas por hacer:

- Hay que intensificar su difusión y conocimiento entre los investigadores de las universidades miembro del Consorcio.
- Hay que abordar un plan de curación, preservación y acceso a largo plazo de los datos depositados en el repositorio.
- Sería conveniente que e-cienciaDatos pudiera dar soporte a distintos datos dentro del ciclo de vida de la investigación, es decir, datos iniciales e intermedios y no únicamente los datos finales como es actualmente.
- Habrá que ampliar la infraestructura y actualizar el software, teniendo en cuenta los desarrollos y tendencias internacionales en este sentido, así como la escalabilidad proyectada de *Dataverse*.
- Hay que estudiar si conviene añadir nuevos formatos de metadatos y vocabularios o Schemes de contenido.
- Conviene valorar el uso de tecnologías que faciliten la compartición de datos entre máquinas.
- Y, por último, es necesario crear nuevos servicios de valor añadido especializados.

Esperamos poder ir desarrollando estos planes de futuro para hacer de e-cienciaDatos el repositorio de datos de referencia entre los investigadores del Consorcio Madroño y un ejemplo de buena práctica para todas las bibliotecas universitarias.