

Terminología + Inteligencia Artificial

INFORME DE LA GESTIÓN DEL PROYECTO TERESIA

(Fase I)_extracto 2024

IG24-1-0-0-20241231

















ÍNDICE

444

1. INTRODUCCIÓN	3
2. GESTIÓN	3
2.1. Gestión científico/técnica	3
2.2. Seguimiento del proyecto	4
3. COLABORACIONES	4
3.1. Reunión con productores de terminologías	4
3.2. Colaboración con el proyecto ComTClar	6
3.3. Colaboración con UNTerm (ONU), a través de la Dirección General del Español en el Mundo	6
3.4. Colaboración con la Dirección General de Traducción de la Comisión Europea	7
4. COMUNICACIÓN	7
5. RESULTADOS	8
5.1. Resultados entregados	8
5.2. Premio	8
6. ACTIVIDADES DE DIFUSIÓN, COMUNICACIÓN PÚBLICA, INTERCAMBIO CIENTÍFICO Y FORMACIÓN	9
6.1. Presentaciones en instituciones e infraestructuras internacionales	9
6.2. Presentaciones en el ámbito académico	11
6.3. Actividades monográficas	12
6.4 Medios de comunicación	13
6.5 Otras actividades con presencia de TeresIA	14
6.6 Actividades de formación	1/

1. INTRODUCCIÓN

El 2 de agosto se publicó en el BOE el **convenio** entre el CSIC y la Universidad Politécnica de Madrid, la Asociación Española de Terminología, el Centro Nacional de Supercomputación, el Instituto Cervantes y la Real Academia de Ingeniería, marco normativo necesario para el desarrollo del proyecto TeresIA.

El convenio especifica, entre otras cuestiones, la relación de tareas y resultados previstos para el tiempo de ejecución del proyecto. Uno de los materiales previstos es un informe de gestión de cada año de ejecución del proyecto. El correspondiente a 2024, ha sido entregado en su versión completa a la Secretaría de Estado de Digitalización e Inteligencia Artificial. El presente documento es un extracto del informe de gestión y recoge la información más relevante para el público en general; es decir, no contiene información de carácter interno o procedimiental. Concretamente, refleja algunas cuestiones de la gestión científico/técnico, el seguimiento del proyecto. En este extracto del informe de gestión de la primera fase del proyecto TeresIA (año 2024) se resumen las principales actividades: las colaboraciones interinstitucionales en TeresIA, las acciones de comunicación, los resultados y las actividades de difusión, comunicación pública, intercambio científico y formación del proyecto durante 2024.

No se detallan en este extracto aquí cuestiones de gestión administrativa o de gestión económica. En relación a la gestión científico/técnica y al seguimiento del proyecto, se apuntan algunas tareas sin entrar en detalles, que solo son relevantes en este momento para el equipo del proyecto y para la entidad financiadora.

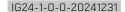
2. GESTIÓN

2.1 Gestión científico/técnica

Se incluyen en este apartado tres acciones clave para el proyecto, que requieren de un gran trabajo de gestión y coordinación: la elaboración del pliego para licitar el desarrollo del metabuscador, el establecimiento y configuración del servidor que albergará el corpus y el propio metabuscador y la constitución del corpus de literatura científica y técnica, base para la extracción de neología terminológica.

La construcción del metabuscador de terminologías es uno de los objetivos clave de TeresIA. Si bien la concepción, el diseño, el trabajo conjunto con productores de terminologías, el establecimiento de acuerdos con ellos y la conversión y protección de sus datos es un trabajo asumido por el equipo de investigación de TeresIA, el desarrollo de la herramienta tecnológica será asumido por una





empresa, por lo que ha de gestionarse una licitación, cuyo pliego se está ultimando en estos momentos.

Se está configurando el servidor en el que se desplegará el repositorio de datos y donde estará albergado el corpus con el que finalmente se va a trabajar y sobre el que se aplicarán los algoritmos de extracción de terminología a gran escala. Ya se ha trabajado con los algoritmos en otros servidores y con partes del corpus. El corpus completo estará compuesto por textos académicos y técnicos del ámbito iberoamericano, en el campo del Derecho y de la Biomedicina.

2.2 Seguimiento del proyecto

La especial naturaleza del proyecto TeresIA, tanto por los fondos que recibe (PRTR), como por la enorme colaboración interinstitucional que comporta, requiere de mecanismos de seguimiento.

Quincenalmente se celebran reuniones de todo el equipo de TeresIA para compartir toda la información del desarrollo del proyecto y para discutir cuestiones metodológicas. Periódicamente se celebran también reuniones monográficas sobre retos científicos y/o técnicos del proyecto que han de abordarse colectivamente (anotación, definición de corpus, diseño de formularios o herramientas para el trabajo, etc.).

Mensualmente se mantienen reuniones de seguimiento con la Secretaría de Estado de Digitalización e Inteligencia Artificial. Representan un ejercicio regular de rendición de cuentas ante la entidad financiadora. Este seguimiento mensual se complementa desde octubre con la cumplimentación de la tabla Excel proporcionada por la SEDIA en la que se va detallando el progreso del proyecto. Además, se están cumplimentando una serie de indicadores KPI propuestos por parte del equipo de TeresIA.

El **convenio** firmado entre las instituciones que forman parte de TereslA establece en su cláusula séptima la constitución de una comisión de seguimiento del convenio y la celebración de reuniones trimestrales de la misma. El pasado 11 de diciembre se celebró la reunión correspondienteⁱ, revisándose de manera detallada el Gantt del proyecto y reajustando cuestiones relativas a la planificación y consecución de objetivos.

3. COLABORACIONES

3.1. Reunión con productores de terminologías

El 3 de octubre se llevó a cabo una reunión clave para el proyecto en el Instituto Cervantes. Estuvo dirigida a los productores de terminologías y el objetivo fue



presentar con detalle el proyecto y el estado de desarrollo, buscando la colaboración de los productores de terminologías para dotar de contenido al metabuscador de TeresIA. En total asistieron 46 personas incluyendo al director del Instituto Cervantes, Luis García Montero. Estuvieron representadas unas 15 instituciones diferentes.



Orden del día de la reunión con productores de terminología

Con cada una de las instituciones se ha de establecer un marco de colaboración y un procedimiento técnico de tratamiento de los datos terminológicos, mediante los conversores desarrollados por la Universidad Politécnica de Madrid. La colaboración interinstitucional es clave en la construcción del metabuscador de TeresIA. La propuesta de integrar en el metabuscador datos terminológicos producidos por otras instituciones pretende facilitar el acceso a terminologías dispersas a usuarios generalistas y especializados (como puedan ser los traductores e intérpretes) y representa –por ello trabajamos- una serie de ventajas a los diferentes colaboradores, tal y como se muestra en la siguiente figura.



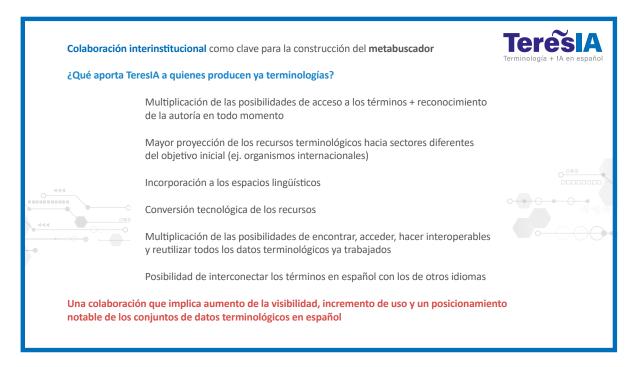


Figura I. Ventajas del metabuscador de TereslA para productores de terminologías

3.2. Colaboración con el proyecto ComTClar

En la formulación del proyecto TeresIA se plantean tres escenarios de uso, uno de ellos dedicado a la generación de terminologías en el ámbito jurídico. Aparte de otros corpus, se está concretando la colaboración con el proyecto **ComTClar**, que dirige Estrella Montolío, catedrática de Lingüística en la UB y directora de la cátedra de lenguaje claro. El trabajo terminológico en el ámbito tributario permitirá no solo avanzar en los objetivos de TeresIA, sino también contribuir a los objetivos del proyecto ComTClar por hacer el lenguaje más claro y facilitar la comunicación entre administración y administrados.

De este modo, estaríamos generando un impacto directo en la ciudadanía, un compromiso que no deberíamos perder de vista como investigadores.

Ya se han celebrado algunas reuniones con el equipo de trabajo de ComTClar, se están definiendo los ámbitos de la colaboración (trabajo con el corpus, anotación y validación, básicamente) y se ha hecho una breve presentación de TeresIA mediante un vídeo en la reunión de arranque del proyecto

3.3 Colaboración con UNTerm (ONU), a través de la Dirección General del Español en el Mundo

La presentación de TeresIA en Ginebra (referida más abajo), gracias a la mediación de la Dirección General del Español en el Mundo, ha permitido establecer contacto con la oficina de la ONU en Nueva York, responsable de la base de datos de terminología UNTerm de la ONU. Se pretende establecer un marco de colaboración con ellos para poder aunar los esfuerzos que se hacen en el ámbito terminológico.



3.4 Colaboración con la Dirección General de Traducción de la Comisión Europea

Desde el comienzo, la Dirección General de Traducción de la Comisión Europea es una institución colaboradora esencial en TeresIA. Aportan la visión de los traductores instituciones, transmiten las necesidades que TeresIA debería resolver y prestan apoyo en las reuniones periódicas, pero también en jornadas monográficas como la celebrada en diciembre en Madrid (véase la referencia más abajo) y en el paquete de trabajo relativo a la validación técnica y funcional de la solución. Su tarea diaria en la identificación de neologismos y en la necesidad de resolver el significando, el alcance y la traducción, valiéndose para ello del criterio de los especialistas disciplinares, está ayudando a definir un esquema de trabajo preciso y ágil.

Esta colaboración es crucial para TereslA y permite que el proyecto crezca con unas acciones de transferencia a públicos amplios muy bien definidas y realistas.

4. COMUNICACIÓN

Las acciones de comunicación en TereslA permiten informar a un público amplio, rendir cuentas de la actividad ante las instituciones que nos respaldan y ante la institución financiadora y, desde luego, son imprescindibles para poder articular todo el trabajo con colaboradores y destinatarios finales de los resultados de TereslA.

Durante esta anualidad, se han diseñado el logotipo del proyecto, la **web** y un **pasquín** informativo. En estos momentos está en producción el vídeo que resumirá TereslA y sus objetivos y que estará disponible en la web, así como en el canal **Youtube** que se ha creado para el proyecto.

También se han creado tres cuentas en otras redes sociales, a través de las cuales se proporciona información regularmente sobre la actividad del proyecto o asuntos relacionados con la terminología, interesantes para el mismo. Con estas redes creamos también una comunidad especializada en terminología, interesante para los objetivos del proyecto, pero también para la propia disciplina.

X: @ProyectoTeresIA (279 seguidores)

Bluesky: @proyectoteresia.bsky.social¹

LinkedIn: https://www.linkedin.com/company/105125823/proyecto-teresia

¹ Recientemente abierta, no tiene aún contenidos, aunque comenzarán a publicarse en 2025.



5. RESULTADOS

5.1. Resultados entregados

En el Excel de seguimiento que se proporciona a la SEDIA desde octubre de 2024 constan todos los detalles sobre los paquetes de trabajo del proyecto, el grado de desarrollo de los mismos, los resultados entregados o los problemas detectados en la ejecución de las distintas tareas. Por ello, en este apartado solo se resumen los resultados entregados hasta el momento o los que están próximos a ser entregados, conforme a la planificación. En la **sección de resultados** de la web de TeresIA y en el repositorio del proyecto en **GitHub** pueden consultarse los resultados entregados y publicados en acceso abierto hasta el momento. Sin embargo, algunos de los informes correspondientes a los entregables no van a ser públicos de manera inmediata debido a que algunas de sus partes son extractos de publicaciones científicas en proceso de revisión o edición. En esos casos, proporcionamos a la SEDIA acceso a los informes a través de nuestro repositorio institucional, SACO, en un enlace privado.

Los resultados en acceso abierto son, actualmente, los siguientes:

E2.1. Inventario inicial de recursos terminológicos (ICRTT-1.0.0-20240405):

Algunos de ellos candidatos a ser incluidos en el portal TeresIA. Entregado a la SEDIA en abril de 2024 y disponible en la **web** y en **GithHub**.

E2.2. Modelo de representación (ontología) de datos en TeresIA (IMROD-1.0.0-20240930):

Entregado a la SEDIA el 19/12/2024, puede consultarse en la sección de resultados de la **web** y en **GitHub**.

E3.3. Algoritmos para la extracción de terminologías y su evaluación (inicial) (APETE-1.0.0-20241230_e3.3):

KeyCARE, desarrollado por el BSC. Publicado el código en **GitHub**, disponible también en la **sección de resultados** de la web.

Extractor de relaciones genérico basado en relaciones de Wikidata (mREBEL), desarrollado por el OEG-UPM.

Los software van acompañados de un informe, disponible en la web y en GitHub.

5.2 Premio

El 17 de mayo de 2024, el Proyecto TereslA obtuvo el **Premio de Internet** (convocado por Google, Telefónica, etc.) en la categoría emprendimiento e investigación. La entrega del premio tuvo lugar durante la Gala de entrega de los Premios de Internet en el Senado.



6. ACTIVIDADES DE DIFUSIÓN, COMUNICACIÓN PÚBLICA, INTERCAMBIO CIENTÍFICO Y FORMACIÓN

Desde la concesión del proyecto, se ha llevado a cabo una estrategia continuada de presentación pública del proyecto, dirigida tanto al público general como a un público más especializado, buscando, en este segundo caso, propiciar el acercamiento y la colaboración de otros proyectos, instituciones, profesionales e investigadores. A continuación se relacionan y detallan las presentaciones:

6.1. Presentaciones en instituciones e infraestructuras internacionales

· Presentación en Bruselas

El 11 de diciembre de 2023 se presentó el proyecto TeresIA en la delegación institucional del CSIC en Bruselas, con la colaboración del Instituto Cervantes de la ciudad belga. Se enmarcó en la Presidencia española del Consejo de la UE y contó con la presencia de representantes de la Dirección General del Español en el Mundo.

Véanse las notas de prensa en diferentes webs institucionales: PRTR, CSIC, Instituto Cervantes, UPM, AETER



Elea Giménez presenta TeresIA en Bruselas. Miguel Marañón (a la derecha) fue también ponente en el acto.

Presentación de TeresIA en la ONU (Ginebra)

En el marco del Día Internacional de la Traducción, celebrado en la sede de la ONU en Ginebra el 30 de septiembre de 2024, el proyecto TeresIA se consolidó como una iniciativa clave para el fortalecimiento del español en los organismos internacionales. La **jornada**, organizada por la Misión de España en Ginebra y la Dirección General del Español en el Mundo (parte del Ministerio de Asuntos



IG24-1-0-0-20241231

Exteriores, Unión Europea y Cooperación), estuvo dedicada a resaltar el papel de la terminología como base esencial para el español como lengua de ciencia y comunicación internacional.

El evento principal se llevó a cabo en el Palais des Nations, reuniendo a un gran número de traductores e intérpretes de la ONU y otros organismos internacionales. En este marco, Guillermo Escribano, titular de la Dirección General del Español en el Mundo, presentó un ambicioso plan de acción enfocado en promover el español en los organismos multilaterales. A continuación, Elea Giménez, investigadora del CSIC, expuso los avances y aplicaciones del proyecto TeresIA, que se centra en el desarrollo de herramientas de inteligencia artificial para la terminología en español, con aplicaciones destacadas en traducción e interpretación institucional.

· Reuniones bilaterales en Ginebra

También el 30 de septiembre de 2024, la delegación española celebró **reuniones** con representantes clave de Naciones Unidas en Ginebra, como los Embajadores del Grupo de Amigos del Español, responsables del Centro de Lenguas y Multilingüismo, y líderes de los servicios de interpretación y traducción de la ONU como la directora de la división de gestión de conferencias de la ONU, Rima Al-Chikh; la jefa de servicio de interpretación, Mónica Varela y la jefa del servicio de traducción, Michelle Keating. Durante estos encuentros se abordaron oportunidades de colaboración con el proyecto TeresIA, con resultados muy positivos tanto en términos de acogida como de las colaboraciones emergentes.

xxx Reunión Técnica Anual de Latindex

En la sesión *Multilingüismo* e *infraestructuras* de *información* del 19 de septiembre, en Casa de América, Elea Giménez presentó el proyecto TeresIA enmarcado en las acciones de impulso al multilingüismo computacional.

OPERAS Conference 2024

OPERAS es la infraestructura europea sobre comunicación académica en abierto en Humanidades y Ciencias Sociales. En su **congreso anual**, celebrado en abril de 2024 en la ciudad de Zadar (Croacia), la investigadora principal de TereslA organizó el seminario Research for mulitilingualism in Science. Durante el mismo se presentaron tres proyectos: TereslA, Mondaecus y Translations and Open Science, estos últimos sobre traducción y ciencia abierta. Se exploraron las confluencias de los tres proyectos. Participaron unas 60 personas y hubo un debate muy relevante sobre el trabajo terminológico en la actualidad, en relación con la traducción. Esa acción ha supuesto la proyección de TereslA hacia una comunidad académica internacional involucrada en el multilingüismo en la ciencia.

Presentación oficial de la infraestructura europea OPERAS en el CSIC

El 7 de octubre, se **presentó** oficialmente la infraestructura europea sobre comunicación académica en abierto en Humanidades y Ciencias Sociales (OPERAS) en



España. En un acto organizado por el CSIC, Elea Giménez intervino como investigadora principal del CSIC en OPERAS, incidiendo en las posibles aplicaciones e interacciones de TeresIA con otros proyectos desarrollados en el marco de esta infraestructura.

Annual Event CLARIN Barcelona 2024

El 15 de octubre, Elena Montiel (UPM), Paula Díez (UPM) y Patricia Martín-Chozas (UPM) hicieron una presentación de TereslA en el **evento anual de CLARIN**. Con el título *The Spanish INESData and TereslA Projects as Potential Contributors to CLARIN* buscaron generar sinergias en el marco de CLARIN, Infraestructura Común de Tecnología y Recursos Lingüísticos, conformada por centros de 23 países.

6.2 Presentaciones en el ámbito académico

 EELISAConnect workshop: Teaching and Learning Language Technologies (text and speech) in light of Generative AI

Los días 7 y 8 de marzo, Elena Montiel organizó en la UPM la **EELISAConnect workshop**, iniciativa de EELISA InnoCORE, rama de investigación e innovación de la European Engineering Learning Innovation and Science Alliance (EELISA), con el título: *Teaching and Learning Language Technologies (text and speech) in light of Generative AI*, en el que pudo presentar TeresIA, junto a Patricia Martín-Chozas (UPM).

 II CONGRESO INTERNACIONAL Traducción y Sostenibilidad Cultural: Retos y nuevos escenarios

El 17 de abril, Nava Maroto y Joaquín García Palacios, ambos de AETER, participaron en este **congreso internacional de traducción**, en la sesión *Retos de la inteligencia artificial en la terminología para la traducción*, presentando el proyecto: «Terminología e Inteligencia artificial: el proyecto TeresIA».

La inteligencia artificial y la industria cultural

El 7 de mayo, Elea Giménez Toledo participó en el Encuentro de profesionales sobre *La inteligencia artificial y la industria cultural*, organizado por la Institución Libre de Eenseñanza, con la ponencia *Inteligencia artificial*, *español y ciencia*, con la que pudo explicar las aplicaciones de la IA en la extracción de terminología a partir de textos científicos

Jornada Científica: Gestión de datos en Humanidades y Ciencias Sociales

Jornada organizada por la PTI Ciencia Digital del CSIC, el 30 de mayo, en la que participaron Pablo Calleja (UPM) y Elea Giménez (CSIC). La jornada abordó la producción y gestión de datos en proyectos de Humanidades y Ciencias Sociales. Se presentó TeresIA y las particularidades del trabajo con los datos terminológicos. La presentación permitió que TeresIA fuera más conocido dentro del CSIC.



xx Jornada Científica REALITER y xxIII Jornada AETER

El 24 y 25 de octubre tuvieron lugar en San Millán de la Cogolla, la XX Jornada Científica REALITER y XXIII Jornada AETER, en las que participaron Joaquín García Palacios (AETER), Nava Maroto (AETER) y Elena Montiel (UPM), presentando TeresIA.

6.3 Actividades monográficas

Seminario en el Instituto de Lengua, Literatura y Antropología

El seminario *Investigación interdisciplinar en terminología: TeresIA* fue impartido por Elea Giménez, el 25 de noviembre de 2024, ante científicos y técnicos del ámbito de las Humanidades y Ciencias Sociales (CCHS), dentro del seminario permanente del Instituto de Lengua, Literatura y Antropología (ILLA).

· Jornada El factor humano en la validación terminológica

La Dirección General de Traducción de la Comisión Europea organiza anualmente una jornada especializada en algún asunto concerniente a la traducción institucional, haciendo partícipes a profesionales y académicos. Para la jornada de 2024, propusieron al equipo de TereslA centrar la jornada en la validación terminológica, un aspecto clave en la traducción en las instituciones europeas y, por otra parte, un asunto central en el proyecto TereslA, del que ellos participan. Como resultado, se organizó conjuntamente la jornada *El factor humano en la validación terminológica* el 13 de diciembre. El equipo de TereslA participó en el diseño del programa e intervino en varias de las presentaciones.

El público objetivo fueron traductores, intérpretes, profesionales de la lengua en general, investigadores en terminología, lingüística computacional, IA, humanidades digitales, etc. Entre los 132 asistentes, la mayoría responden a ese perfil. La concentración de la jornada en una mañana hizo que la práctica totalidad de los asistentes estuvieran en todas las presentaciones. Fue un intercambio muy valioso para concretar y conectar el trabajo cualitativo de los humanos en el trabajo terminológico con los desarrollos que se están haciendo en IA. Además, la intervención de traductores e intérpretes puso de manifiesto hasta qué punto es necesario el trabajo que propone TeresIA para el ejercicio de su profesión, así como la enorme disposición a colaborar para que salga adelante. Durante la jornada se mostraron los principales resultados del proyecto hasta el momento.

El equipo de TeresIA muestra su satisfacción con la jornada, tanto por lo mostrado como por lo aprendido. Los encuentros así resultan clave para perfilar el proyecto en su vertiente de transferencia, pero también para crear la comunidad necesaria para que TeresIA sea realmente un proyecto colaborativo de ciencia abierta, en el que todos hemos de aportar.





Programa de la Jornada El factor humano en la validación terminológica

6.4 Medios de comunicación

Enmarcadas en la estrategia de llegar al público más amplio posible, se han aprovechado las intervenciones en medios de comunicación, para hablar de TereslA y acercar las aplicaciones de las nuevas tecnologías a disciplinas de humanidades a un público más amplio, como son estos dos ejemplos en radio:



IG24-1-0-0-20241231

- Programa «Un idioma sin fronteras» de RNE. Programa monográfico de la relación entre la ciencia y la lengua: Ciencia y lengua: Elea Giménez (20/01/2024).
- Programa «A hombros de gigantes» de RNE (30/12/2024): a partir del minuto 20, Elea Giménez presenta y explica el proyecto TeresIA.
- Programa **«El ojo crítico»** de RNE (29/12/2024): Entrevista a Elea Giménez, que a partir del minuto 19 habla acerca del proyecto TeresIA.

6.5 Otras actividades con presencia de TeresIA

- Mesa redonda del ciclo «Desayuno con palabras»: Desayuno con palabras:
 Ray Loriga y Elea Giménez Toledo. Instituto Cervantes, 03 de octubre de 2024.
- **Presentación del libro** de Fernando A. Navarro, *Medicina en español VII*, prologado por Elea Giménez. Instituto Cervantes, 28 de mayo de 2024.
- **XVIII Jornada MEDES**: Elea Giménez moderó la Sesión 3: «Humanizando la comunicación médico-paciente a través de la tecnología y el arte», Cursos de verano de El Escorial, 16 de agosto de 2024.

6.6 Actividades de formación

 Sesión de formación en la Dirección General de Traducción de la Comisión Europea

El 12 de diciembre de 2023 se desarrolló una sesión de formación destinada a traductores del Parlamento Europeo, el Consejo Europeo o la Comisión, celebrada en la sede de la Dirección General de Traducción de la Comisión Europea de Bruselas, aunque estuvo abierta también a los traductores de Luxemburgo. La sesión estuvo centrada en los objetivos de TeresIA y, sobre todo, en las aplicaciones que tendrá para traductores institucionales. Impartida por Elea Giménez, contó también con el apoyo de José Luis Vega Expósito (DGT), Luis González (DGT) y Miguel Marañón (Instituto Cervantes).

Curso UIMP

El 28 de agosto, Elea Giménez participó en el curso UIMP *Los datos en investigación: retos y oportunidades*, impartiendo la clase *Bases de datos en investigación: Humanidades* y participando en la mesa redonda «Colaboración entre científicos de datos y expertos de áreas, ¿qué se puede mejorar?». Puede consultarse la presentación de la clase en **DigitalCSIC**.



