ENTREGABLE D2.1

Informe sobre la metodología de recogida de datos

CIEDES

Número de Proyecto: 101213796
Convocatoria: CERV-2024-CHAR-LITI
Autoridad concedente: Agencia Ejecutiva Europea de Educación y Cultura
Duración del Proyecto: 24 meses

Cofinanciado por la Unión Europea

Financiado por la Unión Europea. Sin embargo, las opiniones y puntos de vista expresados son exclusivamente los del autor o los autores y no reflejan necesariamente los de la Unión Europea ni los de [nombre de la autoridad que concede el préstamo]. Ni la Unión Europea ni la autoridad que concede el préstamo se responsabilizan de ellos.

Detalles contractuales del Proyecto

Título del proyecto	Red de Tolerancia contra los delitos de odio
Acrónimo del proyecto	ReTo
N.º del acuerdo de subvención	101213796
Fecha de inicio del proyecto	01.06.2025
Fecha de finalización del proyecto	31.05.2027
Duración	24 meses
Website	El proyecto no tendrá una página web específica, pero cada socio creará una página dedicada a este proyecto en la página web de su organización.

Detalles del entregable

Paquete de trabajo No.	WP2	Tarea No.	T2.1	Paquete de trabajo No.
Título del documento	Informe sobre la metodología de recogida de datos			
Nivel de difusión	Publico			
Fecha de entrega	30/09/2025			

Colaboradores del documento

Responsable del documento	Fundación CIEDES		
Colaboradores del documento	Organización	Colaboradores del documento	Organización
Mª Carmen García Peña	Fundación CIEDES	Mª Carmen García Peña	Fundación CIEDES
Elena María Mesa Alcaraz	Fundación CIEDES	Elena María Mesa Alcaraz	Fundación CIEDES
Miriam Fernández Núñez	Fundación CIEDES	Miriam Fernández Núñez	Fundación CIEDES

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

Descargo de responsabilidad

Financiado por la Unión Europea. Sin embargo, las opiniones y puntos de vista expresados son exclusivamente los del autor o los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Ejecutiva Europea en el Ámbito Educativo y Cultural (EACEA). Ni la Unión Europea ni la autoridad que concede el préstamo se responsabilizan de ellos.

Historial del documento

Formulario	Versión	Fecha	Autor	Organización	Revisor
Borrador	1.0	22/09/2025	María del Carmen García Peña Elena María Mesa Alcaraz Miriam Fernández Núñez	CIEDES	Martha Goyeneche Valentín González
	2.0	25/09/2025	María del Carmen García Peña Elena María Mesa Alcaraz Miriam Fernández Núñez	CIEDES	Martha Goyeneche Valentín González
	3.0	30/09/2025	María del Carmen García Peña Elena María Mesa Alcaraz Miriam Fernández Núñez	CIEDES	Martha Goyeneche Alfiya Urazayeva Valentín González

Cofinanciado por la Unión Europea

ReTo
Red de Tolerancia contra los delitos de odio

Contenido

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

1.Introducción

El presente informe técnico constituye el Entregable D2.1 del Work Package 2 (WP2) del proyecto ReTo – Red de Tolerancia, centrado en la "Investigación y Análisis de Datos sobre Delitos de Odio". Su propósito, es detallar de forma rigurosa la metodología adoptada para la recogida, tratamiento y análisis de datos empíricos vinculados al discurso de odio y los delitos motivados por prejuicio, con un foco geográfico inicial en Andalucía y proyección escalable a nivel nacional y europeo.

El diseño metodológico responde a una visión estratégica e interseccional alineada con los objetivos del proyecto, que incluyen: (1) el desarrollo de metodologías sensibles al género para la recolección sistemática de datos; (2) la creación de una base de datos centralizada, desagregada por variables clave (género, edad, origen étnico, discapacidad, entre otras); y (3) la integración de tecnologías avanzadas para la monitorización y análisis de tendencias en tiempo real.

El WP2, liderado por la Fundación CIEDES y en coordinación con CIFAL Málaga y Movimiento Contra la Intolerancia (MCI), pretende desarrollar e implementar metodologías integrales y sensibles al género para la recopilación y el análisis sistemático de datos sobre el discurso y los delitos de odio, completándolo con la creación y el mantenimiento de una base de datos centralizada que registre los incidentes de odio desglosándolos por variables clave como el género, la edad, la etnia y la discapacidad.

Esta metodología no sólo proporciona solidez científica y trazabilidad técnica, sino que permite generar evidencia útil para la formulación de políticas públicas, la capacitación de actores institucionales y comunitarios, y el diseño de intervenciones focalizadas.

Los objetivos específicos que persigue el diseño de esta metodología son:

1. Sistematizar y categorizar la información existente para mejorar su claridad y utilidad.
2. Ampliar y enriquecer el contenido con recomendaciones estratégicas, ejemplos prácticos y consideraciones técnicas adicionales.

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

3. Proponer una arquitectura de datos integrada que conecte la recolección, el almacenamiento, el análisis y la visualización, posicionando la IA como un pilar central.

La complejidad y multidimensionalidad del fenómeno del discurso y los delitos de odio exige una aproximación metodológica rigurosa, interdisciplinar y sensible a las desigualdades estructurales que afectan de manera diferenciada a diversos grupos sociales. Por ello, el enfoque metodológico adoptado en el WP2 del proyecto ReTo no sólo responde a criterios de validez científica, sino también a principios éticos, interseccionales y de utilidad operativa para las políticas públicas.

Este enfoque se justifica por tres razones fundamentales:

1. Necesidad de evidencia desagregada y contextualizada: La falta de datos sistemáticos y desagregados sobre discursos y delitos de odio, especialmente en algunas regiones como Andalucía, limita gravemente la capacidad institucional y comunitaria para prevenir, intervenir y reparar los daños causados por estas formas de violencia. El diseño metodológico de ReTo permite capturar esta información mediante herramientas mixtas y tecnologías de análisis automatizado, ofreciendo una base empírica sólida y replicable.

2. Compromiso con un enfoque sensible al género e interseccional: Las manifestaciones del odio no son neutras. Afectan de manera desproporcionada a mujeres, personas LGTBIQ+, minorías étnicas y religiosas, personas con discapacidad, entre otros colectivos. Por ello, se integran metodologías que visibilizan la discriminación múltiple y estructural, permitiendo el desarrollo de intervenciones más justas y efectivas.

3. Aprovechamiento de tecnologías avanzadas para la monitorización en tiempo real: La integración de herramientas basadas en inteligencia artificial, combinadas con sistemas de visualización interactiva, dota al proyecto de capacidades de análisis predictivo y seguimiento continuo. Esto amplía el impacto de la metodología más

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

allá del ámbito académico, facilitando la toma de decisiones informadas por parte de actores públicos y sociales.

Pero la metodología adoptada en el WP2 no se concibe de forma aislada, sino como un pilar transversal del proyecto ReTo. Este paquete de trabajo actúa como el eje estructurante para generar una base empírica común que alimenta y sustenta el diseño, implementación y evaluación de los demás Work Packages del proyecto. Así, el enfoque metodológico no solo responde a la necesidad de capturar con precisión la realidad del discurso y los delitos de odio en España, sino que también permite generar conocimiento útil y accionable para los socios implicados.

En particular, el WP2 provee los datos y análisis que informan las prioridades temáticas y territoriales del WP1 (Coordinación), asegurando que las estrategias colaborativas se basen en evidencias contrastadas. Del mismo modo, ofrece insumos cruciales para el WP3 (Apoyo a víctimas), ya que permite identificar perfiles de riesgo, tipologías de incidentes y lagunas en los sistemas actuales de notificación y reparación, fomentando un enfoque verdaderamente centrado en la víctima y sensible al género. En el ámbito de la comunicación, los hallazgos del WP2 se vinculan directamente con el WP4 (Periodismo ético), proporcionando evidencias sobre las narrativas de odio que circulan en medios y redes sociales, y sirviendo como base para la formación de profesionales de la información. Asimismo, en el WP5 (Deporte e inclusión), los datos obtenidos permiten localizar espacios deportivos con alta incidencia de discriminación, orientando intervenciones focalizadas para promover entornos seguros e inclusivos. Finalmente, el WP6 (Intervención cultural) se nutre de los análisis del WP2 para segmentar campañas por perfil, canal y sensibilidad temática, optimizando así su impacto y relevancia.

En definitiva, el enfoque metodológico del WP2 no solo legitima su valor técnico y científico, sino que garantiza la coherencia interna del proyecto ReTo, fortaleciendo su impacto y capacidad de réplica a escala nacional y europea. Al situar la recopilación y análisis de datos como núcleo articulador, ReTo se posiciona como una iniciativa innovadora, eficaz y alineada con los principios de la Agenda 2030, el Plan de Acción contra el Racismo de la UE y el marco de derechos fundamentales.

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

2. Marco conceptual: definiciones operativas

2.1. Marco legal internacional y nacional

La Organización para la Seguridad y la Cooperación en Europa (OSCE), en consonancia con el Alto Comisionado de las Naciones Unidas para los Derechos Humanos (ACNUDH), define los delitos de odio (*hate crimes*) como: "Actos delictivos motivados por prejuicios hacia características reales o percibidas de una persona o grupo, tales como la raza, etnia, religión, nacionalidad, orientación sexual, identidad de género, discapacidad u otras características protegidas."

Esta definición tiene dos elementos esenciales:

- Un acto tipificado como delito en el derecho penal (por ejemplo, agresión, amenazas, vandalismo).
- Una motivación basada en prejuicio o intolerancia hacia la víctima debido a su pertenencia (real o percibida) a un grupo social específico.

Es decir, no todo discurso de odio (*hate speech*) es un delito de odio: el discurso puede ser legalmente sancionable o no, pero el delito de odio siempre implica un acto ilícito (violencia, amenazas, daños, etc.) motivado por odio o discriminación.

La definición de discurso de odio que nos proporciona la ONU es: "Cualquier forma de comunicación de palabra, por escrito o a través del comportamiento, que sea un ataque o utilice lenguaje peyorativo o discriminatorio en relación con una persona o un grupo sobre la base de quiénes son, en otras palabras, por su religión, etnia, nacionalidad, raza, color, ascendencia, género u otros factores de identidad."

Son tres las principales características que reúne un discurso de odio:

1. Puede transmitirse a través de cualquier forma de expresión, incluidas imágenes, caricaturas, memes, objetos, gestos y símbolos, y puede difundirse de manera online u offline.
2. Es "discriminatorio" (parcial, intolerante o fanático) o "peyorativo" (prejuiciado, despectivo o degradante) hacia un individuo o grupo.

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

3. Menciona "factores de identidad" reales o percibidos de un individuo o un grupo, entre ellos "religión, etnia, nacionalidad, raza, color, ascendencia, género", pero también características como el idioma, el origen económico o social, la discapacidad, el estado de salud o la orientación sexual, entre muchas otras.

La Comisión Europea establece como marco regulador la Decisión Marco 2008/913/JHA de la UE (*Council Framework Decision on combating certain forms and expressions of racism and xenophobia by means of criminal law*). En su artículo 1 indica: "Cada Estado miembro adoptará las medidas necesarias para asegurar que los siguientes comportamientos intencionales sean sancionables:

(a) incitar públicamente a la violencia o al odio dirigido contra un grupo de personas o un miembro de tal grupo definido por referencia a la raza, el color, la religión, el origen nacional o étnico o el linaje" (...)

La UE ha propuesto recientemente ampliar este marco para cubrir otras características protegidas (género, orientación sexual, edad, discapacidad, etc.) y para incluir directamente tanto delitos de odio como discurso de odio como áreas "EU crimes" bajo el Artículo 83(1) TFEU.

En España, el marco legal de los delitos y discursos de odio está principalmente en el Código Penal, complementado por normativa europea e internacional.

El art. 510 del Código Penal tipifica como delitos de odio:

- Incitar directa o indirectamente al odio, hostilidad, discriminación o violencia contra un grupo, una parte del mismo o contra una persona determinada por razón de su pertenencia a aquel, por motivos racistas, antisemitas, antigitanos u otros referentes a la ideología, religión o creencias, situación familiar, la pertenencia de sus miembros a una etnia, raza o nación, su origen nacional, su sexo, orientación o identidad sexual, por razones de género, aporofobia, enfermedad o discapacidad.
- Difundir informaciones falsas o injuriosas que inciten al odio o la violencia contra esos colectivos.

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

- Negar, trivializar gravemente o enaltecer delitos de genocidio, lesa humanidad o contra las personas en conflicto armado.
- Poseer, distribuir o difundir material que promueva el odio o la violencia.

Estos son delitos autónomos, aunque no exista otro delito añadido. Ejemplo: difundir un panfleto racista ya constituye delito.

El art. 22.4 del Código Penal establece como circunstancia agravante: "Cometer un delito por motivos racistas, antisemitas u otra clase de discriminación referente a la ideología, religión o creencias de la víctima, la pertenencia de sus miembros a una etnia, raza o nación, su sexo, orientación o identidad sexual, género, enfermedad o discapacidad."

Aquí no se crea un delito autónomo, sino que cualquier delito común (agresión, amenazas, daños, etc.) se castiga con pena más grave si la motivación fue el odio o la discriminación.

En cuanto al discurso de odio en España se regula principalmente en el artículo 510 CP, donde dice que no todo discurso de odio es delito: solo lo es cuando cruza los límites fijados en el Código Penal (incitación pública, difusión de mensajes, negacionismo, etc.). El Tribunal Constitucional y el Tribunal Supremo han establecido que se debe equilibrar con la libertad de expresión (art. 20 CE). Solo es punible cuando supone una incitación directa y clara al odio, hostilidad, discriminación o violencia.

A continuación, se presenta alguna normativa complementaria, junto con algunos planes y propuestas de acción, utilizados para perfilar el discurso de odio:

- Ley 4/2015, de Estatuto de la Víctima del Delito: reconoce a las víctimas de delitos de odio como especialmente vulnerables.
- Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación en España.
- Marco estratégico para la ciudadanía y la inclusión, contra el racismo y la xenofobia. (2023-2027)
- Ley de Servicios Digitales (DSA)
- Plan de Acción de Lucha contra los Delitos de Odio 2022-2024 (Ministerio del Interior, ONDOD): establece protocolos policiales y de coordinación institucional.

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

- Plan de Acción de la UE contra el Racismo 2020-2025**:** Este plan promueve medidas coordinadas a nivel europeo para combatir el racismo estructural.
- Recomendación CM/Rec(2022)16 del Consejo de Europa**:** Sobre la lucha contra el discurso de odio. Proporciona un marco no vinculante pero muy relevante para actuaciones públicas, incluyendo herramientas de recogida de datos, políticas educativas, y regulación en plataformas digitales.
- Informe anual de la FRA (Agencia de los Derechos Fundamentales de la UE): Contiene estadísticas comparadas entre Estados miembros, buenas prácticas y enfoques nacionales en materia de delitos de odio. Es útil como benchmark europeo.
- Marco metodológico Hatemedia ya adoptado por ReTo**:** No solo como herramienta tecnológica (monitor de odio), sino también como marco teórico-práctico basado en CRISPDM, con niveles de intensidad y tipologías del discurso de odio.
- Informe ODIHR – Hate Crime Data Collection and Monitoring Guidelines**:** Este documento técnico de la OSCE es una guía clave para estandarizar la recogida de datos, incluyendo la desagregación, el rol de la policía, y la colaboración con sociedad civil.
- Informe del Relator Especial de Naciones Unidas sobre racismo, discriminación racial y xenofobia: Algunos informes recientes han tratado el impacto del discurso de odio en redes sociales y su vinculación con violencia real.

De forma resumida, se recoge en el siguiente esquema las definiciones que se han tenido en cuenta:

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

FUENTE / MARCO LEGAL	DISCURSO DE ODIO	DELITO DE ODIO
NACIONES UNIDAS (ONU)	Cualquier tipo de comunicación que ataque o use lenguaje peyorativo contra una persona o grupo con base en su identidad	Acto criminal motivado por prejuicio hacia la identidad del grupo de la víctima (recomendado para tipificación legal)
COMISIÓN EUROPEA (CE)	Expresiones que difunden, incitan, promueven o justifican el odio basado en la intolerancia (raza, religión, orientación sexual...)	Infracción penal motivada por prejuicio u hostilidad hacia una característica protegida. Requiere sanciones específicas
ESPAÑA (MINISTERIO DEL INTERIOR)	Expresiones que incitan al odio, violencia o discriminación contra grupos vulnerables. Tipificado en el art. 510 del Código Penal	Acciones delictivas con motivación discriminatoria basadas en características protegidas. Definidas en el Código Penal.

Figure 1. Comparativa de definiciones operativas según marco territorial

2.2. Delimitación del concepto de discurso de odio para ReTo

El delito de odio queda claramente delimitado y registrado a través de las denuncias que se realizan, por lo que es fácil rastrearlo a través de medios digitales on line y off line. Sin embargo, el concepto de discurso de odio se mantiene más ambiguo y está sujeto a interpretaciones.

El Departamento de Seguridad Nacional (DSN), por su parte, en Trabajos del Foro contra las Campañas de Desinformación - Iniciativas 2024 señala que la mayoría de los discursos de odio detectados en España comparten cuatro elementos comunes:

- Grupo diana: personas del norte de África, musulmanas y afrodescendientes.
- Tipo de discurso: deshumanización grave y discurso agresivo explícito.

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

- Vinculación a eventos sobre inseguridad ciudadana.
- Divulgación de información falsa (bulos y fake news).

Desde su aprobación, el proyecto ReTo propuso abordar los discursos y delitos de odio conforme a una tipología basada en las formas de discriminación más reconocidas en el contexto europeo y español. Estas categorías, alineadas con el enfoque interseccional y de derechos humanos promovido por el proyecto, son las siguientes:

- Racismo y discriminación étnica: Incluye el odio basado en el grupo étnico, raza, color o ascendencia.
- Antisemitismo
- Antigitanismo
- Discriminación por orientación sexual: Contra personas LGBTQ+, incluyendo lesbianas, gais, bisexuales, transgénero y queer.
- Discriminación por identidad de género: Contra personas trans o no binarias.
- Intolerancia religiosa: Hostilidad hacia personas por sus creencias religiosas o convicciones personales.
- Discriminación por edad: Edadismo, especialmente contra personas mayores.
- Disfobia: Discriminación contra personas con discapacidad.
- Xenofobia: Odio hacia personas extranjeras o percibidas como tales.
- Homofobia: Aversión específica hacia la homosexualidad.
- Discriminación ideológica: Por razones de pensamiento político, filosófico o creencias personales.

Durante la sesión de trabajo celebrada el 17 de septiembre de 2025, se analizó la necesidad de adaptar y ampliar la tipología original del proyecto a partir de la evidencia empírica y los marcos utilizados por observatorios nacionales. En dicha reunión se compartió una clasificación más detallada basada en 18 categorías oficiales, tal como sigue:

- Racismo/Xenofobia
- Antisemitismo
- Islamofobia

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

- Orientación Sexual
- Identidad de Género
- Discriminación por Sexo/Género
- Ideología
- Creencias o Prácticas Religiosas
- Disfobia (Discapacidad)
- Aporofobia
- Antigitanismo
- Discriminación Generacional
- Discriminación por Razón de Enfermedad
- Aspecto Físico
- Situación de Dependencia
- Igualdad de trato
- Cultura
- Estado Civil.

Este listado, más exhaustivo, refleja las formas de odio reconocidas estadísticamente por el Ministerio del Interior en España, y está alineado con las metodologías de análisis actuales empleadas por proyectos como Hatemedia y plataformas de monitoreo de discurso de odio.

Finalmente, y en coherencia con el enfoque participativo del proyecto, se ha previsto una fase de consulta directa a todos los socios del consorcio ReTo, con el objetivo de validar y ajustar la clasificación definitiva que será utilizada para el análisis y la categorización de incidentes de odio. Esta clasificación consensuada se apoyará en los resultados de una encuesta estructurada cuyos resultados permitirán definir las temáticas prioritarias y los criterios técnicos de clasificación.

Al tratarse este informe de un documento vivo, que se irá modificando en el tiempo conforme se siga estudiando y trabajando la temática del proyecto, una vez que se complete la encuesta y se discutan sus resultados en el comité técnico, se integrará la clasificación acordada y se aplicará de forma homogénea en todos los trabajos relacionados con la recolección, análisis, reporte y visualización de datos.

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

2.3. Perspectiva intercultural e interseccional

La perspectiva intercultural adoptada por el proyecto ReTo se basa en el marco del Consejo de Europa sobre Integración Intercultural, que promueve la diversidad como un valor positivo y motor de cohesión social. Esta perspectiva parte de tres principios fundamentales:

- Igualdad de derechos, deberes y oportunidades para todos, independientemente de su origen o pertenencia cultural.
- Reconocimiento y valoración positiva de la diversidad, evitando su exotización o estigmatización.
- Interacción significativa entre personas de diferentes orígenes, como condición necesaria para evitar la segregación y prevenir el conflicto.

La aplicación práctica a este paquete de trabajo implicará que:

- Los datos se deben interpretar considerando el contexto cultural en el que se expresan los mensajes de odio.
- Se debe incluir un análisis semántico y simbólico en redes sociales para detectar referencias culturales locales, chistes discriminatorios, frases hechas o códigos expresivos propios de comunidades concretas.
- Se validará el etiquetado y la clasificación de mensajes con participación de actores sociales y comunidades afectadas, para evitar sesgos etnocéntricos o interpretaciones descontextualizadas.

2.4. Fundamentos éticos y legales

El enfoque interseccional está inspirado en los trabajos de Kimberlé Crenshaw y se utiliza para entender cómo múltiples factores de discriminación (como género, etnia, orientación sexual, discapacidad, clase, edad…) se entrelazan y amplifican en los discursos y delitos de odio.

Este enfoque considera que una persona puede estar sujeta a discriminaciones simultáneas y no aditivas, por lo que:

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

- Las experiencias de odio no son homogéneas dentro de un mismo colectivo.
- Las respuestas institucionales deben ser sensibles a la combinación de identidades.

La aplicación práctica en este paquete de trabajo supondrá:

- La base de datos integrará variables que permiten identificar patrones de discriminación múltiple (por ejemplo: mujer + migrante + LGTBI).
- Los análisis se diseñarán para captar diferencias de impacto del odio en función de múltiples ejes simultáneos.
- Se incluirán mecanismos de protección reforzada en la visualización y publicación de datos que podrían estigmatizar aún más a colectivos con múltiples vulnerabilidades.

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

3. Marco experiencial y estado del arte

El diseño metodológico del proyecto ReTo no parte de un vacío, sino que se construye sobre una base sólida de experiencias previas, iniciativas afines y referencias internacionales en la lucha contra los discursos y delitos de odio.

El WP2 capitaliza aprendizajes derivados de experiencias anteriores como el proyecto Hatemedia, que desarrolló un *Monitor de Odio* con integración de IA generativa para detectar, clasificar y visualizar mensajes de odio en tiempo real. ReTo adapta y amplía esta metodología mediante la incorporación de una librería propia de términos discriminatorios y entrenamiento de modelos supervisados con perspectiva de género e interseccional.

Asimismo, se toman como referencia los marcos conceptuales de iniciativas de organismos como la OSCE-ODIHR, la ECRI del Consejo de Europa, y los planes nacionales (p. ej., el II Plan de Acción contra los Delitos de Odio del Ministerio del Interior), reconociendo sus contribuciones al desarrollo de definiciones operativas y tipologías de odio, al tiempo que se identifican limitaciones persistentes en términos de cobertura territorial, enfoque interseccional y sistematización de datos.

El proyecto incorporará herramientas probadas en el ámbito del análisis del discurso de odio, permitiendo avanzar desde la mera detección hacia un análisis automatizado con capacidad explicativa, adaptado a las especificidades lingüísticas y culturales del contexto español.

Los socios del consorcio de ReTo han acumulado una trayectoria significativa en iniciativas relacionadas con derechos humanos, cohesión social y lucha contra la intolerancia. Este capital institucional garantiza una implementación metodológica informada por la realidad social, y capaz de movilizar redes locales para la validación y transferencia de resultados.

Limitaciones detectadas en el estado del arte

El análisis del estado del arte revela debilidades comunes: fragmentación de los datos, subregistro de incidentes, escasa interoperabilidad entre fuentes, y limitada integración de

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

variables interseccionales. La metodología de ReTo se propone responder a estas carencias mediante la creación de una arquitectura de datos robusta y replicable, con base en principios de calidad, anonimización, desagregación y trazabilidad técnica.

En este contexto, la propuesta metodológica del WP2 aporta una doble innovación. Por un lado, en términos tecnológicos, al diseñar un pipeline de análisis automatizado basado en scraping, IA y visualización dinámica. Por otro lado, desde el punto de vista conceptual, al articular un enfoque verdaderamente interseccional y sensible al género que permita segmentar tanto víctimas como tipologías de odio, y generar evidencia útil para el diseño de políticas públicas, intervenciones comunitarias y campañas de sensibilización.

Algunos de los proyectos y experiencias que se han tenido en cuenta para formular la metodología final que se propone en este documento son:

Nombre del Proyecto	HateLab
Entidad Impulsora	Universidad de Cardiff
Fecha de Creación / Vigencia	Activo (referencias recientes hasta 2025)
Breve Descripción	Plataforma académica que recopila y analiza datos globales sobre discurso y delitos de odio mediante visualización interactiva
Objetivos	Monitorear tendencias temporales y espaciales del discurso de odio, servir de apoyo a políticas públicas y sensibilización
Recursos / Herramientas Utilizadas	Paneles de visualización de datos, algoritmos de análisis lingüístico, recolección automatizada de datos sociales

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

Nombre del Proyecto	TILT
Entidad Impulsora	Moonshot CVE
Fecha de Creación / Vigencia	Activo
Breve Descripción	Sistema de monitoreo en tiempo casi real que detecta señales de radicalización y discurso de odio en línea para actuar de forma preventiva
Objetivos	Prevenir conflictos y violencia mediante detección temprana; proveer a gobiernos y ONG de alertas predictivas
Recursos / Herramientas Utilizadas	Aprendizaje automático, scraping de redes sociales, geolocalización de contenidos

Nombre del Proyecto	Perspective API
Entidad Impulsora	Jigsaw (Google)
Fecha de Creación / Vigencia	Desde 2017
Breve Descripción	Herramienta de análisis de lenguaje que evalúa la "toxicidad" de comentarios en línea usando modelos de IA
Objetivos	Identificar lenguaje ofensivo y fomentar entornos digitales más seguros
Recursos / Herramientas Utilizadas	Procesamiento de lenguaje natural, API accesible para desarrolladores, métricas de toxicidad

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

Nombre del Proyecto	Monitor de Odio – Hatemedia
Entidad Impulsora	Universidad Complutense de Madrid / Medialab UCM
Fecha de Creación / Vigencia	Desde 2022
Breve Descripción	Herramienta que detecta, clasifica y analiza el discurso de odio en medios digitales y redes sociales en España
Objetivos	Generar datos empíricos sobre odio en línea para intervención pública, activismo y estudios académicos
Recursos / Herramientas Utilizadas	Librería de 7.210 términos, modelos supervisados, dashboard, metodología CRISPDM

Nombre del Proyecto	HateXplain
Entidad Impulsora	Allen Institute for AI / Georgia Tech
Fecha de Creación / Vigencia	2020
Breve Descripción	Conjunto de datos anotados con explicaciones para clasificar discurso de odio, ofensivo o no ofensivo
Objetivos	Mejorar la explicabilidad de los sistemas de detección automática de odio
Recursos / Herramientas Utilizadas	Datos anotados por humanos, etiquetas multilabel, explicaciones justificadas

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

4. Enfoque Metodológico General

La distinción entre discurso de odio y delitos de odio es fundamental, tanto legal como operativamente. Integrarla correctamente es lo que hará que el proyecto ReTo sea preciso y útil para todos los actores (ONGs, periodistas, policía...).

El enfoque metodológico del WP2 se articula en torno a un ciclo de vida del dato, estructurado de manera integral desde la recolección hasta la visualización pública, pasando por el análisis y el aprendizaje automático. Para ello se adopta, adapta y amplía el modelo CRISP-DM (Cross-Industry Standard Process for Data Mining), que es ampliamente utilizado en procesos de ciencia de datos por su flexibilidad y aplicabilidad en proyectos interdisciplinarios.

Esta adaptación se alinea tanto con los objetivos del proyecto ReTo como con los aprendizajes obtenidos del proyecto Hatemedia, que constituye un referente metodológico relevante en el ámbito del monitoreo automatizado del discurso de odio.

En la siguiente tabla se recogen los elementos de adaptación de esta metodología al proyecto ReTo:

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

Fase CRISP-DM	Aplicación en WP2 – ReTo
1. Comprensión del negocio	Análisis de la problemática del odio en España, con foco en Andalucia. Alineación con marcos legales y sociales.
2. Comprensión de los datos	Identificación y evaluación de fuentes de datos (online/offline), estructura, formatos, sesgos y limitaciones.
3. Preparación de los datos	Limpieza, anonimización, estructuración y etiquetado manual/semi-automático de contenido sensible.
4. Modelado	Uso de librerias [ej Hatemedia], clasificación de contenido por tipologia e intensidad. Implementación de IA.
5. Evaluación	Validación participativa con expertos y comunidad. Comparación entre modelos IA y criterios humanos.
6. Despliegue	Visualización interactiva mediante dashboards. Generación de informes periódicos y alertas automáticas.

Figure 2. Adaptación del modelo CRISP-DM al WP2 del proyecto ReTo

El proyecto ReTo adopta y adapta varios elementos metodológicos clave desarrollados en Hatemedia, que se centró en la detección de expresiones de odio en medios digitales mediante técnicas de procesamiento del lenguaje natural e inteligencia artificial. A continuación, se detallan los aspectos más relevantes de esta adaptación:

a) Librería de Lemas y Expresiones

- Se parte de la librería de Hatemedia (7.210 lemas), la cual se enriquece con expresiones específicas de contexto andaluz, localismos, y variantes culturales.
- Se incorporan sinonimias contextuales, N-gramas y análisis sintagmático, claves para mejorar la precisión semántica.

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

b) Etiquetado y control de calidad

- El etiquetado sigue el enfoque Hatemedia, basado en niveles de intensidad (incívico, insulto, amenaza…) y tipología (xenófobo, machista, político, etc.).
- Se mejora mediante validación cruzada entre IA y humanos, además de revisión por pares, como mecanismo de control de sesgos.

c) Modelado IA

- Hatemedia empleó redes neuronales y análisis supervisado. ReTo añade el uso de APIs como Google Perspective y la posibilidad de entrenar modelos propios en español.
- Se busca integrar modelos como BiCapsHate o HateXplain, que permiten explicabilidad y mejor adaptación cultural.

d) Visualización y despliegue

- Siguiendo a Hatemedia, se propone una plataforma de visualización pública, con dashboards dinámicos alimentados en tiempo real.
- Se contempla la incorporación de segmentación por territorio y por tipología de odio, útil para administraciones y OSC.

Las principales ventajas que se han observado que aportarían el uso y adaptación en el proyecto ReTo del modelo desarrollado en Hatemedia son:

- Rapidez de implementación: uso de componentes ya desarrollados por Hatemedia como punto de partida.
- Contextualización: adaptación de modelos y taxonomías al territorio andaluz y español.
- Calidad y validación: combinación de modelos automáticos con revisión experta y comunitaria.
- Replicabilidad y sostenibilidad: todo el enfoque se documenta con vistas a su escalado nacional.

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

5. Fases del proceso metodológico

Esta propuesta metodológica por fases busca proporcionar una hoja de ruta clara y técnica para transformar el proyecto ReTo en una plataforma de referencia para el análisis y la prevención de los delitos de odio en España.

Figure 3. Fases del proceso metodológico

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

Fase 1: Recolección de Datos (Fuentes y Herramientas)

a) Selección de fuentes de datos y modelos de recogida de información off line.
b) Analizar los métodos y seleccionar el modelo de recogida de información online y su tratamiento: scraping, monitoreo, redes neuronales, librerías previas de datos.
c) Establecer criterios de calidad para seleccionar datos fiables o con valor analítico (p. ej., frecuencia, volumen, diversidad temática).
d) Establecer el modelo final de recogida de datos y alimentar la base de datos (BD) con el modelo mejorado.
e) Configurar fuentes de datos oficiales (OBERAXE, observatorio de datos del Ministerio del Interior).
f) Implementar un dashboard básico en Tableau/Power BI conectado a la BD.

Fase 2: Etiquetado, almacenamiento y Gestión (Diseño de Base de Datos)

a) Implementar la recolección automática de datos de redes sociales.
b) Integrar una API de IA (Google Perspective) para clasificación inicial.
c) Incluir módulo de anonimización y privacidad si se va a trabajar con datos sensibles.
d) Incorporar validación cruzada entre etiquetadores humanos e IA (control de calidad).
e) Afinar los dashboards con los KPIs definidos.

Fase 3: Procesamiento, análisis y limpieza (IA, Modelos y Conjuntos de Datos para Entrenamiento)

a) Validación de librerías con talleres lingüísticos y etiquetado de expertos.
b) Integración de N-gramas y expresiones culturales.

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

c) Generación de nuevas librerías de entrenamiento de IA.
d) Análisis y selección de indicadores simples y compuestos a generar de forma periódica.
e) Elaboración de documentación técnica de modelos entrenados y su replicabilidad.

Fase 4: Visualización y Publicación (Herramientas de BI y Dashboards)

a) Desarrollar y entrenar modelos de IA propios para el español.
b) Implementar un sistema de alertas tempranas automatizado.
c) Publicar informes periódicos automáticos.
d) Valorar la posibilidad de segmentación territorial y por tipología de odio en los dashboards (útil para OSC, ayuntamientos...).

Fase 5: Automatización e Integración (Conexión de la IA con la Base de Datos)

a) Propuesta de arquitectura integrada que permita la automatización en la actualización de datos y su visualización.
b) Conexión con bases de datos de distintos socios y entidades interesadas.
c) Establecer algún estándar de intercambio de datos o protocolo de interoperabilidad (por ejemplo, JSON, APIs RESTful seguras, etc.)

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

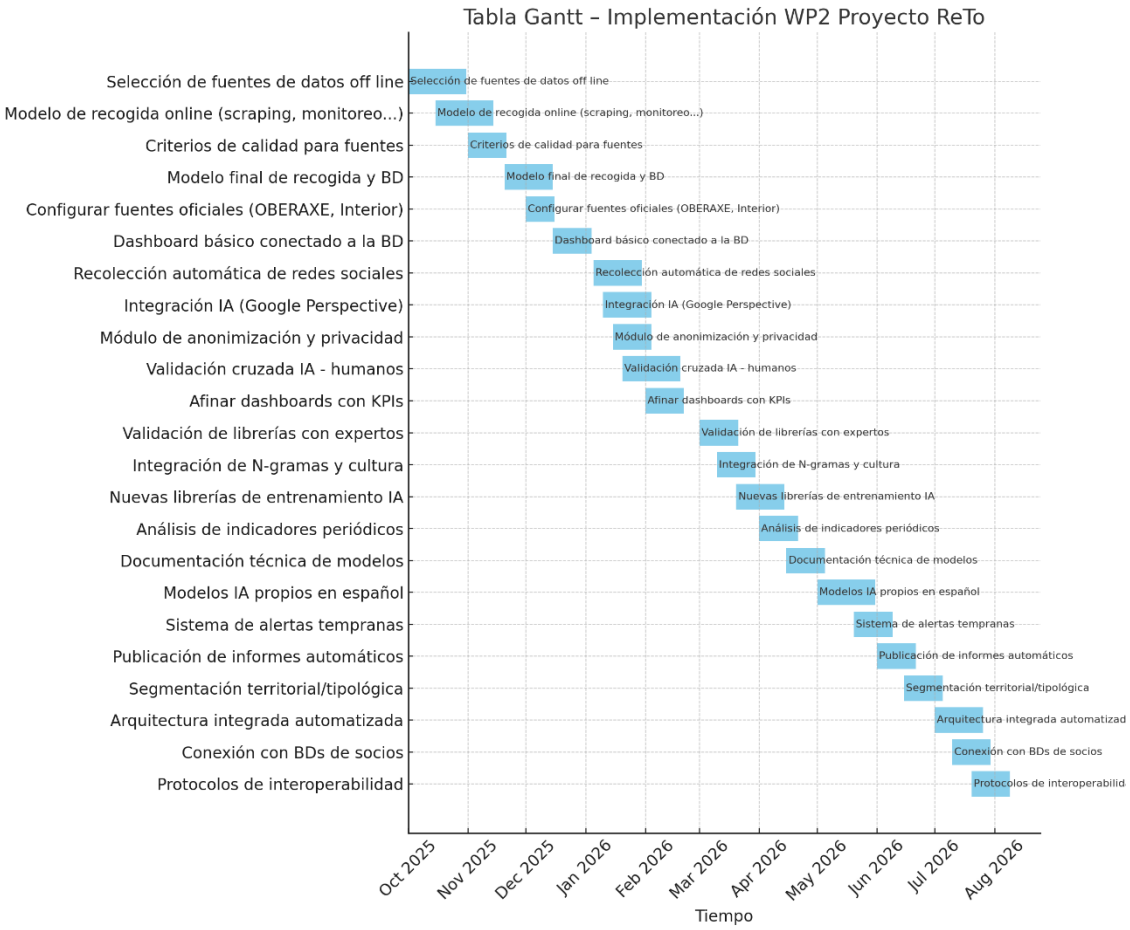

Figure 4. Diagrama de Gantt del WP 2

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

6.Validación participativa

La validación participativa en ReTo es una fase esencial del WP2, articulada conforme al enfoque metodológico CRISP-DM y a los compromisos recogidos en el Grant Agreement. Este proceso se ejecutará entre los meses 16 y 18 del proyecto e incluirá talleres co-diseñados con organizaciones de la sociedad civil, periodistas, instituciones públicas y comunidades marginadas, con el objetivo de revisar y afinar los resultados del análisis de datos sobre delitos y discursos de odio.

La validación incluirá:

- Revisión de resultados preliminares con enfoque interseccional.
- Verificación de la pertinencia cultural y lingüística de los términos de la librería de odio.
- Retroalimentación sobre la clasificación automática de discursos de odio (nivel de toxicidad, tipologías, etc.).
- Ajustes consensuados al dashboard y a los modelos de IA entrenados.

Esta fase busca asegurar que los productos del WP2 no solo sean técnicamente sólidos, sino también socialmente legítimos y útiles para las comunidades a las que se dirigen.

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

7. Indicadores clave de evaluación (KPIs)

El seguimiento y evaluación del WP2 se articula en torno a una batería de Indicadores Clave de Rendimiento (Key Performance Indicators – KPIs) que permiten medir de manera objetiva y verificable el grado de cumplimiento de los objetivos previstos. Estos indicadores han sido definidos en coherencia con el enfoque metodológico del proyecto ReTo, sus compromisos en la propuesta aprobada, y los estándares de calidad exigidos por el programa CERV.

La selección de KPIs incorpora tanto dimensiones cuantitativas como cualitativas, con especial atención a la trazabilidad de los procesos, la participación de los actores clave y la utilidad operativa de los productos generados.

Estos indicadores serán objeto de seguimiento periódico y actualizaciones y revisiones internas durante las reuniones técnicas de coordinación del WP2, y sus resultados alimentarán los informes de progreso y el informe final de resultados del proyecto ReTo.

Indicador	**Meta prevista a 24 meses**
Plataforma de monitoreo funcional operativa	**1**
Nº de entidades colaboradoras que suministran datos	**≥ 10**
Variables clave con datos desagregados (sexo, edad, etnia, etc.)	**≥ 5**
Nº de talleres participativos de validación desarrollados	**≥ 3**
Nivel de accesibilidad del dashboard a los socios y partes interesadas	**100%**
Nivel de satisfacción de los actores implicados en la validación participativa	**≥ 80% de satisfacción**

Los indicadores que se proponen se irán actualizando conforme avance el proyecto.

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

8. Consideraciones éticas

El WP2 integra salvaguardas éticas estrictas, en línea con el Reglamento General de Protección de Datos (RGPD), el marco ético del Programa CERV, y los principios internacionales (Declaración de Helsinki). Las medidas específicas incluyen:

- Anonimización de datos: Eliminación de identificadores directos e indirectos.
- Consentimiento informado: Formularios accesibles, adaptados a colectivos vulnerables.
- Revisión por pares: Supervisión metodológica en todas las fases de etiquetado y análisis.
- Sesgos algorítmicos: Monitorización continua de sesgos en los modelos de IA usados para detectar discurso de odio, con retroalimentación de colectivos afectados.
- Accesibilidad universal: Todos los productos (informes, dashboard, base de datos) se diseñarán bajo principios de accesibilidad digital y cognitiva.

Estas medidas aseguran que el WP2 no solo cumpla con estándares éticos, sino que los convierta en un eje estratégico del proyecto.

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

9. Escalabilidad y replicabilidad

La metodología desarrollada en el marco del WP2 está diseñada explícitamente para ser escalable a nivel nacional y replicable en distintos contextos territoriales y sociales. Esta intención se alinea con la estrategia general del proyecto ReTo, que utiliza Andalucía como región piloto con el objetivo de extender el modelo metodológico y tecnológico al resto de España y, potencialmente, al ámbito europeo.

La replicabilidad se sustenta en tres ejes fundamentales:

1. Documentación sistemática del proceso metodológico: Todos los procedimientos de recolección, tratamiento y análisis de datos se documentarán en un manual técnico abierto, asegurando su accesibilidad y comprensibilidad por parte de otras entidades e instituciones interesadas.
2. Infraestructura modular y adaptable: La arquitectura tecnológica va a ser diseñada para adaptarse a distintos niveles de complejidad, volumen de datos y recursos técnicos, permitiendo su implementación en entornos con diferentes capacidades.
3. Validación participativa e iteración contextual: La validación de la metodología se realizará mediante talleres y consultas con actores locales en Andalucía, cuyos resultados serán incorporados en el diseño final del modelo. Este enfoque permitirá transferir aprendizajes clave a otras regiones, asegurando una contextualización adecuada del modelo replicado.

La sostenibilidad del modelo se garantiza mediante:

- El uso de software de código abierto y componentes reutilizables.
- La compatibilidad con marcos legales y técnicos europeos.

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

- La implicación de actores públicos y privados desde la fase piloto.

En definitiva, el enfoque metodológico de ReTo está diseñado no solo para abordar eficazmente el fenómeno del discurso de odio en Andalucía, sino para servir de referente operativo a nivel nacional y europeo en el ámbito de la lucha contra la intolerancia, la discriminación y los delitos motivados por prejuicio.

Cofinanciado por la Unión Europea
ReTo Red de Tolerancia contra los delitos de odio

10. Referencias y fuentes de información

Artículos científicos

Alkhoury, L., Filippova, A., & Ustalov, D. (2024). *Emojis Trash or Treasure: Utilizing Emoji to Aid Hate Speech Detection*. ACL Anthology.

Aroyehun, S. T., & Gelbukh, A. (2020). *When Sarcasm Hurts: Irony-Aware Models for Abusive Language Detection*. SpringerLink.

Bassignana, E., Basile, V., & Patti, V. (2021). *HateXplain: A Benchmark Dataset for Explainable Hate Speech Detection*. arXiv.

Caselli, T., et al. (2023). *Exploring Boundaries and Intensities in Offensive and Hate Speech*. arXiv.

Chiruzzo, L., et al. (2023). *Spanish MTLHateCorpus: Multi-task learning for hate speech*. ScienceDirect.

Mozafari, M., et al. (2021). *EmojiRoBERTa: A Contextual Emoji Representation for Hate Speech Detection*. SpringerLink.

Paramita, M. L., et al. (2021). *DeepHate: Hate Speech Detection via Multi-Faceted Text Representations*. arXiv.

Xu, K., et al. (2020). *MMHS: Multimodal Model for Hate Speech Intensity Prediction*. SpringerLink.

Zhou, B., & Jurgens, D. (2021). *AngryBERT: Joint Learning Target and Emotion for Hate Speech Detection*. arXiv.

Fuentes y recursos en internet

Agencia de Derechos Fundamentales de la UE (FRA). **https://fra.europa.eu**

Departamento de Seguridad Nacional (DSN). **https://www.dsn.gob.es**

Hate Crime Data – OSCE. **https://hatecrime.osce.org/hate-crime-data**

Cofinanciado por la Unión Europea

ReTo Red de Tolerancia contra los delitos de odio

Ministerio del Interior (ONDOD). **https://www.interior.gob.es/opencms/es/servicios-al-ciudadano/ondod/**

Oberaxe – Ministerio de Inclusión. **https://www.inclusion.gob.es/oberaxe/**

Proyecto Hatemedia. **https://editorial.tirant.com**

RAXEN Reports – EUMC/FRA. **https://fra.europa.eu/en/project/2007/raxen**

RiSSC – Emore Project. **https://www.rissc.it/homepage/our-projects/emore-project/**

Tilt Monitor. **https://www.tiltmonitor.com**

UNDP iVerify. **https://www.undp.org/digital/iverify**

WeLiveSecurity (2023). *5 herramientas OSINT gratuitas para redes sociales*. **https://www.welivesecurity.com/la-es/2023/02/28/5-herramientas-osint-gratuitas-redes-sociales/**

Instituto de la Paz y los Conflictos (Universidad de Granada). **https://ipaz.ugr.es/proyectos-de-investigacion/**

Informe Raxen de Movimiento contra la Intolerancia

https://www.educatolerancia.com/informe-raxen/

Materiales Didácticos Movimiento contra la Intolerancia

https://www.educatolerancia.com/materiales-didacticos/

Cofinanciado por la Unión Europea

ReTo Red de Tolerancia contra los delitos de odio