

# PROYECTO HÉRCULES

## Aspectos Técnicos del proyecto: Enriquecimiento de Datos y Métodos de Análisis (EDMA)



**Jesualdo Tomás Fernández Breis**

CATEDRÁTICO, UNIVERSIDAD DE MURCIA



# Índice

- 01** **Introducción**
- 02** **Tecnologías propuestas**
- 03** **Enriquecimiento de datos:  
Alcance técnico**
- 04** **Métodos de análisis:  
Alcance técnico**
- 05** **Componentes I+D**
- 06** **Cooperación entre proyectos**

# Índice

- 01** **Introducción**
- 02 Tecnologías propuestas
- 03 Enriquecimiento de datos:  
Alcance técnico
- 04 Métodos de análisis:  
Alcance técnico
- 05 Componentes I+D
- 06 Cooperación entre proyectos



# Introducción al Proyecto

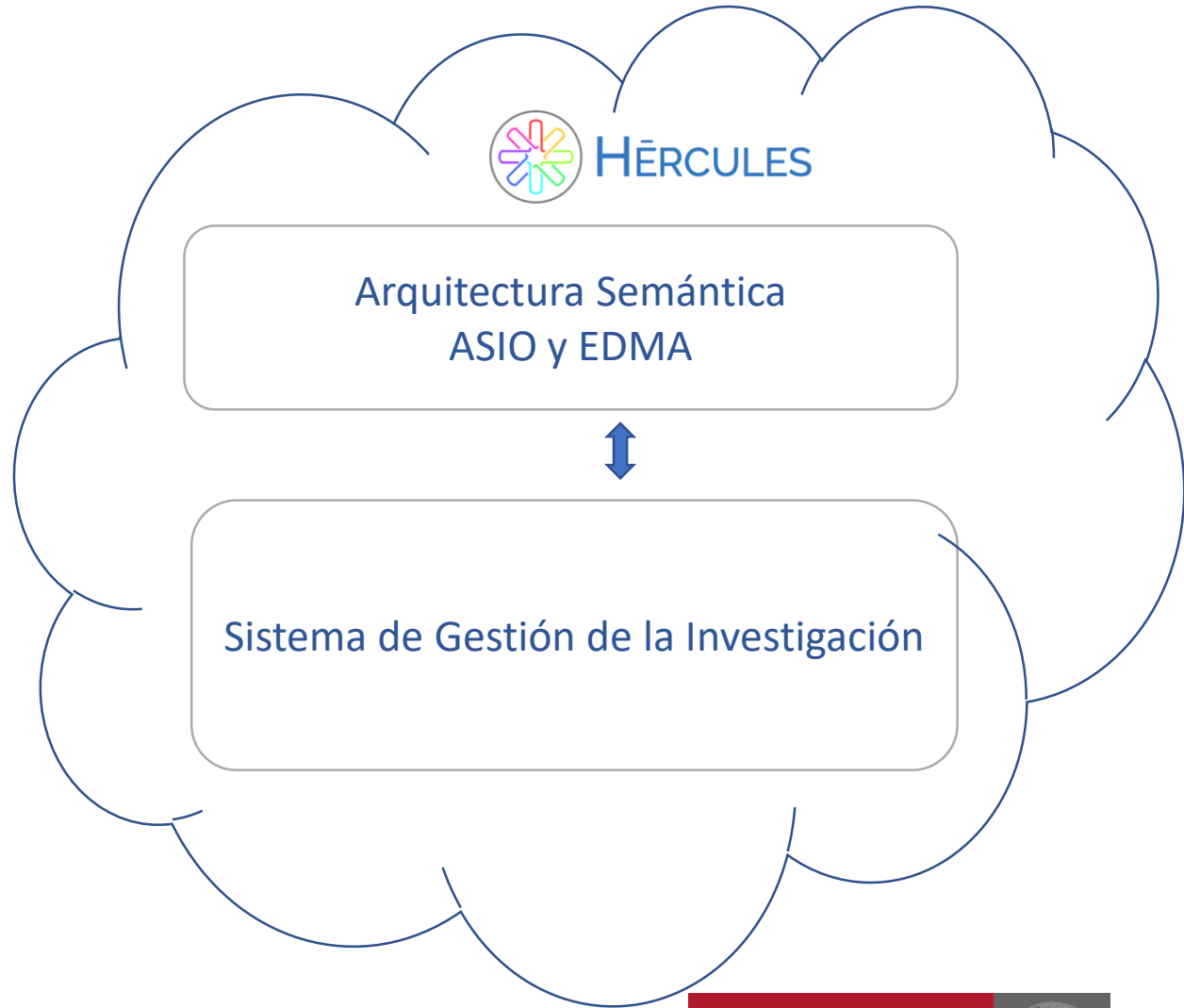
## Proyecto Hércules

El objetivo de Hércules es crear un Sistema de Gestión de la Investigación basado en **datos abiertos semánticos** que ofrezca una visión global de los datos de investigación **para mejorar la gestión, análisis y posibles sinergias entre universidades y gran público.**

El proyecto Hércules se compone de los siguientes proyectos:

- Arquitectura semántica (ASIO y EDMA)
- Sistema de gestión de la investigación (SGI)

## Introducción al Proyecto



**Enriquecer, aprovechar y explotar los datos producidos por la Arquitectura Semántica y la Infraestructura Ontológica de la Información del Sistema Universitario Español**

## Retos relacionados



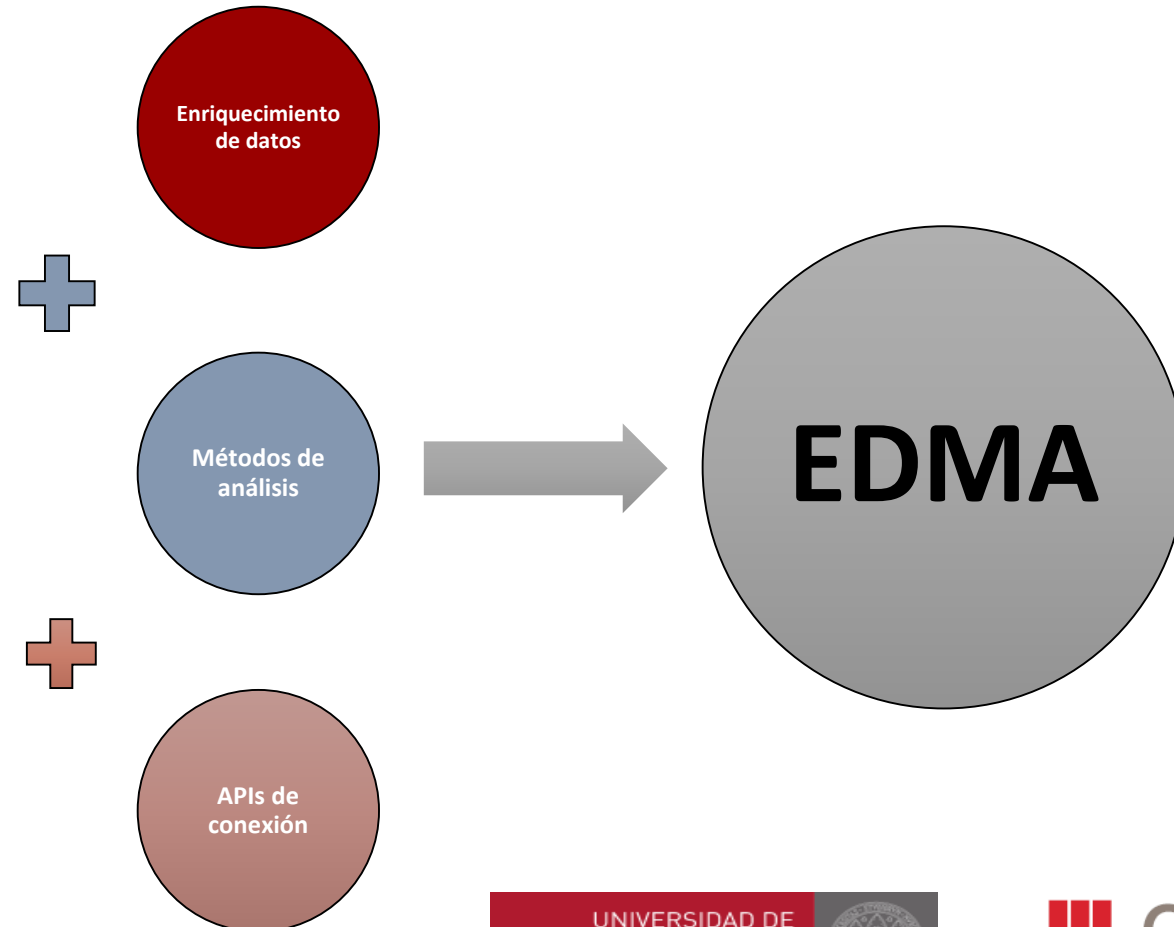
**Open/ Linked Data**

**Tecnologías del Lenguaje**

**Principios FAIR**

## El proyecto EDMA

EDMA



**Lote 1. Enriquecimiento de datos:**  
Facilitar el enriquecimiento mediante el uso de fuentes de información disponibles en internet y que son de uso habitual.

**Lote 2. Métodos de Análisis:**  
Permitir la explotación y la analítica de los datos con propósitos de inclusión y participación de los agentes implicados.

**Las aplicaciones desarrolladas deben proveer métodos vía API** que permitan establecer una conexión directa con el SGI y demás funcionalidades de Hércules.

# Índice

- 01 Introducción
- 02 Tecnologías propuestas**
- 03 Enriquecimiento de datos:  
Alcance técnico
- 04 Métodos de análisis:  
Alcance técnico
- 05 Componentes I+D
- 06 Cooperación entre proyectos





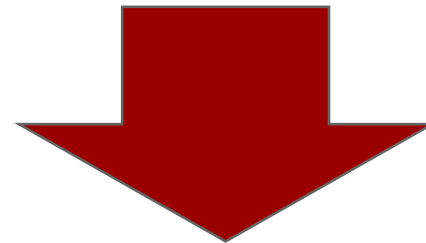
## Tecnologías Propuestas

Open Data

Web Semántica

Principios FAIR

Tecnologías del Lenguaje



**Modelo de software común y libre para facilitar la compartición de datos entre Universidades.  
Inclusión de capacidades semánticas al Sistema de Gestión de la Investigación.**

Es de gran interés la reutilización de desarrollos financiados por el **Plan de Impulso de las Tecnologías del Lenguaje** del Gobierno de España, así como estar alineados con iniciativas internacionales como la **European Open Science Cloud**.

# Índice

- 01 Introducción
- 02 Tecnologías propuestas
- 03 Enriquecimiento de datos:  
Alcance técnico**
- 04 Métodos de análisis:  
Alcance técnico
- 05 Componentes I+D
- 06 Cooperación entre proyectos



## Enriquecimiento de datos: Alcance técnico

### GESTIÓN DE FAIR RESEARCH OBJECTS (RO)

#### Módulo 1: Gestión de Research Objects

El investigador tiene sus 'Objetos de Investigación' en varios repositorios, y vía APIs se podrán reclamar los RO de las distintas aplicaciones, registrando la trazabilidad de la procedencia de RO.

#### Módulo 2: Gestión integrada de RO

Los RO son agregaciones de varios RO, y existen RO de muy diferentes tipos, existiendo matices propios de cada RO. Se deben poder manejar integradamente varios RO.

#### Módulo 3: Valorización de RO

Establecer un valor para todos los RO producidos en una investigación, con capacidad de determinar su procedencia.

### Necesidad de desarrollar un esquema de metadatos para RO

# Enriquecimiento de datos: Alcance técnico

## GESTIÓN DE FAIR RESEARCH OBJECTS (RO) – TIPOS DE RO

### Módulo 4: Referencias bibliográficas

Se deben poder resolver las referencias en las publicaciones, armando al investigador con un mapa conceptual donde se pueda ver la relación entre sus RO y los que cita.

### Módulo 6: Módulo de código

El código generado por un investigador debe ser cuantificable dentro del conjunto de su producción, y este módulo debe permitir dicha cuantificación.

### Módulo 5: Protocolos experimentales

Disponer de una relación explícita entre el protocolo experimental y las publicaciones relacionadas. Estándares que apoyen el registro de protocolos experimentales.

### Módulo 7: Anotaciones manuales

Se debe permitir la anotación de un documento, que posibilite generar una referencia, ampliar o explicar un elemento narrativo y generar una discusión sobre el elemento narrativo.

# Enriquecimiento de datos: Alcance técnico

## PROCESAMIENTO Y ANÁLISIS DE RO

### Módulo 8: Extracción de tópicos

Con el objetivo de poder mejorar los tópicos de experiencia derivados de las publicaciones de un investigador, se extraerán gracias a un algoritmo inteligente los temas directamente de las publicaciones. El algoritmo que lo sugerirá aprenderá de cambios realizados por el investigador en el establecimiento de los tópicos.

### Módulo 9: Similitud semántica

Se espera que se puedan indicar al usuario aquellas publicaciones que son similares en el contenido, comunicando de manera clara cómo y por qué los documentos encontrados son similares a uno dado por el usuario.

# Enriquecimiento de datos: Alcance técnico

## PERFIL DEL INVESTIGADOR (Researcher Dashboard)

### Módulo 10: Módulo Researcher Dashboard

El usuario debe ser capaz de administrar los datos de su perfil, incluyendo su producción sobre diferentes aplicaciones web (enriqueciéndose). Necesidades que se deben satisfacer:

- Ingresar y editar los datos del CV, cargando vía API los datos directamente de fuentes externas y del SGI.
- Obtener el listado de los trabajos que ha dirigido o codirigido, sea cual sea la naturaleza del mismo.
- Obtener el listado de congresos y eventos de divulgación científica en los que haya participado, indicando el rol que se haya tenido.
- Obtener el listado de patentes, diseños industriales, que haya registrado como titular o cotitular.
- Obtener el listado de proyectos en los que se haya participado, incluyendo el rol desempeñado (investigador principal por ejemplo).
- Obtener el listado de la producción científica
- Obtener el listado de startup o spin-off que se hayan fundado o cofundado
- Obtener los indicadores de producción científica
- Visualizar la trayectoria según la línea de tiempo, siendo parametrizable.
- Saber si el usuario es apto para solicitar una evaluación relativa al nuevo sexenio de transferencia de conocimiento
- Introducir Ofertas tecnológicas dirigidas a empresas

## Enriquecimiento de datos: Alcance técnico

### SERVICIOS CENTRALES

#### Módulo 11: Single Sign On

Métodos que permitan a los usuarios registrados acceder a las aplicaciones de Hércules con un único usuario y clave.

#### Módulo 12: Single login Entry Point

En el caso que determinadas aplicaciones de Hércules tengan restricciones de seguridad para algunos usuarios, será gestionado por la interfaz, creando un registro de identidades para el usuario.

#### Módulo 13: Administración de usuarios

Desarrollo de un módulo que permita crear, editar, borrar y suspender usuarios. Definición de roles, permisos para cada tipo de usuario de Hércules, gestionando la seguridad asociada a cada perfil de usuario.

# Índice

- 01 **Introducción**
- 02 **Tecnologías propuestas**
- 03 **Enriquecimiento de datos:  
Alcance técnico**
- 04 **Métodos de análisis:  
Alcance técnico**
- 05 **Componentes I+D**
- 06 **Cooperación entre proyectos**





# Métodos de análisis: Alcance técnico

## BUSQUEDA DE INVESTIGADORES

### Módulo 1: Research Synergy Finder

El usuario debe poder realizar búsquedas de investigadores y producción científica con la finalidad de facilitar la detección de posibles nichos de colaboración y alianzas estratégicas, mediante una interfaz capaz de filtrar de acuerdo a distintos criterios los intereses.

- Obtención de listados de centros de investigación que trabajan en una disciplina o área específica
- Listado de investigadores de un área o disciplina específica
- Top 10 de los investigadores en función de número de citas, etc
- Top 10 de centros de investigación en función de sellos de calidad
- Listado de centros de investigación que hayan realizado proyectos por ejemplo H2020, o de Plan Estatal
- Obtener un listado de producción científica en un determinado rango de fechas de un centro de investigación.
- Obtener una visualización en la que se recoja la distribución de la producción científica española.
- Comparar comunidades autónomas, universidades, grupos de investigación en un área o disciplina.
- Obtener un listado de patentes, diseños industriales, en un área.
- Obtener un listado de proyectos adjudicados o desarrollados de un centro, de un área, en un determinado año de búsqueda, obteniendo detalles.
- Siendo un usuario no investigador, se debe poder conocer el tamaño y experiencia de un área de investigación a distinta escala
- Conocer el porcentaje de participación de un centro de investigación en proyectos nacionales, europeos etc.

# Métodos de análisis: Alcance técnico

## ANÁLISIS DE PROYECTOS DE INVESTIGACIÓN

### Módulo 2: Gestión y análisis de proyectos

El usuario será capaz de **analizar los resultados** de los proyectos de investigación que se están llevando a cabo en un centro/estructura de la organización.

- Posibilidad de insertar/modificar los datos relacionados con los proyectos de investigación. Se tendrá acceso a esta información según su nivel de acceso.
- Visualización que permita explorar la información de cada proyecto por filtros.
- Identificación de proyectos por temática y objetivos científicos similares.
- Visualización que permita analizar la evolución de un investigador, conjunto de investigadores o líneas de investigación a través de los resultados de los proyectos realizados.

## Métodos de análisis: Alcance técnico

### ANÁLISIS DE PROYECTOS DE INVESTIGACIÓN

#### Módulo 3: Análisis de indicadores de la investigación

Se proveerá al usuario de un mínimo de 20 indicadores de investigación e innovación que le permitan medir la investigación nacional a partir de la información disponible en la red HÉRCULES y en fuentes externas.

- Posibilidad de obtener el listado de indicadores con su respectivo valor y unidad de medida, calculados en un periodo de tiempo para toda la Universidad o cada centro.
- **Visualización de la evolución** de indicadores según la línea del tiempo.
- **Predicción de la evolución** de indicadores a partir de la serie temporal de sus valores y de las variables existentes en el sistema.
- **Detección de tendencias** en áreas y líneas de investigación a partir de los datos disponibles en HÉRCULES.
- Posibilidad de cuantificar la contribución de cada investigador, línea de investigación, área de conocimiento, centro de investigación, comunidad autónoma,... a cada indicador.
- Posibilidad de hacer clusters de líneas y áreas de investigación, investigadores, centros,... usando como criterio de clasificación uno o varios indicadores de productividad.
- Conocimiento de que líneas y áreas de investigación, investigadores, centros,... presentan desviaciones significativas respecto a la media de los indicadores de productividad científica.
- Conocimiento del **perfil** y evolución de la **relación** de investigación y transferencia de una **empresa** con un conjunto de **centros de investigación** en un periodo de tiempo.

# Índice

- 01 Introducción
- 02 Tecnologías propuestas
- 03 Enriquecimiento de datos:  
Alcance técnico
- 04 Métodos de análisis:  
Alcance técnico
- 05 **Componentes I+D**
- 06 Cooperación entre proyectos



## Componentes I+D

COMPONENTE	Lote
Modelos de metadatos para descripción y anotación de FAIR Research Objects	1
Librería de enriquecimiento de FAIR Research Objects	1
Librería de extracción de tópicos	1
Librería de similitud semántica	1
Librería de métodos de análisis, visualización y predicción	2
Librería de detección de sinergias	2

# Índice

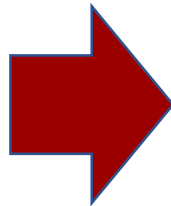
- 01 Introducción
- 02 Tecnologías propuestas
- 03 Enriquecimiento de datos:  
Alcance técnico
- 04 Métodos de análisis:  
Alcance técnico
- 05 Componentes I+D
- 06 **Cooperación entre proyectos**

## Cooperación entre proyectos

Los proyectos del presente pliego:

- Dependen de la Infraestructura Ontológica y la Arquitectura Semántica de Datos.
- Están relacionado con el Sistema de Gestión de la Investigación.
- Están estructurados en dos lotes.

Es por eso que entra dentro del alcance de los adjudicatarios de todos los pliegos **interactuar con los desarrolladores de los demás proyectos para:**



Resolver dudas.

Modificar los artefactos generados, siempre que tal modificación esté justificada y mantenga el nivel de excelencia técnica.

Asegurar la interoperabilidad y coherencia de las soluciones desarrolladas en todos los pliegos y lotes.



FONDO EUROPEO DE DESARROLLO REGIONAL  
(FEDER)  
*Una manera de hacer Europa*



Unión Europea

