

Thesaurus Technology

MARAFd

Open Source – Código Abierto

MEMORIA TÉCNICA ELABORADA PARA PORTAL **HISPANA.ES**

Gestión inteligente de fondos documentales digitales

[Site oficial del proyecto](#)

1. El concepto

MARAFd puede definirse como un gestor de fondos digitales de segundo nivel que se sitúa entre el usuario y el servicio básico de hosting web, sin necesidad de sistemas intermedios de software ni la hiperestructuración propia de los sistemas de bases de dato documentales, priorizando la preservación documental así como una accesibilidad en línea permanente e inmediata al cien por cien del contenido indizado y respetando los principios de procedencia y de orden como pilares fundamentales de la archivística.

Técnicamente, consiste en un ecosistema de tecnologías de código abierto diseñado para gestionar fondos documentales digitales mediante índice-tesauro automatizado y que aprovecha principalmente los recursos del sistema operativo para el indizado de archivos. Su arquitectura, basada en estándares abiertos como *PHP*, *Batch-script*, *FTP*, *JavaScript* y *HTML5*, elimina la dependencia de complejos y costosos sistemas de bases de datos tradicionales, priorizando siempre la preservación documental así como la accesibilidad, escalabilidad y, muy especialmente, la reducción máxima en cuanto a costes de implementación se refiere. El índice-tesauro, generado a partir de la propia estructura de carpetas y nombres de archivo, permite búsquedas instantáneas incluso en volúmenes de información del orden de varios TeraBytes y sin otro límite que el derivado del servicio de hospedaje contratado (*hosting-web*).

La transición digital ha exacerbado la necesidad de soluciones innovadoras para gestionar un crecimiento exponencial de datos que puede acabar colapsando a

muchas organizaciones. Por ello se exigen soluciones disruptivas capaces de afrontar el **crecimiento exponencial** de datos y documentos digitales. Pero, ¿podemos considerar esta sencilla tecnología como realmente disruptiva?

MARAFd permite a cualquier usuario, con un editor de texto, prácticamente sin conocimientos técnicos y con **apenas 10 Kilobytes de código abierto**, manejar grandes fondos documentales digitales (cientos o miles de *Gigabytes* de peso) accesibles en línea.

En este ámbito, la tecnología **MARAFd** constituye una herramienta multitecnológica de código abierto dirigida fundamentalmente a la gestión Inteligente de fondos documentales digitales y que basa su funcionamiento principal en un **índice-tesauro**. Este **tesauro**, como auténtica columna vertebral del fondo documental en línea, rompe literalmente los esquemas convencionales preestablecidos en cuanto a la estructuración de la información se refiere prescindiendo de sistemas externos de gestión de bases de datos y otros métodos más complejos de estructuración de la información empleados en la mayoría de fondos documentales digitales.

La solución **MARAFd** supone por tanto un **cambio de paradigma** en la gestión de archivos digitales y la gestión del conocimiento al superar la complejidad inherente a la **hiperestructuración** de datos que acaba dificultando el acceso a la información y disparando los costes estructurales de soporte.

La gestión de grandes fondos documentales de archivos digitales con un elevado grado de escalabilidad y portabilidad mediante **MARAFd** es posible. Repositorios de millones de archivos e incluso con varios *Terabytes* de información son perfectamente manejables con esta tecnología prácticamente sin requisitos especiales de *hardware* ni un soporte técnico más allá del derivado de las propias necesidades del *hosting web*.

2. Características técnicas

La solución **MARAFd** surge a partir de una integración estratégica de tecnologías abiertas y gratuitas como *PHP*, *Batch-Script*, *Bootstrap*, *JavaScript* y *HTML5*. **MARAFd** no es realmente un software al uso sino lo que podríamos denominar un ecosistema de código abierto.

MARAFd ofrece una forma segura para la gestión de repositorios con una escalabilidad prácticamente ilimitada y unos costes ínfimos en lo que a implementación se refiere.

Nos encontramos, en definitiva, ante una solución que prioriza la accesibilidad a la información transformando laboriosos procesos manuales en flujos de trabajo seguros y automatizados, manteniendo la importancia de la preservación y la integridad, el valor histórico de los documentos y la incalculable valía del tiempo de búsqueda.

3. Cómo está diseñado

En el desarrollo e implementación de **MARAFd** se utilizan las siguientes tecnologías/lenguajes:

- **PHP** : Adaptación de las rutinas de búsquedas y navegación (*backend*) por directorios (*finder.php, navega.php*)
- **BATCH-SCRIPT** : Generación local de tesoro y transferencia a servidor (*reindizar.cmd*)
- **SFTP** : Transferencia de archivos al servidor y sincronización de carpetas.
- **HTML5** : Diseño del *frontend* de acceso al fondo (*index.html*)
- **JAVASCRIPT** : Diseño del *frontend* de acceso al fondo (*index.html*)
- **BOOTSTRAP** : Diseño del *frontend* de acceso al fondo con compatibilidad multiplataforma
- **IA GENERATIVA** : Mecanismo de análisis y ataque a la información mediante modelos externos de inteligencia artificial entrenados a partir del contenido del fondo documental (*MyGPT, etc.*)

El **tesoro-índice es actualizado y generado de forma automatizada** mediante *Batch-Script* a partir de la estructura de carpetas y archivos locales, que a su vez se encuentra duplicada en el servidor mediante sincronización **SFTP**. De esta forma conservamos siempre un *mirror* (copia espejo) de todos los directorios y archivos del fondo sin necesidad de procesos complementarios de *backups*, implicando ello una **preservación y accesibilidad inmediata en línea** de cualquier documento indizado en el fondo.

El **tesoro** consiste en un archivo de texto plano que aprovecha el propio **nombre** y **path** (ruta completa) de cada archivo para codificar una cantidad importante de información relativa a éste, por ello, el nombre de todos los ficheros debe ser descriptivo y ajustarse a ciertos criterios formales preestablecidos por el

administrador, descartando por ejemplo el uso de espacios, tildes y/o símbolos que puedan comprometer la homogeneidad del tesoro y permitan mantener cierta estructura informativa acerca de algunas características importantes relativas al contenido del archivo, como puede ser el año de elaboración del documento, su tipología, título, autor, lugar, etc.

El fichero **índice-tesoro** se forma a partir del **path** y el **nombre** de cada uno de los ficheros indizados y con cada nueva incorporación al fondo, que puede realizarse de forma individual o en bloques ilimitados de ficheros, el tesoro se actualiza, regenera y transfiere automáticamente al servidor mediante **Batch-Script**.

Para mantener la integridad del mismo, debe respetarse un límite máximo por **url** de unos **2 mil caracteres**, longitud suficiente para codificar una parte importante de la información relativa a cada registro y que nos permite el manejo y acceso a los archivos de forma nativa sin necesidad de recurrir a sistemas de gestión de base de datos.

A continuación se muestra ejemplo del registro de un fichero ya indizado en el fondo y en el que se han codificado datos como el año de elaboración, título del documento, autor/es, idioma, etc.

```
/ebiblio_melilla_riff/PUBLICACIONES-REVISTAS-DIARIOS/1907_Tipo-documento_Titulo-  
subtitulo_del_documento_autor_lugar-es_referido_etc_ES_EN_FR_CA_IT_DE.pdf
```

A partir del propio **tesoro** y mediante dos sencillas **rutinas personalizables PHP** de código abierto, el sistema **MARAFd** crea los *hiperenlaces* a los documentos tanto en el sistema de **búsqueda rápida** (*finder.php*) como en la **navegación por directorios** (*navega.php*) de forma dinámica.

4. Qué necesito

Como administrador:

Prácticamente, la puesta en marcha de un fondo documental en línea por parte del administrador mediante la tecnología **MARAFd** solo viene a requerir la capacidad de almacenamiento necesaria, tanto en el equipo local (*local hosting*), como en el servidor (*hosting web*), para albergar la totalidad de archivos contenidos en el fondo documental.

Dicho servicio de **hosting web** deberá ofrecer conexión por **SFTP** y posibilidad de ejecutar en el servidor código **PHP** para las rutinas de **búsqueda rápida** (*finder.php*) y

de **navegación por directorios** (*navegar.php*), encargadas de rastrear el fichero **índice-tesauro** y generar de forma dinámica y en tiempo real los *hiperenlaces* a los documentos contenidos en el fondo.

La elaboración, adaptación y subida en línea de estas sencillas rutinas **PHP** necesarias para su puesta en marcha, así como la edición del código **HTML5** necesario para montar el *frontend* que dará acceso al fondo pueden realizarse a través de sencillas herramientas como el *Bloc de notas*, *VSCode*, *Notepad++*, *FileZilla*, etc.

Como usuario:

El **usuario/lector/investigador** que accede al fondo únicamente requiere de un dispositivo con conexión a internet y un navegador de internet. A partir de ahí, los requerimientos de *hardware/software* del cliente pueden variar en función del tipo de archivos que descarguemos/visualicemos así como el tipo de tareas a realizar sobre ellos.

5. Quién puede implementar **MARAFd**

MARAFd ofrece sobrada capacidad para vertebrar enormes repositorios en prácticamente cualquier ámbito del conocimiento y con una inversión de capital y recursos humanos prácticamente nula, permitiendo una sencilla implementación por parte de:

- **Archivistas y gestores de información digital.** Con un costo de *hardware* muy reducido **MARAFd** nos permite acceder a un potencial enorme y una escalabilidad total.
- **Desarrolladores de software** interesados en la gestión documental y que decidan apostar por la implementación de **MARAFd**.
- **Profesionales independientes del sector de las tecnologías** que buscan soluciones innovadoras y asequibles en un mundo cada vez más digitalizado y complejo.
- **Pequeñas y medianas empresas** con un problema real organizativo y de estructuración de archivos digitales acumulado a lo largo de años.
- **Entes públicos y/o privados** sin posibilidades de inversión en un sistema de gestión de archivos de alto coste y con necesidad urgente de **digitalización documental**.

- **Equipos científicos multidisciplinares de investigación** en cualquier rama del conocimiento con necesidades de análisis y escrutinio grupal y estructuración de grandes volúmenes de información documental digitalizada.

6. Por qué implementar **MARAFd**

Dadas sus características técnicas y los objetivos perseguidos, podemos afirmar que **MARAFd** ofrece notables ventajas frente a otros sistemas alternativos, especialmente en cuanto simplicidad técnica y costos se refiere, destacando entre otras:

- **PRESERVACIÓN INMEDIATA** del conocimiento y del patrimonio digital cultural y/o científico.
- **INCREMENTO NOTABLE DE LA PRODUCTIVIDAD** en los diferentes ámbitos de aplicación al reducirse drásticamente los tiempos de acceso a la información.
- **ACCESIBILIDAD** total en línea y escalabilidad ilimitada desde cualquier tipo de plataforma y/o sistema operativo.
- **VERSATILIDAD** y gestión flexible frente a cualquier tipo de contenido digital.
- **INTERFAZ ÁGIL E INTUITIVA** absolutamente customizable.
- **Tecnología basada en un índice-tesauro** actualizado de forma automatizada y que vertebra en tiempo real la totalidad de archivos del fondo ofreciendo un rendimiento inmejorable y tiempos de respuesta inmediatos incluso con volúmenes del orden de varios CIENTOS de *GIGABYTES* de información y decenas de miles de archivos.
- La implementación de un **SISTEMA EXPERTO ENTRENADO** del tipo **MyGPT®** entrenado a partir del contenido de los fondos dispara de forma exponencial el potencial de acceso al contenido en investigaciones avanzadas y/o consultas rápidas por parte de cualquier usuario.
- **OPTIMIZACIÓN Y SEGURIDAD** en los flujos de trabajos de equipos multidisciplinares posibilitando compartir fondos en línea de forma transparente y sencilla.
- **REDUCCIÓN NOTABLE DE COSTES** de implementación, mantenimiento y soporte.

- **COMPATIBILIDAD TOTAL Y COEXISTENCIA CON OTROS SISTEMAS DE GESTIÓN DE FONDOS**, permitiendo mantener ambas soluciones funcionales de forma simultánea.

7. Cómo puedo probar el potencial de **MARAFd**

Actualmente existen dos casos de aplicación de la tecnología **MARAFd** con resultados francamente satisfactorios como herramienta de apoyo en el ámbito de la investigación historiográfica y la preservación:

- **Fondo Documental Digital MELILLA-RIF** : Con más de 9 mil ficheros multimedia ya indizados y un volumen total de datos que supera los 100 GB de información digital completamente accesible en línea, este fondo temático indizado en el **Directorio de Colecciones** del portal **HISPANA** se ha convertido en apenas dos años en un referente para la investigación y la divulgación sobre la historia y la cultura de Melilla y el Rif.
- **Fondo Documental Digital sobre PROTOHISTORIA E HISTORIA ANTIGUA DEL MEDITERRANEO OCCIDENTAL** : este fondo de **ACCESO RESTRINGIDO** ofrece un vastísimo repertorio de documentos relativos a las primeras civilizaciones y culturas del Mediterráneo Occidental, pretendiendo facilitar la investigación académica y la consulta especializada.

MARAFd ha permitido a estos fondos digitales ofrecer un acceso más rápido y profundo a los contenidos garantizando además una experiencia de usuario adaptada a las demandas de cualquier investigación, estudiantes y/o cualquier perfil de usuario. Con un enfoque sencillo y escalable y un coste de implementación difícilmente superable, **MARAFd** puede convertirse en una herramienta imprescindible para todos aquellos proyectos que requieren una gestión rápida, sencilla y eficaz de archivos digitales y un volumen del orden de miles e incluso millones de entradas.

[By Rafael Lomena Varo]

Contact/admin/author: info@calentamientoglobalacelerado.net

1993 ©© 2025

Creative Commons Attribution-NonCommercial-ShareAlike 4.0

18/02/2025 - [2502060809787](#)