

# Manual de uso de MMAD (Migración de Metadatos y Archivos Digitales)



v1.0 29/07/2019

## **Autores**

**Alexis Febre - OCA**  
**Tomás Cohen Arazi - PSI**

## **Equipo Oficina de Conocimiento Abierto**

**Alejandra Nardi**  
**Emilio Di Domenico**  
**Lucrecia García**  
**Mario Pizzi**

## **Equipo Prosecretaría de Informática**

**Lorena Orcellet**  
**Carlos Iván Scándolo**  
**Nicolás Salvai**



Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial 4.0 Internacional](https://creativecommons.org/licenses/by-nc/4.0/).

**Contacto:** Alejandra M. Nardi: [amnardi@unc.edu.ar](mailto:amnardi@unc.edu.ar)

# Índice

<b>Introducción</b>	<b>4</b>
<b>Descripción</b>	<b>6</b>
<b>Funcionamiento</b>	<b>6</b>
<b>Requisitos de Instalación</b>	<b>6</b>
<b>Estructura</b>	<b>7</b>
<b>Proceso de Migración</b>	<b>8</b>
<b>Modificaciones futuras</b>	<b>13</b>
<b>Conclusiones</b>	<b>13</b>
<b>Bibliografía</b>	<b>13</b>
<b>Anexo</b>	<b>14</b>
Unidades Académicas	14
Metadatos	14

## Introducción

MMAD (Migración de Metadatos y Archivos Digitales) es un programa desarrollado por la Oficina de Conocimiento Abierto (OCA) y la Prosecretaría de Informática, ambas pertenecientes a la Universidad Nacional de Córdoba (UNC). El programa facilita la migración de metadatos y archivos almacenados en el Módulo Memoria SIGEVA<sup>1</sup>-UNC a su Repositorio Digital Universitario (RDU).

Con esta acción la Universidad Nacional de Córdoba, avanza con el cumplimiento de los plazos de adecuación establecidos por la autoridad de aplicación de la Ley Nacional [Nº 26.899/2013](#) “Repositorios Digitales Institucionales de Acceso Abierto” y su Reglamento Operativo Resolución [753 - E/2016](#).

El módulo Memoria SIGEVA fue adaptado a las necesidades de nuestra Universidad. Su objetivo es mostrar una síntesis anual de las actividades y publicaciones de los docentes investigadores de cada unidad académica. Está integrado por los siguientes ítems: publicaciones (artículos, partes de libros, libros, trabajos presentados en eventos de CyT, tesis); desarrollos tecnológicos, organizacionales y socio-comunitarios; servicios; redes, gestión editorial y eventos (trabajos en eventos CyT no publicados); trayectoria - formación de recursos humanos y financiamiento CyT. De esta manera, “Memoria” resulta una síntesis anual de las actividades científicas, tecnológicas y artísticas acontecidas en un año en concreto.

Más adelante se muestra una imagen de la pantalla principal de una Memoria SIGEVA correspondiente a una Unidad Académica de la UNC del año 2015, donde se destacan aquellos apartados objetos de migración.

La tarea comienza con una unificación de registros en el Módulo MEMORIA SIGEVA. Estas labores las realizan personas con el rol “colaborador MEMORIA SIGEVA”. Dicho rol, en la mayoría de los casos recae en personal de las Bibliotecas de cada Unidad Académica que, a su vez, operan como *Nodos OCA* (labores de permanente cooperación y coordinación de acciones junto a la Oficina de Conocimiento Abierto).

Una vez unificados los registros, la persona con el rol “Director de MEMORIA” (que recae generalmente sobre quien fuese titular de la Secretaría de Ciencia y Técnica de la Unidad Académica en cuestión) realiza el “envío de la presentación”, que equivale a señalar que las tareas de unificación están finalizadas.

De esta manera, el personal informático de la OCA comienza a trabajar con la herramienta MMAD. Una vez importados los datos al Repositorio Institucional, la Oficina y

---

<sup>1</sup> SIGEVA - Sistema Integral de Gestión y Evaluación. Dirección de Informática. Gerencia de Organización y Sistemas - Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). MÓDULO DE MEMORIA. Versión 1. Octubre de 2015. Disponible en:

[http://sigeva.conicet.gov.ar/wp-content/uploads/2015/10/SIGEVA-Modulo-Memoria\\_v1.pdf](http://sigeva.conicet.gov.ar/wp-content/uploads/2015/10/SIGEVA-Modulo-Memoria_v1.pdf).

los Nodos OCA proceden a la curatoría de los datos; esto implica la optimización de los datos bibliográficos. En esta etapa se atienden particularmente las políticas editoriales de cada una de las revistas donde han sido publicados originalmente los trabajos, cuidando particularmente los periodos de embargo.

MEMORIA		MEMORIA 2015					
Información Institucional							
Información				Estado			
FAC.DE							
Publicaciones	TOTALES		POSIBLES SEMEJ.		IGNORADAS		
	Total	%	Total	%	Total	%	
Artículos	159	100.0	0	0.0	0	0.0	
- Partes de libro	60	100.0	0	0.0	0	0.0	
- Libros	23	100.0	0	0.0	0	0.0	
- Trabajos en eventos C-T publicados	184	100.0	0	0.0	0	0.0	
- Tesis	31	100.0	0	0.0	0	0.0	
- Publicaciones en revistas	38	100.0	0	0.0	0	0.0	
- Revistas científicas	2	100.0	0	0.0	0	0.0	
Desarrollos tecnológicos, organizacionales y socio-comunitarios	TOTALES		POSIBLES SEMEJ.		IGNORADAS		
	Total	%	Total	%	Total	%	
- Desarrollo de tecnologías, procesos tecnológicos y sistemas de gestión	3	100.0	0	0.0	0	0.0	
- Desarrollo tecnológico organizacional	0	0.0	0	0.0	0	0.0	
- Desarrollo tecnológico de gestión empresarial	0	0.0	0	0.0	0	0.0	
- Desarrollo tecnológico de gestión pública	0	0.0	0	0.0	0	0.0	
Servicios	TOTALES		POSIBLES SEMEJ.		IGNORADAS		
	Total	%	Total	%	Total	%	
- Servicios	5	100.0	0	0.0	0	0.0	
Redes, gestión editorial y eventos	TOTALES		POSIBLES SEMEJ.		IGNORADAS		
	Total	%	Total	%	Total	%	
- Trabajos en eventos C-T no publicados	267	100.0	0	0.0	0	0.0	
Trayectoria	TOTALES		POSIBLES SEMEJ.		IGNORADAS		
	Total	%	Total	%	Total	%	
- Asociación de investigadores	125	100.0	0	0.0	0	0.0	
- Asociación de investigadores científicos	1016	100.0	0	0.0	0	0.0	
- Asociación de investigadores tecnológicos	134	100.0	0	0.0	0	0.0	
- Asociación de investigadores científicos	316	100.0	0	0.0	0	0.0	
- Asociación de investigadores científicos de la salud	23	100.0	0	0.0	0	0.0	
- Comunicación pública de la ciencia y la tecnología	150	100.0	0	0.0	0	0.0	
- Eventos científicos	16	100.0	0	0.0	0	0.0	
- Eventos científicos organizados por investigadores	133	100.0	0	0.0	0	0.0	
- Eventos científicos organizados por investigadores	16	100.0	0	0.0	0	0.0	
- Eventos científicos organizados por investigadores	76	100.0	0	0.0	0	0.0	
Financiamiento CyT	TOTALES		POSIBLES SEMEJ.		IGNORADAS		
	Total	%	Total	%	Total	%	
- Becas de posgrado	314	100.0	0	0.0	0	0.0	
- Becas de posgrado otorgadas a investigadores	5	100.0	0	0.0	0	0.0	
- Becas de posgrado otorgadas a investigadores	0	0.0	0	0.0	0	0.0	
- Becas para investigadores	7	100.0	0	0.0	0	0.0	
- Becas para investigadores de posgrado	6	100.0	0	0.0	0	0.0	
PRESENTACION. Fecha Límite:				(*) Enviar Presentación			

## Descripción

La herramienta MMAD está desarrollada en lenguaje de programación Perl. Está conformado por un menú principal y distintos módulos cuyas funcionalidades se describen más adelante.

## Funcionamiento

En términos generales, la herramienta se conecta a la base de datos de Memoria SIGEVA, extrae los datos de los docentes investigadores y sus producciones y luego son codificados como Dublin Core. Cada resultado es almacenado en directorios bien definidos. Para que los registros sean importados desde el Repositorio Digital Universitario, es necesario comprimirlos. En esta primera versión del programa, la operación de compresión se realiza con programas estándares.

## Requisitos de Instalación

Para el funcionamiento la herramienta es necesario instalar:

- [Perl](#) >= v5.26.1
- [MySQL](#) >= v5.7.26

Más información para crear un usuario y base de datos.

Instalar MySQL: <https://dev.mysql.com/downloads/mysql>

Crear usuario: <https://dev.mysql.com/doc/refman/5.7/en/create-user.html>

Crear base de datos: <https://dev.mysql.com/doc/refman/5.7/en/creating-database.html>

# Estructura

El programa está organizado como se muestra en la imagen:

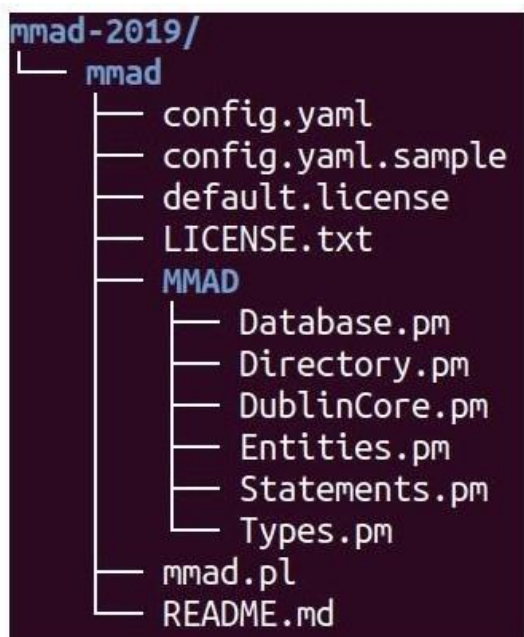


Figura 1 – Estructura de directorios de MMAD

- **config.yaml.sample:** Es un modelo de archivo de configuración. Se debe realizar una copia del mismo (Ej.: **config.yaml**) y registrar configuraciones como la conexión a la base de datos o la creación del directorio de salida de los registros migrados.
- **default.license:** Es la **Licencia de Distribución No Exclusiva** de la Universidad. Esta licencia es incluida en un archivo de contenidos que se genera al realizar el proceso de migración de metadatos y archivos digitales. Cada universidad deberá incluir su propia licencia.
- **LICENSE.txt:** Es una licencia [AGPL](#).
- **MMAD:** Es el directorio donde residen los módulos del proyecto. Estos se describen a continuación:
  - ❑ **Database.pm:** Se encarga de establecer la conexión a la base de datos.
  - ❑ **Directory.pm:** Se crean los directorios de los registros migrados y se almacenan en un directorio de salida establecido en el archivo de configuración. El directorio creado tiene la siguiente forma:  
**[Unidad Académica] / [Tipo de Material] / [Año de Publicación]**  
Ejemplo: **Odontología/Artículo/2013**

- ❑ **DublinCore.pm:** En este módulo se codifican los registros obtenidos de la base de datos en un esquema de [Dublin Core](#). Es importante definir los metadatos dentro del repositorio digital, preferentemente [DSpace](#).
  - ❑ **Entities.pm:** Este módulo contiene las unidades u entidades académicas, cuyos valores corresponden a Memoria Sigeva UNC, por lo tanto cada universidad deberá adecuarlos a sus particularidades.
  - ❑ **Statements.pm:** Contiene las sentencias SQL que se requieren para la obtención de los registros. Están basados en un entorno de test.
  - ❑ **Types.pm:** En este módulo se realiza la consulta de los registros de acuerdo al tipo de sentencia elegida. En este proyecto se estableció cinco tipos de consultas. Artículos, Libros, Capítulos de Libros, Congresos y Tesis.
- **mmad.pl:** Es el archivo principal, desde el cuál se ejecuta el programa. Es necesario pasarle tres argumentos. Éstos son: año de publicación, unidad académica y tipo de material.
  - **README.md:** Describe el proyecto.

## Proceso de Migración

Para el proceso de migración se debe ejecutar el siguiente comando:

***\$perl [mmad-src]/mmad.pl - - date [arg1] - - unit [arg2] - - type [arg3]***

Donde:

- **date:** Fecha de publicación (Ej.: 2013 en un rango de años de 2000 a 2018. Estos pueden ser modificados)
- **unit:** Unidad académica (Ej.: 187 corresponde a Facultad de Odontología. Ver anexo)
- **type:** Tipo de material:
  - 1 → Artículos científicos
  - 2 → Libros
  - 3 → Congresos
  - 4 → Capítulos de Libros
  - 5 → Tesis

Como resultado se obtiene un directorio de salida cuyo contenido es una unidad académica con sus registros catalogados por tipo y año.

Para explicar este proceso se muestra un ejemplo:

1. Ejecutamos:

***perl [mmad-src]/mmad.pl - - date 2013 - - unit 187 - - type 1***



2. Nos dirigimos al directorio recién creado:

**cd Odontología/Artículo/2013**

Se observa la siguiente salida:

```
Odontologia
├── Artículo
│   └── 2013
│       ├── item_114410
│       │   ├── contents
│       │   ├── dublin_core.xml
│       │   ├── license.txt
│       │   └── Senn_et_al_2011.rar
│       ├── item_126367
│       │   ├── contents
│       │   ├── dublin_core.xml
│       │   ├── license.txt
│       │   └── Suitability_of_ART_for_ACTA_revised_manuscript.doc
│       └── item_127346
│           ├── contents
│           ├── dublin_core.xml
│           ├── license.txt
│           └── Malignancy_Risk_Models_for_Oral_Lesions.pdf
```

**Figura 2 – Estructura de directorios**

Los ítems que se encuentran dentro del directorio **año**, corresponden a los registros extraídos de Memoria Sigeva. Cada uno de ellos contiene cuatro archivos, los cuales se describen a continuación:

- **contents:** Es un índice de contenidos, en el cual se muestra cada uno de los archivos que contiene el directorio.
- **dublin\_core.xml:** Este archivo contiene un conjunto de metadatos que describen el registro. El esquema que utiliza es Dublin Core.
- **license.txt:** Es una copia de la **Licencia de Distribución No Exclusiva** (default.license).
- **\*.pdf:** Es el documento en extensión pdf. Es recomendable que los registros que se carguen en el repositorio tengan extensión **.pdf**.

Antes de realizar la importación, es importante comprobar la existencia de estos archivos. La falta de algunos de ellos genera error en la importación de los registros.

3. Comprimos el directorio de los registros, especialmente el directorio que corresponde al año. Esta función se realiza por fuera del programa, utilizando los programas estándares. Se muestra como ejemplo, una manera de realizar tal operación.

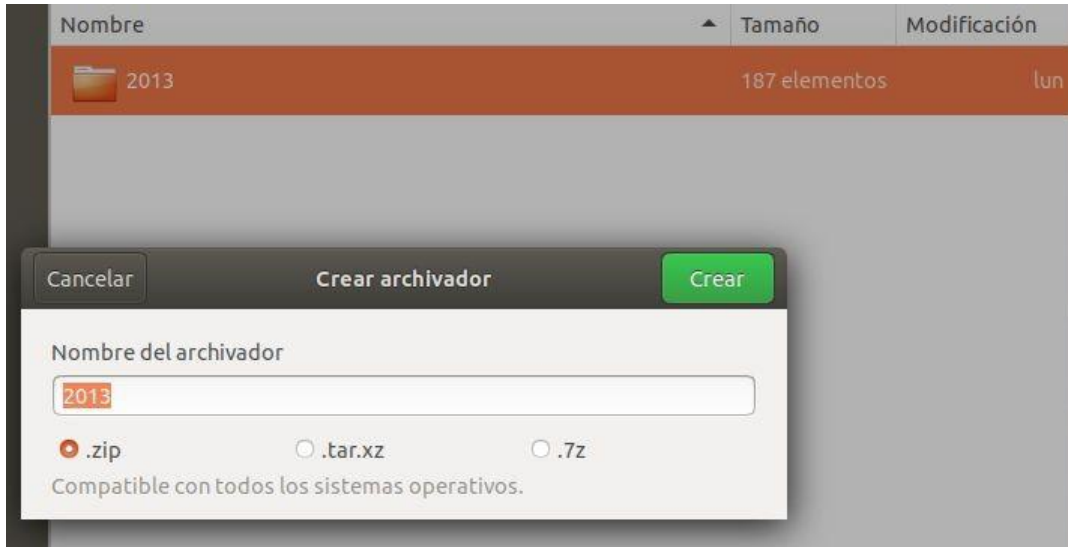


Figura 3 - Compresión del directorio año

4. Dentro del repositorio DSpace, creamos los metadatos que se utilizarán para importar los nuevos registros.



ID	Espacio de Nombres
1	<a href="http://dublincore.org/documents/dcmi-terms/">http://dublincore.org/documents/dcmi-terms/</a>

## Añadir un nuevo campo de metadato

Nombre del campo: dc .

journal

title

Nota de alcance:

Título del artículo

**Figura 4 - Edición de metadatos en DSpace**

Más información sobre edición de metadatos:

<https://wiki.duraspace.org/display/DSDOC5x/Batch+Metadata+Editing>

5. Creamos la colección de destino de los registros recién creados. Es importante que dicha colección tenga permisos de sólo lectura, debido a que el contenido debe ser curado antes de ser publicado.

CONTEXTO

Editar Comunidad
Exportar comunidad
Exportar metadatos
<b>Crear colección</b>
Crear subcomunidad

## Introduzca metadatos para una nueva colección de Memoria-SIGEVA

Nombre:

Artículo

Descripción breve::

## Editar colección: Artículos

Editar metadatos	<b>Asignar roles</b>	Origen del Contenido	Curar
------------------	----------------------	----------------------	-------

---

Administradores	ninguno	Crear...
Los administradores de la colección deciden quién puede enviar ítems a la colección, retirarlos, editar los metadatos (después del envío) y añadir (relacionar) ítems existentes de otras colecciones a esta colección (dependiendo de la autorización de esta colección).		

---

Publicadores	ninguno	Crear...
Usuarios y grupos que tienen permiso de envío de ítems nuevos a esta colección		

---

<b>Acceso por defecto de lectura</b>	El permiso por defecto de lectura de los ítems y archivos está actualmente asignado a Anónimo.	Restringir...
Usuarios y grupos que pueden leer los nuevos ítems enviados a esta colección. Los cambios en este rol no son retroactivos y los antiguos ítems permanecerán visibles para aquellos que hayan tenido acceso en el momento de su incorporación.		

Figura 5 - Crear una colección y otorgar permisos de sólo lectura.

### 6. Realizamos la importación de los registros.

<b>Administración de contenidos</b>
Ítems
Ítems retirados
Ítems privados
Importar metadatos
<b>Importación por lotes (ZIP)</b>

## Importar carga Batch (ZIP)

Seleccione un colección..

Colección:

Test > Sigeva-Memoria > Tesis

Seleccione la colección a la que desea enviar un ítem.

Seleccionar archivo 2013.zip

**Subir un ZIP SimpleArchiveFormat**

Figura 6 – Selección de colección de destino y registro

Seleccionada la colección destino y el correspondiente registro se oprime **Subir un ZIP SimpleArchiveFormat**. Si no surge algún inconveniente, DSpace devuelve el siguiente mensaje:



Figura 7 - Carga exitosa de los registros

## Modificaciones futuras

- Crear una interfaz gráfica para el uso de personal bibliotecario.
- Crear tareas programadas para la migración automática.
- Agregar en DSpace una funcionalidad de migración automática.

## Conclusiones

Debido a que este módulo es una primera versión, es probable que se produzcan errores en algunos pasajes. Es importante aclarar que este módulo puede ser modificado en el tiempo para obtener una versión estable y responder a las necesidades que requieran las instituciones.

## Bibliografía

- [1] Perl Tutorial: <https://www.perltutorial.org/>  
[2] Dublin Core: <http://dublincore.org/>  
[3] MySQL: <https://dev.mysql.com/doc/refman/5.7/en/>  
[4] DSpace 5: <https://wiki.duraspace.org/display/DSDOC5x/DSpace+5.x+Documentation>

## Anexo

### Unidades Académicas

Los valores que se muestran a continuación pueden variar según cómo se definan las unidades académicas en otras instituciones.

Memoria SIGEVA UNC:

- 185 = 'Rectorado'
- 186 = 'Medicina'
- 187 = 'Odontología'
- 188 = 'Ciencias-Económicas'
- 189 = 'Psicología'
- 190 = 'Derecho-Ciencias-Sociales'
- 191 = 'Ciencias-Agropecuarias'
- 192 = 'Ciencias-Exactas'
- 193 = 'FAMAF'
- 194 = 'Ciencias-Químicas'
- 195 = 'Filosofía-Humanidades'
- 196 = 'Observatorio'
- 197 = 'Arquitectura'
- 198 = 'Centro-Estudios-Avanzados'
- 200 = 'Lenguas'
- 201 = 'Artes'
- 206 = 'Ciencias-Sociales'
- 207 = 'Comunicación'

### Metadatos

Filiación del autor	dc.description.fil
---------------------	--------------------

### Artículos

EISSN	dc.identifier.eissn
Volumen	dc.journal.volume
Tomo	dc.journal.tome
Número	dc.journal.number
Páginas	dc.journal.pagination

País de edición	<b>dc.journal.country</b>
Ciudad de edición	<b>dc.journal.city</b>
Revista	<b>dc.journal.title</b>

### **Capítulos de Libros**

Volumen	<b>dc.book.volume</b>
Tomo	<b>dc.book.tome</b>
Número	<b>dc.book.number</b>
Página inicial	<b>dc.book.firstpage</b>
Página final	<b>dc.book.lastpage</b>
País de edición	<b>dc.book.country</b>
Ciudad de edición	<b>dc.book.city</b>
Cantidad de páginas	<b>dc.book.pages</b>
Editorial	<b>dc.book.editorial</b>

### **Congresos**

Tipo de publicación	<b>dc.conference.publication</b>
Tipo de trabajo	<b>dc.conference.work</b>
Revista	<b>dc.conference.magazine</b>
País de edición	<b>dc.conference.country</b>
Ciudad de edición	<b>dc.conference.city</b>
Editorial	<b>dc.conference.editorial</b>
Evento	<b>dc.conference.event</b>
País de evento	<b>dc.conference.eventcountry</b>
Ciudad de evento	<b>dc.conference.eventcity</b>

Fecha de evento	<b>dc.conference.eventdate</b>
Institución organizadora	<b>dc.conference.institution</b>

### **Libros**

Cantidad de volúmenes	<b>dc.book.volumes</b>
Cantidad de páginas	<b>dc.book.pages</b>
País de edición	<b>dc.book.country</b>
Ciudad de edición	<b>dc.book.city</b>
Editorial	<b>dc.book.editorial</b>

### **Tesis**

***Sin cambios***