

**EGICyT**  
Especialización en  
Gestión de Información  
Científica y Tecnológica

**FaHCE**  
FACULTAD DE HUMANIDADES Y  
CIENCIAS DE LA EDUCACIÓN



UNIVERSIDAD  
NACIONAL  
DE LA PLATA

**Universidad Nacional de La Plata**

**Facultad de Humanidades y Ciencias de la Educación**

**Secretaría de Posgrado**

**Especialización en Gestión de Información Científica y Tecnológica**

**Herramienta para la incorporación de recursos del repositorio  
institucional CONICET Digital a Memoria Académica**

**TRABAJO FINAL INTEGRADOR**

**ESPECIALISTA EN GESTIÓN DE INFORMACIÓN CIENTÍFICA Y TECNOLÓGICA**

**Alumna: Patricia Genovés**

**Tutora: Prof. Claudia González**

**La Plata, mayo de 2024**

**TÍTULO.** Herramienta para la incorporación de recursos del repositorio institucional CONICET Digital a Memoria Académica

**TÍTULO EN INGLÉS.** Tool for incorporating resources from the Digital Institutional Repository of the CONICET into Memoria Académica

**RESUMEN.** Se presenta un trabajo final integrador consistente en el desarrollo de dos programas informáticos para la cosecha e incorporación de recursos del repositorio CONICET Digital a Memoria Académica, el repositorio institucional de la Facultad de Humanidades y Ciencias de la Educación de la Universidad Nacional de La Plata y del Instituto de Investigaciones en Humanidades y Ciencias Sociales (IdIHCS) de doble dependencia CONICET-UNLP. Los objetivos son evitar que los investigadores nucleados en el IdIHCS deban autoarchivar en Memoria Académica trabajos que ya han incorporado al Sistema Integral de Gestión y Evaluación (SIGEVA) de CONICET, y contribuir al propósito de Memoria Académica de reunir toda la producción científico-académica generada en el ámbito de la Facultad, lo cual dará lugar a la provisión de perfiles de autor más completos que fortalezcan iniciativas en curso, tales como las estadísticas a nivel autor, los informes estadísticos a nivel departamentos y centros de investigación, y el seguimiento personalizado de la producción almacenada en cumplimiento a la Ley nacional 26.899 de Repositorios digitales institucionales de acceso abierto.

**PALABRAS CLAVE.** Recursos digitales; Repositorios institucionales; Interoperabilidad; OAI-PMH; Memoria Académica; CONICET Digital.

**ABSTRACT.** This final integrating work consists of the development of two computer programs for the harvest and incorporation of resources from the Repositorio Institucional CONICET Digital (the Digital Institutional Repository of the CONICET) to Memoria Académica (the Institutional Repository of the Faculty of Humanities and Educational Sciences of the National University of La Plata and of the CONICET-UNLP dual dependency Research Institute in Humanities and Social Sciences -IdIHCS-). The objectives are: to prevent researchers affiliated with the IdIHCS from having to self-archive in Memoria Académica works that they have already incorporated into the Sistema Integral de Gestión y Evaluación –SIGEVA (the Integrated Management and Evaluation System of the CONICET), and to contribute to the Memoria Académica's purpose of bringing together all the scientific-academic production generated in the scope of the Faculty. This will lead to the provision of more complete author profiles that will strengthen ongoing initiatives, such as statistics at the author level, statistical reports at the level of departments and research centers, and personalized monitoring of stored production in compliance with the National Law 26,899 on Open Access Institutional Digital Repositories.

**KEYWORDS.** Digital resources; Institutional repositories; Interoperability; OAI-PMH; Memoria Académica; CONICET Digital.

## TABLA DE CONTENIDO

Introducción .....	5
Justificación .....	8
Marco teórico .....	10
Sistemas de información documental e interoperabilidad .....	10
Repositorios digitales abiertos y protocolo OAI-PMH .....	12
Repositorios institucionales .....	16
CONICET Digital .....	19
Memoria Académica .....	20
Materiales y método .....	23
Herramientas utilizadas .....	26
BASH .....	26
Shell-oaiharvester .....	27
xsltproc .....	28
CISIS .....	29
yaz-marcdump .....	30
YAD – Zenity .....	30
Desarrollo .....	32
Diagramas de flujo de datos .....	32
Descripción de CDMA_tecno .....	34
Descripción de CDMA_procesos .....	39
Limitaciones a considerar .....	42
Resultados obtenidos .....	44
Adaptaciones futuras .....	46
Bibliografía y fuentes .....	47
Anexo A. Datos y metadatos relevantes según formato de cosecha .....	51
Anexo B. Código fuente de CDMA_tecno .....	57
Anexo C. Código fuente de CDMA_procesos .....	62
Anexo D. Manuales de uso .....	66
Anexo E. Contenido del conjunto de datos provisto .....	76

## ÍNDICE DE TABLAS Y FIGURAS

### Tablas

Tabla 1. Lista de formatos provistos para la cosecha por el repositorio CONICET Digital .....	25
Tabla 2. Definición de la base de datos <i>conicet</i> , conteniendo los datos cosechados .....	36
Tabla 3. Distribución de los recursos cosechados de CONICET Digital según tipo de documento y de acceso .....	44
Tabla 4. Distribución de los recursos cosechados de CONICET Digital según tipo de correspondencia con recursos de Memoria Académica .....	44

### Figuras

Figura 1. Ejemplo de configuración de cosecha en archivo <i>config.xml</i> de Shell-oaiharvester .....	28
Figura 2. Diagramas de flujo de datos de los programas desarrollados .....	32
Figura 3. Primeras líneas de un registro de metadatos cosechado en formato XOAI .....	38
Figura 4. Primeras líneas de la hoja de estilo <i>xoai.xsl</i> .....	38
Figura 5. Registro de ejemplo de la base de salida <i>conicet</i> .....	39
Figura 6. Interfaz de selección de tipo documental .....	41
Figura 7. Interfaz de selección de unidad de investigación .....	41
Figura 8. Interfaz de muestra secuencial de recursos .....	42
Figura 9. Interfaz de ingreso manual de datos .....	42

## INTRODUCCIÓN

El tema general del trabajo final integrador es la gestión de información científica y tecnológica en repositorios digitales institucionales, siendo también mi área de desarrollo profesional. En esta Especialización, los repositorios institucionales fueron tema central del seminario cursado en 2016 y dictado por la Prof. Mónica Pené, “Planificación, desarrollo y sostenimiento de repositorios institucionales”.

En dicho curso, y siguiendo la definición de Alonso Arévalo, se consideró como repositorio institucional de acceso abierto toda colección de objetos digitales basada en la Web, de material académico producido por los miembros de una institución (o varias), con una política definida, cuyas características más importantes son: a) El auto-archivo: el contenido es depositado por el creador, propietario o una tercera parte en su nombre; b) la interoperabilidad: el uso de procesos normalizados que posibilitan la interconexión con otros archivos abiertos; c) la libre accesibilidad: la ausencia de restricciones para acceder al contenido de forma permanente, y d) la preservación a largo plazo.

También se consideró la definición proporcionada en la resolución 469/11 de creación del Sistema Nacional de Repositorios Digitales, en la que se definieron los repositorios digitales como “ (...) aquellas colecciones digitales de la producción científico-tecnológica de una institución en las que se permite la búsqueda y la recuperación para su posterior uso nacional e internacional. Un repositorio digital contiene mecanismos para importar, identificar, almacenar, preservar, recuperar y exportar un conjunto de objetos digitales, normalmente desde un portal web. Esos objetos son descritos mediante etiquetas o metadatos que facilitan su recuperación. A su vez, los repositorios digitales son abiertos e interactivos, pues cumplen con protocolos internacionales que permiten la interoperabilidad entre ellos.”

Estas definiciones permiten apreciar la variedad de subáreas de interés que contiene el área temática elegida, dada la característica de los repositorios digitales abiertos de ser sistemas complejos que conjugan funciones propias de las bibliotecas digitales y de los archivos tradicionales, todo ello bajo el paraguas de las tecnologías propias de los archivos abiertos. De ellas se desprenden cuatro grandes ámbitos de

gestión, que a su vez incluyen subáreas de posible interés para la investigación y el diseño:

- gestión administrativa (planeamiento, financiamiento, diseño de políticas, administración de recursos, coordinación de tareas y proyectos, aspectos legales, entre otras);
- gestión de colecciones (diseño de colecciones, descripción y análisis de recursos bibliográficos, procesamiento documental, representación de la información, esquemas de metadatos y uso de vocabularios controlados, entre otras);
- gestión de servicios (interfaces web de búsqueda, navegación y consulta, diseño de servicios puntuales tales como autoarchivo, estadísticas, etc., y aspectos comunicacionales, entre otras);
- gestión tecnológica (aspectos técnicos en general, servidores y terminales de trabajo, implementación de protocolos, interoperabilidad con otros sistemas, mantenimiento y preservación de archivos bibliográficos, documentales y administrativos, administración de bases de datos, desarrollo de aplicaciones, entre otros).

Para el desarrollo del trabajo final integrador he seleccionado, dentro del área temática de los repositorios institucionales abiertos, la gestión tecnológica en apoyo a la gestión de colecciones digitales, con el fin de desarrollar una herramienta informática para Memoria Académica<sup>1</sup>, el repositorio institucional de la Facultad de Humanidades y Ciencias de la Educación de la Universidad Nacional de La Plata (FaHCE-UNLP) y del Instituto de Investigaciones en Humanidades y Ciencias Sociales, de doble dependencia CONICET-UNLP. Creado en 2009, el IdIHCS<sup>2</sup> pertenece al Centro Científico Tecnológico CONICET La Plata de la Red Institucional del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). La producción de los integrantes del IdIHCS, debido a la doble dependencia del Instituto, es incluida tanto en el repositorio Memoria Académica como en CONICET

---

1 <http://www.memoria.fahce.unlp.edu.ar>

2 <https://idihcs.fahce.unlp.edu.ar/>

Digital<sup>3</sup>, el repositorio institucional del CONICET; pero ambos conjuntos no están sincronizados, ya que sus procesos de trabajo siguen modalidades distintas. Así, mientras en Memoria Académica la ingesta ocurre principalmente por autoarchivo, carga retrospectiva y flujo interno, en CONICET Digital se aborda la curaduría de datos registrados por los investigadores en el Sistema Integral de Gestión y Evaluación<sup>4</sup> (SIGEVA).

Se presenta entonces un trabajo final integrador consistente en el desarrollo de una herramienta que sirve de apoyo tecnológico a la gestión de colecciones digitales, permitiendo la incorporación de recursos digitales mediante la cosecha, verificación e ingesta de metadatos y documentos del set IdIHCS del repositorio CONICET Digital al repositorio institucional Memoria Académica.

---

3 <https://www.conicet.gov.ar/conicet-digital/>

4 <https://sigeva.conicet.gov.ar/>

## JUSTIFICACIÓN

El trabajo final integrador se enmarca en la política general llevada adelante por Memoria Académica<sup>5</sup>, consistente en la reunión, organización, difusión y preservación de toda la producción científica, académica y tecnológica creada y editada en el ámbito de la FaHCE-UNLP y del IdIHCS; y en particular, en las tareas llevadas a cabo en los últimos años, consistentes en:

- la creación en 2016 del servicio de perfiles de autor<sup>6</sup>, que reúne en forma personalizada toda la producción científico-académica de quienes desempeñan funciones docentes, de investigación, de gestión y/o técnicas en la FaHCE/IdIHCS; un servicio muy valorado por toda la comunidad FaHCE/IdIHCS y sostén a su vez de nuevos servicios;
- durante 2018-2019, la búsqueda, identificación e incorporación manual con carácter retrospectivo de todos los artículos de revistas disponibles en acceso abierto de los investigadores y becarios del IdIHCS;
- desde 2022, la generación de informes estadísticos dirigidos a los centros, áreas y laboratorios de investigación del IdIHCS, centrados en la tipología, el volumen y el uso (descargas) de las publicaciones de sus integrantes almacenadas en el repositorio;
- desde 2023, el seguimiento personalizado de la producción reciente alojada en el repositorio perteneciente a los investigadores y becarios de investigación del IdIHCS.

Para poder seguir brindando estos y nuevos servicios en forma consistente, y dado que todos se basan en la producción científico-académica alojada en Memoria Académica, es necesario reunir toda la producción faltante de los miembros de la comunidad, en especial de los investigadores y becarios de investigación. La incorporación de documentos faltantes desde CONICET Digital permitirá así:

---

5 En adelante Memoria Académica, o Memoria.

6 <https://www.memoria.fahce.unlp.edu.ar/memoria/menus/autores-fahce-idihcs/perfiles-autor>

- no sobrecargar de tareas a los investigadores, evitando que vuelvan a depositar en Memoria documentos que ya depositaron en otros sistemas;
- reunir toda la producción científica y técnica del personal docente, técnico y de investigación de la FaHCE/IdIHCS en Memoria;
- proveer perfiles de autor más completos;
- proveer informes estadísticos y encarar tareas de seguimiento más consistentes;
- utilizar los recursos disponibles para dar apoyo tecnológico a la gestión de colecciones digitales.

Una tarea similar de articulación de contenidos se inició en Memoria Académica en 2019 con el Servicio de Difusión de la Creación Intelectual<sup>7</sup> (SeDiCI), perteneciente a la Universidad Nacional de la Plata en su conjunto. Para tal fin, diseñamos un programa (*script*) que permite incorporar a las bases de datos de Memoria Académica los metadatos de los recursos alojados en el set FAHCE de SeDiCI; pero en este caso, tanto el chequeo de correspondencias entre registros como la descarga de documentos y la generación de informes son procesos no automatizados.

Una incorporación inicial de recursos de CONICET Digital a Memoria Académica se realizó en 2020, a partir de una planilla de datos de artículos ya incorporados a CONICET Digital provista por el IdIHCS, a cargo de la curaduría de datos. Este trabajo inicial, manual en su mayor parte, reveló la necesidad de automatizar los procesos a fin de incorporar toda la producción curada por el IdIHCS de un modo más eficiente.

Con ese fin, y tomando como antecedente el *script* desarrollado anteriormente, se presenta el desarrollo de un nuevo programa con más funcionalidades centrado esta vez en la incorporación de recursos (metadatos y documentos), no encarada en forma sistemática hasta el presente, del repositorio CONICET Digital a Memoria Académica.

---

7 <http://sedici.unlp.edu.ar/>

## MARCO TEÓRICO

### Sistemas de información documental e interoperabilidad

Los sistemas se definen como aquel conjunto de elementos que interactúan entre sí en pos de un objetivo común, de modo tal que sus propiedades derivan no de sus elementos constituyentes sino de la interacción e interdependencia existente entre ellos, conformando un todo cuyas propiedades son inherentes al conjunto. Constan de entradas, salidas y procesos, además de un carácter abierto, es decir, de interacción con el entorno (Senn, 1992:20; Ponjuán et al., 2004).

Los sistemas de información son aquellos destinados en su uso a procesar y proveer elementos informativos (datos, documentos, objetos, información), de allí que las bibliotecas y museos, así como también los sistemas de gestión de bases de datos, son considerados tipos de sistemas de información (Moreiro González, 1995; Ponjuan et al., 2004). Cabe hacer la distinción entre sistema de información en tanto entidad abstracta –representando el medio por el cual los datos fluyen dentro de una organización-, es decir, desde un punto de vista social y no determinista, y sistema de información en tanto sistema informático, determinado por algoritmos y sustentado en *hardware* y *software* (Senn, 1992:20; Lapiedra Alcamí, Devece Carañana y Guiral Herrando, 2011:14). En esta segunda acepción del término, puede definirse un sistema de información como aquel “desarrollo informático que involucre un procesamiento de datos e información, y que apoye a las organizaciones en la optimización, mejoramiento y desarrollo de sus procesos” (Gómez, 2007).

Siguiendo esta línea, los sistemas de información documental son un tipo de sistema de información caracterizado por involucrar procesamiento de datos y de documentos u objetos digitales, e incorporación de estándares para su descripción e identificación unívoca. Su función es “permitir la recuperación y acceso a la información documental almacenada, realizando para ello un proceso de representación, organización, análisis y administración para que aquella pueda llegar al usuario que la necesita, de acuerdo con sus necesidades, propósitos e intereses” (Naranjo Vélez, 2014:58). Pueden incluir tipologías documentales de muy variado tipo y ser en sí mismos sistemas complejos; su heterogeneidad está dada por el uso

variado de idiomas, datos, metadatos, protocolos y tecnologías, así como también por las características de sus usuarios. Es decir, su heterogeneidad está dada básicamente por los sistemas informáticos en que se gestionan y por la información y los documentos que albergan (Gómez, 2007).

Esta heterogeneidad atenta contra la capacidad de interactuar de los sistemas documentales entre sí y con otros sistemas de información, es decir, contra la capacidad de intercambiar información y proveer nuevos servicios y formas de recuperación conjunta. A esta capacidad se la conoce como interoperabilidad, y es definida como la capacidad de dos o más sistemas o componentes de intercambiar información y utilizarla (IEEE, 1990). En particular, se la puede entender como la capacidad de un sistema de información de compartir información y/o documentos digitales con otros sistemas de forma automática y con bajo o nulo nivel de pérdida de datos, sin consideración sobre las condiciones de *hardware* y *software* de cada uno de dichos sistemas.

Es evidente que la interoperabilidad, para ser fluida, requiere de una tecnología de base compartida, referida tanto a los requerimientos tecnológicos de la conexión a establecer como a la capacidad de interpretación de la información provista u obtenida; es decir que requiere la adopción común de protocolos de interoperabilidad y esquemas de metadatos para la descripción documental. Esta interoperabilidad incluye así distintos elementos o niveles de acción: la sintaxis (por ejemplo, el uso de lenguajes estructurados que permitan interfaces comunes de acceso y recuperación de la información), la semántica (sobre todo en el uso compartido de esquemas de metadatos a la descripción), la infraestructura (normalmente Internet y sus protocolos de acceso e intercambio entre sistemas), y la estructura (es el caso, por ejemplo, de la provisión de datos enlazados de la Web Semántica) (Castañeda de León, 2004; Gómez, 2007).

Las bibliotecas, tanto tradicionales como digitales, tienen una larga tradición en el establecimiento de condiciones técnicas y metodológicas para compartir su información, ya sea a través de la integración de sus sistemas como mediante la generación de catálogos colectivos y el uso de protocolos (como, por ejemplo, el protocolo Z39.50) para compartir recursos de información. Los repositorios digitales,

por su parte, siendo iniciativas llevadas adelante mayormente por bibliotecas de instituciones educativas y de ciencia y técnica nucleadas en torno al movimiento por el acceso abierto (*open access*), han tenido desde sus inicio a fines del siglo XX al uso de software abierto, a la implementación de estándares para la descripción de sus recursos y a la interoperabilidad de sus sistemas como elementos constituyentes. Si bien funcionan mayoritariamente en sistemas públicos, caracterizados por la heterogeneidad de sus sistemas informativos y documentales y por los desafíos tanto técnicos como sociales y culturales que implica implementar sistemas interoperables (Míguez, 2021), han logrado un consenso inicial que los fortaleció y dio lugar a un desarrollo creciente en los últimos 20 años (COAR, 2011).

El uso que los repositorios hacen de sistemas y estándares abiertos, base de su existencia, los pone en una situación excepcional en la cual pueden compartir sus recursos y dar lugar a servicios de valor agregado y a redes nacionales y regionales de repositorios sin excesivas inversiones en recursos humanos y tecnología.

### **Repositorios digitales abiertos y protocolo OAI-PMH**

El movimiento de acceso abierto es la iniciativa internacional surgida en la década de 1990 en el seno de la comunidad científica y formalizada a inicios de este siglo mediante una serie de declaraciones públicas, cuyo objetivo es promover la remoción de las barreras económicas y de algunas barreras legales que históricamente han restringido el acceso libre y gratuito a la literatura científica. Las dos estrategias inicialmente recomendadas a los científicos y académicos para lograrlo aún continúan vigentes, y son: el autoarchivo en archivos electrónicos abiertos y la publicación en revistas de acceso libre y gratuito (Alonso-Arévalo, Subirats-Coll y Martínez-Conde, 2008; Suber, 2015)

Los repositorios digitales constituyen una revolución reciente en el sistema de comunicación científica. Surgen a partir de la Iniciativa de Archivos Abiertos OAI<sup>8</sup> (*Open Archives Initiative*), la cual desarrolló en 2002 la tecnología necesaria para mantener archivos electrónicos interoperables mediante un protocolo de red, el Protocolo OAI para la Recolección de Metadatos OAI-PMH (*Protocol Metadata Harvester*). En el contexto OAI, al igual que en el contexto OA, el término archivo

---

8 <http://www.openarchives.org/>

(*archive*) es sinónimo de repositorio de trabajos académicos (*repository of scholarly papers*), pero el término abierto (*open*) refiere ya no al acceso libre y gratuito a sus contenidos sino a su cualidad de ser interoperable. Así, los repositorios digitales abiertos son definidos en este contexto como aquellos servidores proveedores de datos (*data providers*) accesibles en red capaces de procesar correctamente peticiones OAI-PMH de proveedores de servicios (*service providers*), que usan los metadatos así recogidos para desarrollar servicios de valor agregado como, por ejemplo, una interfaz de búsqueda unificada (Alonso-Arévalo et al., 2008; Lagoze, Van de Sompel, 2008; Suber, 2015)

El protocolo OAI-PMH define un mecanismo de bajo costo y aplicación técnica sencilla para la recolección de registros de metadatos de los repositorios digitales de acceso abierto, sirviendo así de sustento a un nuevo modelo de comunicación científica basado en los estándares abiertos HTTP (*Hypertext Transport Protocol*) y XML (*Extensible Markup Language*), y en el uso no exclusivo del esquema de metadatos Dublin Core no calificado<sup>9</sup> (DC) para proporcionar un nivel básico de interoperabilidad. OAI-PMH ha devenido en el protocolo de interoperabilidad estándar en la implementación de repositorios. Esta opción por tecnologías de uso extendido y sencilla aplicación, sumado al crecimiento sostenido de los repositorios digitales en todo el mundo y a la existencia conjunta de repositorios institucionales (nacionales y regionales), temáticos y de sets de datos, es lo que ha permitido contar con una inmensa cantidad de información científica estructurada y disponible en la Web acorde a los estándares de la W3C (*World Wide Web Consortium*), susceptible de ser integrada a otros sistemas y generar nuevos servicios (COAR, 2011; Lagoze, Van de Sompel, 2008; Suber, 2015).

La especificación del protocolo (OAI, 2002) establece las siguientes definiciones:

- Un cosechador es una aplicación cliente que gestiona peticiones OAI-PMH.
- Un repositorio es un servidor accesible en red que puede procesar peticiones OAI-PMH.

---

9 <https://www.dublincore.org/specifications/dublin-core/dces/>

- Proveedores de datos (*data providers*) son aquellos repositorios que exponen metadatos estructurados vía OAI-PMH.
- Proveedores de servicios (*service providers*) son aquellos que cosechan metadatos vía OAI-PMH, como base para la creación de servicios de valor añadido.

Así, un repositorio es gestionado como un proveedor de datos para exponer metadatos a los cosechadores, en uno o más formatos; mientras que un cosechador es operado por un proveedor de servicios como medio de reunir colecciones de metadatos de los repositorios.

OAI-PMH soporta seis peticiones o verbos que pueden ser invocados mediante protocolo HTTP, devolviendo en todos los casos archivos formateados en XML:

1. *Identify*, recupera información sobre el proveedor de datos.
2. *ListMetadataFormats*, recupera la lista de formatos bibliográficos que utiliza el proveedor.
3. *ListIdentifiers*, recupera los encabezamientos de los registros en lugar de los registros completos.
4. *ListRecords*, recupera los registros completos.
5. *GetRecord*, recupera un registro concreto.
6. *ListSets*, recupera un conjunto de registros agrupados previamente por el proveedor de datos.

Los registros de metadatos son extraídos gracias a la asignación de un identificador único a cada ítem, entendido éste como cada uno de los contenedores de metadatos, en múltiples formatos, acerca de un único recurso; de tal modo que los ítems de un mismo recurso, independientemente de su formato de cosecha, comparten el mismo identificador. Como respuesta a una petición OAI-PMH, se obtienen uno o más registros en un solo formato, previamente definido en la petición, gracias al identificador de cada ítem que sigue una sintaxis URI (*Uniform Resource Identifier*).

Los registros obtenidos mediante petición OAI-PMH se identifican unívocamente mediante el identificador único del ítem y el prefijo de metadatos (*metadata\_prefix*) en que deben estar formateadas las respuestas *ListRecords* y *GetRecord*. Otro dato relevante de la petición, en especial en cosechas incrementales, es la fecha (*datestamp*) en que se ha creado, modificado o eliminado el registro.

Ejemplos de sintaxis URI de peticiones OAI-PMH, en protocolo HTTP:

- Un registro concreto en formato Dublin Core:

```
https://ri.conicet.gov.ar/oai/request?  
verb=GetRecord&identifier=oai:ri.conicet.gov.ar:11336/179477&metadataPrefix=oai_dc
```

- Un conjunto de registros pertenecientes a un set determinado en formato Dublin Core:

```
https://ri.conicet.gov.ar/oai/request?  
verb=ListRecords&metadataPrefix=oai_dc&set=snr
```

- Una lista de sets disponibles de un repositorio:

```
https://ri.conicet.gov.ar/oai/request?verb=ListSets
```

Cada registro de metadatos cosechado mediante OAI-PMH tiene la siguiente estructura:

1. Una cabecera (*header*) conteniendo el identificador único del ítem, el *datestamp*, el estado (*status*) del registro, y el o los sets (*SetSpec*) a los que pertenece el ítem.
2. Un cuerpo de metadatos descriptivos del recurso, que si bien puede estar en muchos formatos diferentes debe estar al menos expuesto en formato Dublin Core sin calificar.
3. Un conjunto opcional de metadatos adicionales (*about*), usualmente sobre responsables de creación del registro, derechos y licencias de uso, y datos de preservación.

Los sets son agrupamientos de ítems establecidos por el proveedor de datos a los fines de una cosecha selectiva. Sin límites en la cantidad de sets que pueden

establecerse, generalmente siguen criterios basados en tipos documentales o bien comunidades o áreas académicas, así como también en función de cosechadores puntuales. En este sentido, un requisito de los sistemas nacionales y regionales de repositorios es que los repositorios integrantes definan un set específico que agrupe los ítems que han de ser cosechados por el sistema.

### **Repositorios institucionales**

Los repositorios institucionales de acceso abierto son archivos digitales que reúnen y dan acceso libre y gratuito a la producción científica y tecnológica generada en el ámbito de una institución de ciencia y técnica. Dar visibilidad a dicha producción y preservarla a largo plazo son sus principales objetivos. El protocolo OAI-PMH ha sido hasta ahora la herramienta utilizada por los repositorios digitales (tanto institucionales, como temáticos y de datos) para la provisión de datos en Internet cosechables por parte de otros sistemas proveedores de servicios, permitiendo un nivel básico de interoperabilidad –entendida como la capacidad de interactuar entre sí de uno o más sistemas- gracias al uso no exclusivo del esquema de metadatos Dublin Core. El uso extendido del protocolo, sumado al crecimiento sostenido de los repositorios de acceso abierto en todo el mundo, ha permitido contar con una inmensa cantidad de información científica estructurada y disponible en la Web, basada en estándares abiertos y susceptible de ser integrada a otros sistemas y generar nuevos servicios.

Se entiende aquí por repositorio institucional de acceso abierto "una colección de objetos digitales basada en la Web, de material académico producido por los miembros de una institución (o varias), con una política definida, cuyas características más importantes son: a) El auto-archivo: el contenido es depositado por el creador, propietario o una tercera parte en su nombre. b) La interoperabilidad: el uso de procesos normalizados que posibilitan la interconexión con otros archivos abiertos. c) La libre accesibilidad: la ausencia de restricciones para acceder al contenido de forma permanente. d) La preservación a largo plazo." (Alonso-Arévalo et al.)

La gestión de la información, entendida como acción del profesional de la información y como proceso o actividad llevada adelante en una organización, tiene

por objetivo lograr eficiencia en los procesos y eficacia en los resultados obtenidos, a través de actividades propias tales como la identificación de necesidades y demandas de información; la selección y filtrado de la información; su análisis, organización y almacenamiento, y la elaboración de productos y servicios informativos (Valentim, 2017). En el caso de los repositorios institucionales, la gestión de la información es central e involucra no sólo datos bibliográficos sino también de personas y organizaciones, así como de preservación, acceso y uso de los recursos; mientras que sus modos de representación y provisión, aún para un mismo repositorio, pueden asumir distintas formas, según los requerimientos planteados por cada servicio o sistema (Testa, 2013).

La aplicación de perfiles de autor a los repositorios institucionales surge inicialmente como un servicio de valor agregado destinado a promover la práctica del autoarchivo y el uso general del repositorio entre los miembros de una institución, dentro de una estrategia más amplia orientada al crecimiento y la sostenibilidad de los repositorios. La incorporación de las utilidades propias de las redes sociales científicas a los repositorios digitales dió lugar a los perfiles de autor como un servicio de valor agregado, que reúne la producción de un autor almacenada en el repositorio, enriquecida con estadísticas de uso, información curricular y de contacto, etc. El beneficio que representa para un autor contar con un perfil público en el repositorio de su institución es equivalente al beneficio que para una institución representa contar con un repositorio digital, en términos de ventaja competitiva, valor añadido, incremento de la visibilidad y facilidades de recuperación y acceso a la producción generada. (Barrueco, 2013; Bongiovani, 2013; Genovés, 2017).

En la región iberoamericana, las universidades públicas y las agencias nacionales de política y financiamiento científicos son quienes llevan adelante la iniciativa de acceso abierto a los resultados de las investigaciones financiadas con fondos públicos, pues son los principales actores involucrados. Existe una importante infraestructura de acceso abierto no comercial para dar visibilidad a las publicaciones en revistas dirigidas por instituciones académicas, que es modelo en todo el mundo por funcionar sin costos para el lector ni para el autor. Entre los

portales de revistas de acceso abierto en la región destacan LATINDEX<sup>10</sup>, Redalyc<sup>11</sup>, SciELO<sup>12</sup>, Dialnet<sup>13</sup>, Red Iberoamericana de Innovación y Conocimiento Científico (REDIB)<sup>14</sup> y e-Revistas<sup>15</sup>. Pero las políticas y legislaciones en materia de ciencia y técnica, tanto nacionales como institucionales, han privilegiado la vía verde, esto es, el depósito de las publicaciones en repositorios digitales de acceso abierto. Así, los repositorios institucionales han sido la vía principal de implementación de políticas de acceso abierto tanto a nivel institucional como nacional en la región, destacándose entre ellos LA Referencia, la red de repositorios digitales institucionales de América Latina (Babini y Rovelli, 2020).

LA Referencia conforma una infraestructura federada de nodos nacionales para dar acceso abierto a la producción científica generada en las universidades e instituciones de investigación de la región, y una plataforma técnica basada en el protocolo OAI-PMH. Actualmente da acceso a más de 5 millones de documentos (artículos, tesis, reportes) y está conformada por doce países latinoamericanos: Argentina, Brasil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, España, México, Panamá, Perú y Uruguay; los nodos nacionales son operados por instituciones de ciencia y tecnología y en algunos casos por un consorcio de universidades o red académica.

En nuestro país, es nodo nacional el Sistema Nacional de Repositorios Digitales (SNRD), una iniciativa del Ministerio de Ciencia, Tecnología e Innovación y el Consejo Interinstitucional de Ciencia y Tecnología (CICyT), establecido con la sanción de la Ley 26.899 de Creación de Repositorios Digitales Abiertos y su posterior reglamentación en 2016. Actualmente nuclea 48 repositorios institucionales argentinos y da acceso a unos 500 mil recursos, proveyendo a LA Referencia con alrededor de 350 mil. Su infraestructura técnica se basa también en el protocolo OAI-PMH para la cosecha de recursos, requiriendo de los repositorios la conformación de un set OAI SNRD y el cumplimiento de estándares comunes para

---

10 <http://latindex.org>

11 <http://www.redalyc.org>

12 <http://www.scielo.org/php/index.php>

13 <http://www.scielo.org/php/index.php>

14 <https://www.redib.org/>

15 [https://ddd.uab.cat/pub/ciencias/ciencias\\_a2012m3n21/suplement/index.html.4](https://ddd.uab.cat/pub/ciencias/ciencias_a2012m3n21/suplement/index.html.4)

la creación de metadatos a fin de lograr la interoperabilidad entre sistemas (SNRD, 2015)

### **CONICET Digital**

CONICET Digital es el Repositorio Institucional de Acceso Abierto del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), creado en 2015 para dar cumplimiento a la Ley 26.899. Es una plataforma digital, de acceso abierto y gratuito, con la producción científico-tecnológica de los investigadores y becarios que integran las instituciones agrupadas en el conjunto de Centros Científicos Tecnológicos distribuidos en todo el país y que conforman, a su vez, la Red Institucional del CONICET. El repositorio se nutre del Sistema Integral de Gestión y Evaluación (SIGEVA) en el cual todos los años, los investigadores y becarios del CONICET archivan su producción científico-tecnológica.

CONICET Digital contiene actualmente más de 224 mil recursos documentales (artículos de revistas en su mayoría, pero también documentos de conferencia, libros y capítulos de libros y tesis), pertenecientes a los casi 12 mil investigadores y 11 mil becarios que integran el CONICET; y aporta unos 158 mil recursos al SNRD. El set OAI del IdIHCS en CONICET Digital, por su parte, contiene unos 3000 recursos documentales, correspondientes a un conjunto aproximado de 820 investigadores y becarios que integran el IdIHCS.

La modalidad de trabajo de CONICET Digital difiere del resto de los repositorios institucionales en su forma de trabajo, que es descentralizada, y en que sus colecciones provienen del Banco de Datos del SIGEVA CONICET, en el cual los investigadores, becarios y personal de apoyo CONICET autoarchivan sus trabajos. Esta producción es distribuida por la Oficina Técnica del repositorio a todo el país para su procesamiento, a través de una Red Federal de Especialistas distribuidos en las distintas dependencias del CONICET, encargados de verificar y procesar los datos y evaluar su disponibilidad en acceso abierto de la producción correspondiente a su unidad ejecutora. Hoy esta Red cuenta con 85 miembros, todos ellos bibliotecarios y profesionales de la información, quienes realizan una curaduría de datos, a fin de cumplir con los estándares descriptivos de la Web y sistemas específicos, como el SNRD, y un análisis de las políticas editoriales, a fin de

determinar el tipo de acceso que se le dará al documento en el repositorio (abierto, embargado o restringido, si hubiera convenios previos).

El ciclo de depósito de publicaciones en CONICET Digital comprende así las siguientes fases: 1) autoarchivo, en SIGEVA; 2) compilación, a fin de identificar registros semejantes y preparar los lotes que se distribuirán posteriormente; 3) unificación, para una revisión manual de los registros y la unificación en uno solo de los registros semejantes, en base al mejor construido; 4) curatoría, hecha desde la plataforma CONICET Digital, para controlar los metadatos y agregar datos en base a vocabularios controlados, y 5) depósito, para la puesta a disposición pública de los recursos desde el sitio web del repositorio (CONICET, 2023)

En CONICET Digital el contenido está organizado en comunidades que reflejan la Red Institucional del CONICET: Centros Científicos Tecnológicos (CCT) y sus Unidades Ejecutoras (UE) y Unidades Asociadas. Luego, cada Unidad Ejecutora contiene colecciones por tipo de documento. Se incluyen los siguientes tipos documentales: a) artículos, comprende todos aquellos textos publicados en una revista científica o académica; b) capítulos de libro: partes de un libro tomado en forma independiente; c) libros: monografías publicada por una editorial; d) eventos científicos: trabajos publicados que han sido presentados en eventos científicos; e) audiovisuales: objetos de audio o video; f) tesis de doctorado, y g) conjuntos de datos. Dentro del CCT La Plata, el IdIHCS, por su parte, contiene actualmente todos los tipos documentales mencionados, excepto objetos de audio o video.

### **Memoria Académica**

Memoria Académica es el repositorio digital de la Facultad de Humanidades y Ciencias de la Educación de la Universidad Nacional de La Plata (FaHCE-UNLP) y del Instituto de Investigaciones en Humanidades y Ciencias Sociales (IdIHCS) gestionado por la Biblioteca Prof. Guillermo Obiols (BIBHUMA) y creado en 2006 con el propósito de reunir, registrar y dar acceso abierto y público a la producción académica de la institución. Con este propósito reúne, además de los artículos, libros y capítulos y documentos de conferencia de docentes e investigadores realizados en el ámbito de sus actividades en la FaHCE/IdIHCS, las tesis de doctorado y maestrías y las tesinas de licenciatura y trabajos de final de carrera; las

publicaciones de todo tipo editadas por la Facultad; los programas de materias, cursos y seminarios, y los planes de estudio de las carreras de grado y posgrado que se cursan en su ámbito (Resolución Nro. 527/2021; Mallo et al., 2005; Fushimi et al., 2006). Actualmente registra casi 60 mil archivos de metadatos y da acceso libre y gratuito a más de 47 mil documentos a texto completo, de los cuales unos 36 mil -correspondientes a tesis de grado y posgrado, publicaciones en revistas, libros y capítulos de libros, informes, proyectos de investigación y documentos de conferencia- son cosechados por el SNRD.

El servicio de perfiles de autor de Memoria Académica, lanzado en mayo de 2016, fue muy bien recibido por los docentes e investigadores FaHCE/IdIHCS. No sólo logró el objetivo de promover el autoarchivo sino que, además, fortaleció el lazo con toda la comunidad académica; dió lugar además a una nueva etapa, caracterizada por el intercambio de información entre el repositorio y distintas áreas académicas proveedoras de datos necesarios para el desarrollo del servicio, y por un consecuente trabajo interno de curación de datos recogidos de diversas fuentes. Actualmente está conformado por alrededor de 1340 perfiles de docentes, investigadores y personal técnico de la FaHCE-UNLP, de los cuales unos 800 corresponden a investigadores y becarios de la institución en actividad, lo cual implica en ese segmento una cobertura del 97% y dar visibilidad a un volumen de producción de más de 13 mil recursos.

Los investigadores y becarios del IdIHCS tienen como lugar de trabajo o bien el IdIHCS en su conjunto, o bien algunas de las 18 unidades de investigación que conforman actualmente el Instituto. Estas unidades son 14 centros de investigación: Centro de Historia Argentina y Americana (CHAyA), Centro de Estudios Helénicos (CEH), Centro de Estudios e Investigaciones Lingüísticas (CEIL), Centro de Estudios Latinos (CEL), Centro de Estudios de Literaturas y Literaturas Comparadas (CELyLC), Centro de Estudios de Sociedades Precapitalistas (CESP), Centro Interdisciplinario Cuerpo, Educación y Sociedad (CICES), Centro de Investigaciones Geográficas (CIG), Centro Interdisciplinario de Metodología en Ciencias Sociales (CIMECS), Centro Interdisciplinario de Investigaciones en Género (CInIG), Centro de Investigaciones Socio - Históricas (CISH), Centro de Investigaciones en Filosofía (CieFi), Centro de Estudios de Teoría y Crítica Literaria (CTCL), Centro de Estudios

e Investigaciones de Ed. Física (CEIdEF); 3 laboratorios: Laboratorio de Estudios en Sociología y Economía del Trabajo (LESET), Laboratorio de Investigaciones en Traductología (LIT), Laboratorio de Investigación e Innovación en Educación en Ciencias Exactas y Naturales (LIIECEyN), y un área de investigación: Área de Español como Lengua Segunda y Extranjera (ELSE).

Los perfiles de autor, gracias al trabajo de curaduría de datos que conlleva, fueron la base de nuevos servicios que se desarrollaron en años siguientes. Así, durante 2018-2019 se abordó la búsqueda e incorporación masiva y manual de artículos de revistas de investigadores del IdIHCS, labor que permitió ingresar 780 artículos en acceso abierto a Memoria Académica. En 2022, se inició un proceso de acercamiento a los centros, áreas y laboratorios de investigación del IdIHCS para fortalecer los lazos con Memoria Académica e interiorizarse de su situación respecto al acceso abierto y el autoarchivo, para lo cual se elaboraron informes estadísticos por unidad de investigación, basados en la producción de sus integrantes registrada y almacenada en Memoria Académica. La propuesta fue muy bien recibida y continúa en 2023, acrecentada por una labor de seguimiento personalizado de la producción almacenada en Memoria de los investigadores y becarios, discriminada por unidad de investigación, con el propósito de contar con perfiles de autor más completos y actualizados, y fortalecer el cumplimiento de la ley de acceso abierto y el compromiso de Memoria Académica con la comunidad FaHCE/IdIHCS en pos de dar visibilidad, difusión y preservación a la producción científica generada por sus miembros.

## MATERIALES Y MÉTODO

El presente trabajo tuvo por objetivo principal incorporar la producción del IdIHCS registrada en CONICET Digital y faltante en Memoria Académica, mediante el desarrollo de una herramienta (programa informático) que automatiza los procesos y sirve de apoyo tecnológico a la gestión de colecciones digitales y a la provisión de servicios basados en perfiles de autor.

La metodología seguida para su desarrollo siguió en general el modelo de análisis y diseño estructurado, basado en funciones, procesos y flujo de datos (Senn, 1992:38). Durante la elaboración del plan del trabajo final integrador, se abordó el análisis de requerimientos y de recursos tecnológicos y el diseño general de la herramienta, consistente en dos programas funcionalmente independientes; posteriormente, aprobado el plan, se elaboraron los distintos componentes de ambos desarrollos en forma separada y basado en funciones, efectuando pruebas y rectificaciones en la escritura del código hasta su ensamblaje final. Se elaboraron primeras versiones de ambos programas para su evaluación, incompletas en su alcance aunque funcionales, y tras su puesta a prueba se adoptó un enfoque ágil en la corrección de errores y la mejora, incorporando nuevas funcionalidades en base a los nuevos requerimientos planteados, hasta su implementación final.

En primer término, y siguiendo los objetivos establecidos en el plan de trabajo y la metodología empleada, se analizaron los requerimientos funcionales, las posibilidades técnicas y las herramientas de software disponibles para el desarrollo de la herramienta, la cual debía automatizar los siguientes procesos:

- Cosecha de metadatos del set OAI del IdIHCS en el repositorio CONICET Digital, identificado como *com\_11336\_147* ;
- identificación de recursos ya existentes en Memoria Académica;
- transformación de los metadatos cosechados a los requerimientos de las bases de datos de Memoria Académica;
- incorporación de registros y documentos a los servidores y bases de datos de Memoria Académica;

- elaboración de informes de resultados.

En cuanto a las posibilidades técnicas y herramientas de software, se consideró lo siguiente:

- Memoria Académica trabaja con bases de datos en CDS/ISIS para artículos, documentos de conferencia, tesis y proyectos de investigación, y con base de datos Koha, para libros, capítulos y otros trabajos monográficos. Las herramientas a utilizar deben ajustarse a estos sistemas, y los registros a ingresar, a los formatos específicos de las bases ISIS y al formato MARC21 de Koha.
- Ya hemos desarrollado numerosos *scripts* en apoyo a la gestión de colecciones en BASH, el lenguaje de comandos de Linux de uso sencillo pero potente, que ha permitido escribir aplicaciones que han resultado fáciles de usar y muy funcionales a necesidades específicas del sector que procesa registros y documentos en Memoria. Junto con BASH, se han trabajado dichos *scripts* con los utilitarios CISIS, indispensables para el trabajo con las bases bibliográficas.
- Existen servidores de procesamiento y de gestión técnica en Memoria montados de tal forma que permiten el intercambio de documentos.

Se constató, además, que todos los tipos documentales del set IDIHCS de CONICET Digital eran susceptibles de incluirse en Memoria Académica, a saber: artículos de revistas, libros y capítulos de libros, documentos de conferencia, tesis doctorales y conjuntos de datos.

Para desarrollar el componente relativo a la cosecha, se cosecharon en primer término todos los formatos expuestos por CONICET Digital, a fin de seleccionar el que resultara más adecuado a las necesidades de Memoria Académica. El repositorio CONICET Digital utiliza el software de gestión DSPACE, y provee sus registros en todos los formatos establecidos por este sistema. Estos formatos corresponden a esquemas de metadatos y son los siguientes:

Esquema	URL del esquema
oaidc	<a href="http://www.openarchives.org/OAI/2.0/oai_dc.xsd">http://www.openarchives.org/OAI/2.0/oai_dc.xsd</a>
mets	<a href="http://www.loc.gov/standards/mets/mets.xsd">http://www.loc.gov/standards/mets/mets.xsd</a>
xoai	<a href="http://www.lyncode.com/schemas/xoai.xsd">http://www.lyncode.com/schemas/xoai.xsd</a>
didl	<a href="http://standards.iso.org/ittf/PubliclyAvailableStandards/MPEG-1_schema_files/didl/didl.xsd">http://standards.iso.org/ittf/PubliclyAvailableStandards/MPEG-1_schema_files/didl/didl.xsd</a>
dim	<a href="http://www.dspace.org/schema/dim.xsd">http://www.dspace.org/schema/dim.xsd</a>
ore	<a href="http://tweety.lanl.gov/public/schemas/2008-06/atom-tron.sch">http://tweety.lanl.gov/public/schemas/2008-06/atom-tron.sch</a>
rdf	<a href="http://www.openarchives.org/OAI/2.0/rdf.xsd">http://www.openarchives.org/OAI/2.0/rdf.xsd</a>
etdms	<a href="http://www.ndltd.org/standards/metadata/etdms/1.0/etdms.xsd">http://www.ndltd.org/standards/metadata/etdms/1.0/etdms.xsd</a>
mods	<a href="http://www.loc.gov/standards/mods/v3/mods-3-1.xsd">http://www.loc.gov/standards/mods/v3/mods-3-1.xsd</a>
qdc	<a href="http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd">http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd</a>
marc	<a href="http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd">http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd</a>
uketd_dc	<a href="http://naca.central.cranfield.ac.uk/ethos-oai/2.0/uketd_dc.xsd">http://naca.central.cranfield.ac.uk/ethos-oai/2.0/uketd_dc.xsd</a>

Tabla 1. Lista de formatos provistos para la cosecha por el repositorio CONICET Digital.

Para cada tipo documental distinto, se analizaron los datos provistos a través de todos los formatos cosechados, considerándose todos los datos relevantes para la descripción requerida por los registros de Memoria y la descarga de los documentos. Se pudo advertir, así, la existencia de un conjunto de datos comunes a todos los esquemas, y diferencias sustanciales relativas a la presencia de algunos datos relevantes, tales como la filiación de los autores, el grado académico y la universidad otorgante en el caso de las tesis, los datos de fuente en caso de capítulos de libros y documentos de conferencia, y la URL del recurso en acceso abierto. Una transcripción completa del análisis de los metadatos considerados relevantes según formato de cosecha se incluye en el Anexo A.

El análisis permitió establecer que el formato más completo es el XOAI, el formato del módulo de gestión de colecciones OAI de DSPACE. Se trabajó entonces sobre este formato de cosecha, sobre el que se aplicaron las transformaciones necesarias para obtener todos los datos relevantes a registrar en Memoria Académica.

En base a este análisis inicial de requerimientos, herramientas y formatos, y siguiendo el plan inicial, se optó por el desarrollo de dos aplicaciones, a ser ejecutadas por dos áreas distintas de la Biblioteca: el área de Tecnologías, y el área de Procesos Digitales, a fin de mantener separadas las tareas de cosecha y

transformación de aquellas centradas en la incorporación específica de registros y documentos a las bases y servidores de Memoria. Esta separación permite la ejecución de cosechas periódicas y a la vez –en paralelo-, ser utilizada por Procesos Digitales, adaptándose a los flujos de trabajo específicos de esta área y sin necesidad de esperar a que se efectúe la cosecha (que actualmente demora unos 15 minutos) para comenzar a incorporar recursos.

Tras su implementación, se requirió una funcionalidad adicional en el programa destinado a Procesos Digitales consistente en la selección de recursos por centro, laboratorio o área de investigación perteneciente al IdHICS, con el fin de agilizar el proceso de seguimiento de autores, el cual se realiza por unidad de investigación. Se escribió entonces este agregado al código, buscando correspondencias entre autores de los documentos de CONICET Digital y personas registradas en la base de autoridades de nombres personales de Memoria Académica, la cual recoge información actualizada de los investigadores y becarios actualmente en funciones, así como sus lugares de trabajo.

En la escritura del código se priorizó la sencillez y eficiencia de las líneas de comandos y de su secuencia, así como también en su interfaz de usuario su facilidad de uso, a fin de eliminar o bien reducir la necesidad de apoyo técnico al mínimo indispensable. Se incluyó también en el código fuente de ambos programas documentación interna, esto es, la clarificación a través de comentarios de las acciones que van siendo realizadas. Los programas se completaron con la documentación externa, en este caso, dos manuales de usuario destinados al área de Procesos Digitales: uno sobre el uso del programa destinado a ella, y otro sobre la importación manual de registros al sistema Koha.

### **Herramientas utilizadas**

A continuación, se describen brevemente las herramientas seleccionadas y utilizadas en el desarrollo de ambos programas.

#### **BASH**

BASH (un acrónimo de *Bourne Again SHell*) es un intérprete de comandos de los sistemas Linux, es decir, un programa que funciona como interfaz de programación

para servir de nexo entre el sistema operativo y el usuario, siendo capaz de leer y ejecutar instrucciones. BASH es sólo uno de los intérpretes de comandos (*shells*) que operan en Linux; viene preinstalado en todas las distribuciones, siendo el *shell* por omisión. Pero además tiene muy poderosas capacidades de programación, lo que le permite no sólo interpretar sus propios comandos sino todos aquellos contruidos en este lenguaje, pudiendo ejecutar de este modo programas (*scripts*) contruidos por los usuarios para resolver necesidades específicas. Es decir, BASH es tanto un intérprete de comandos para interactuar con el sistema como un lenguaje de programación en sí mismo (Tackett, Gunter y Brown, 1995:237; Bash Reference Manual).

Desde el Area de Tecnologías de Bibhuma hemos desarrollado para Memoria Académica numerosos *scripts* en BASH, tanto para automatizar tareas específicas del área, entre otras la ingesta masiva de registros, el control y mantenimiento de colecciones, y la creación de perfiles de autor; como así también para automatizar tareas destinadas al Area de Gestión de Colecciones, entre otros, la validación de registros y la gestión de autoarchivos. Por todo ello, sumado a su sencillez y potencia, es que lo elegimos nuevamente como lenguaje de programación base para el desarrollo de los programas que permiten la cosecha e incorporación de registros y documentos de CONICET Digital al repositorio.

### **Shell OAI Harvester**

*Shell OAI Harvester*<sup>16</sup> es un programa escrito en BASH para la cosecha de metadatos vía OAI-PMH mediante el intérprete de comandos. Permite configurar varios repositorios junto con las características de la cosecha a realizar, tales como set y formato de metadatos, en su archivo de configuración *config.xml*.

Tiene también la posibilidad de validar las cosechas realizadas gracias a un sistema de almacenamiento de errores en la carpeta *tmp* del usuario, es decir, en archivos temporales.

Si bien permite la cosecha incremental basada en la impresión de valores *hash* (sumas de verificación) a través del comando *md5sum*, éstas no se aplican a los

---

16 <https://github.com/wimmuskee/shell-oaiharvester>

registros de metadatos descargados sino a cada uno de los directorios en los que son alojados. Se hicieron pruebas para comprobar si estas sumas de verificación son susceptibles a cambios en los registros de metadatos, con resultado negativo.

Se eligió este cosechador por ser muy fiable y fácil de configurar. Se optó por no usar la cosecha incremental del set correspondiente al IdIHCS sino realizar la cosecha completa cada vez, a fin de no perder información de los registros que pudieran presentar modificaciones en el futuro, como es el caso de recursos bajo embargo.

```
<?xml version="1.0" encoding="UTF-8" ?>
<config xmlns="http://github.com/wimmuskee/shell-harvester/">
  <!-- default record save location -->
  <recordpath>/home/pgenoves/cosechas</recordpath>
  <!-- temporary files are stored here -->
  <temppath>/home/pgenoves/tmp/oaiharvester</temppath>
  <!-- location of log file -->
  <logfile>/home/pgenoves/tmp/oaiharvester-log.csv</logfile>
  <!-- store datestamp and identifier for each record found -->
  <!-- <recordlogfile>/tmp/oaiharvester-recordlog.csv</recordlogfile> -->
  <!-- options for the curl command -->
  <curlopts>--retry 5 --silent --limit-rate 200k --user-agent "shell-oaiharvester"</curlopts>
  <!-- command to execute before delete -->
  <!-- <deletecmd>updateMyDatabase.sh ${path}</deletecmd> -->
  <!-- command to execute after update -->
  <!-- <updatecmd>echo ${identifier} >> /tmp/harvested.txt</updatecmd> -->

  <repository id="conicet_xoai">
    <baseurl>https://ri.conicet.gov.ar/oai/request</baseurl>
    <metadataprefix>xoai</metadataprefix>
    <set>com_11336_147</set>
    <!-- By default the harvesting starts from Identify earliestDatestamp,
    use this only for debugging purposes. -->
  </repository>
</config>
```

Figura 1. Ejemplo de configuración de cosecha en archivo *config.xml* de Shell-oaiharvester.

## xsltproc

*xsltproc*<sup>17</sup> es una herramienta que corre en BASH para ejecutar transformaciones XSL, un lenguaje XSLT que se expresa en forma de hoja de estilo para transformar archivos XML. *xsltproc* es parte de *libxslt*, la biblioteca XSLT C para GNOME, pero puede instalarse de forma independiente.

*xsltproc* se invoca desde el intérprete de comandos o desde un *script* BASH con el nombre de la hoja de estilo que se utilizará para la transformación, seguido del nombre del archivo o archivos a los que se aplicará la hoja de estilo. Las salidas pueden ser formateadas según las necesidades del usuario, usualmente son

---

17 <https://gnome.pages.gitlab.gnome.org/libxslt/xsltproc.html>

formatos HTML, TXT o incluso XML. En este caso, se aplicó una hoja de estilo a todos los registros de metatados en el formato elegido de cosecha, para generar una salida única a archivo TXT, preformateada para ser procesada con los utilitarios CISIS y convertirse en base de datos ISIS.

## **CISIS**

Los utilitarios CISIS son un conjunto de aplicaciones que permiten la manipulación a bajo nivel de bases de datos ISIS (*Integrated Scientific Information System*), creados por BIREME/OPS/OMS a fines de los años 80 bajo el patrocinio de la UNESCO. CDS/ISIS, en sus diferentes versiones (Microisis, Winisis, Genisis), es esencialmente un software administrador de bases de datos, orientado a la recolección, almacenamiento y recuperación de información textual estructurada (Rueda Vildoso, 2006), de amplia difusión y uso en nuestro medio en los últimos 30 años. Si bien actualmente ha venido siendo reemplazado por sistemas más modernos, tales como Koha, aún se utilizan en muchas bibliotecas.

Mediante CISIS, de los cuales existen versiones tanto para DOS como para Linux, es posible realizar operaciones de todo tipo con las bases ISIS desde el intérprete de comandos, tales como búsquedas y generación de archivos invertidos, modificación y agregado de datos, importación y exportación de archivos ISO 2709, además de la creación de bases a partir de archivos de texto, la salida de bases a texto, y la unión de bases de datos, entre otras muchas funciones (BIREME, s.f).

En Memoria Académica se utilizan bases ISIS para los registros bibliográficos de distintos tipos documentales, excepto los libros y otros materiales monográficos, registrados en Koha. Por esta razón, y por su enorme versatilidad, se incorporan los utilitarios CISIS (en versiones DOS y Linux 5.2) a la programación desarrollada en BASH, lo que permite la salida a base ISIS de los registros cosechados y otras tareas básicas del desarrollo como el chequeo de correspondencias, la ingesta de registros a bases ISIS, la generación de archivos MARC de importación, y la impresión de informes de resultados.

## **yaz-marcdump**

Yaz<sup>18</sup> es un conjunto de herramientas para el desarrollo de software centrado en el estándar ANSI Z39.50/ISO23950 para la recuperación de información, que permite el desarrollo y configuración de servidores Z.39.50. Incluye además herramientas de línea de comandos para trabajar con registros MARC, entre los que se encuentra *yaz-marcdump*. A diferencia de otros programas para manejar registros MARC, tales como MarcEdit, *yaz* y sus herramientas son muy flexibles y fáciles de automatizar.

*yaz-marcdump*<sup>19</sup> es una herramienta de conversión de registros MARC. Lee registros MARC de uno o más archivos, los analiza y permite su salida en formato de línea, ISO 2709, MARCXML, MARC-in-JSON, MarcXchange y salida hexadecimal (Hammer et al., 2003)

En nuestro caso, se utilizó para convertir registros de libros, capítulos y datasets en formato de línea a formato MARC (.mrc), para su importación al sistema Koha.

## **YAD - Zenity**

YAD<sup>20</sup> es una herramienta para el desarrollo de interfaces gráficas de usuario que corre en BASH, una bifurcación (*fork*) del programa Zenity creado por Victor Ananjevsky. Zenity<sup>21</sup>, por su parte, es responsabilidad del GNOME Project y tiene la misma finalidad, crear diálogos simples que interactúen gráficamente con el usuario, entre otros, diálogos de entrada de texto, calendarios, barras de proceso, de selección de archivos, listas y formularios.

Se utilizó YAD en la segunda aplicación, destinada a ser ejecutada por el área de Procesos Digitales, del siguiente modo: diálogos de entrada de texto (*entry dialog*), para solicitar datos al usuario; diálogos de lista (*list - radiolist dialog*), para seleccionar tipo de documento; diálogos de texto (*text dialog*), para mostrar referencias en cuadro de diálogo, y cuadros simples de notificaciones (información,

---

18 <https://www.indexdata.com/resources/software/yaz/>

19 <https://software.indexdata.com/yaz/doc/yaz-marcdump.html>

20 <https://github.com/v1cont/yad/>

21 <https://help.gnome.org/users/zenity/stable/index.html.es>

error, advertencia). Para las barras de progreso (*progress dialog*) de ambas aplicaciones, se utilizó Zenity.

## DESARROLLO

El desarrollo se nombró CDMA (Conicet Digital a Memoria Académica) y consiste en dos aplicaciones: **CDMA\_tecno**, destinada a ser ejecutada por el área de Tecnologías, y **CDMA\_procesos**, destinada al área de Procesos Digitales, cada una de ellas con su archivo ejecutable homónimo.

### Diagramas de flujo de datos de los programas desarrollados

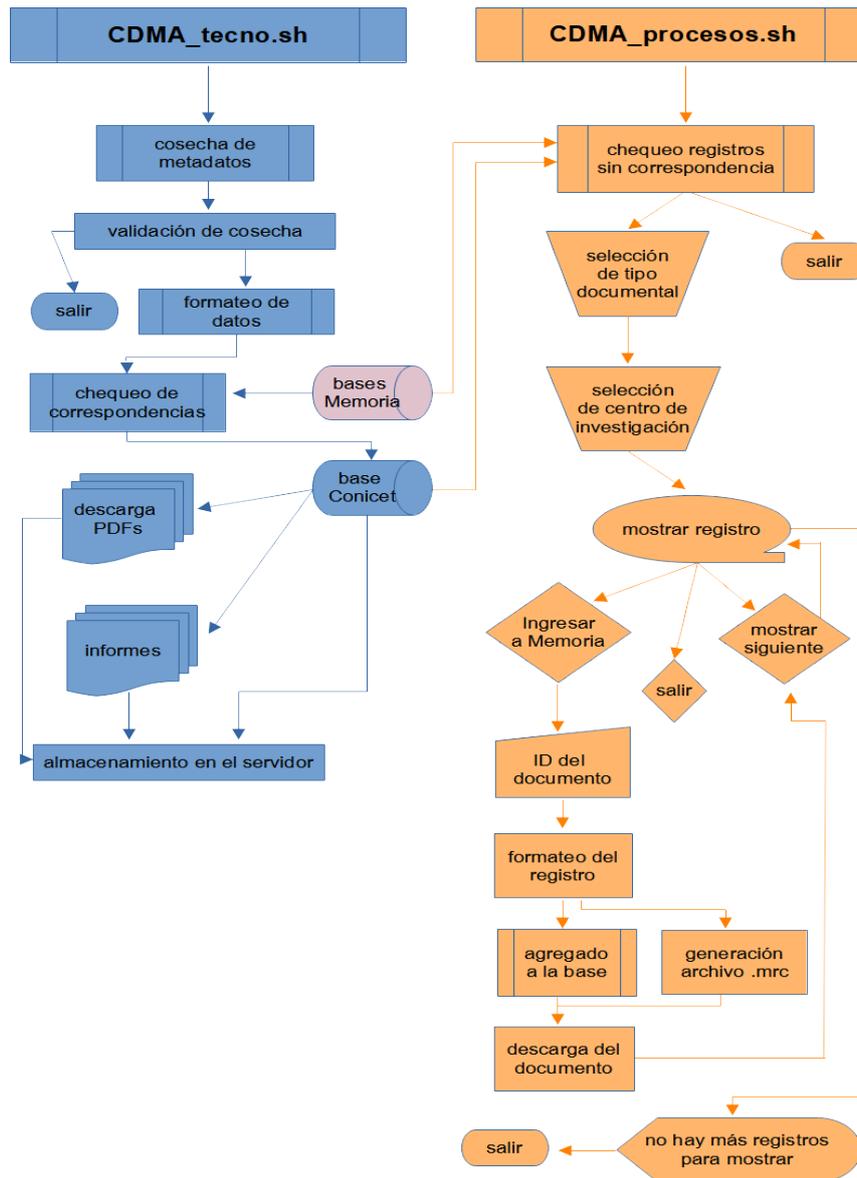


Figura 2. Diagramas de flujo de datos de los programas desarrollados.

La secuencia de procesos del programa **CDMA\_tecno** es el siguiente:

- En primer lugar, elimina registros de la cosecha anterior; cosecha el set IdIHCS de Conicet Digital; comprueba que no haya habido errores; reúne todos los archivos XML cosechados en un sólo archivo XML.
- Luego, si no hubo errores de cosecha, llama a los siguientes scripts:
  - **formateo.sh**: aplica una transformación XSL; crea una base de datos ISIS auxiliar; procesa los campos y subcampos para adecuarlos a los requerimientos de Memoria;
  - **chequeo.sh**: identifica y marca los registros y documentos de CONICET Digital ya existentes en Memoria, en base a URL CONICET Digital, y a correspondencias de tipo de documento, primer autor, título completo y año; marca registros que deban ser ignorados; genera una base de datos final con datos relevantes más datos de control .
- Finalmente, descarga documentos faltantes e imprime informes de resultados, copiando informes, documentos faltantes y base final al servidor de trabajo.

La secuencia de procesos del programa **CDMA\_procesos** es la siguiente:

- En primer lugar, actualiza el chequeo de correspondencias por URL CONICET en registros de Memoria; lee la base final generada por **CDMA\_tecno** y da acceso secuencial a los registros sin correspondencias en Memoria, según tipo documental y unidad de investigación; también es posible procesar todos los recursos de un tipo documental, seleccionando *Todos* los centros.
- En base a elecciones del usuario, ingresa registros y documentos a las bases de Memoria y al servidor de Procesos Digitales, respectivamente, o bien

continúa con el acceso secuencial a los registros, guardando registro de los ingresos hechos en ingresos.log

### **Descripción de CDMA\_tecno.sh**

Permite su ejecución periódica por parte del personal de Tecnologías de Memoria Académica. Ejecuta la cosecha de metadatos del set IDIHCS de CONICET Digital y, previa verificación, aborta el proceso si se presentan errores. Si no hay errores, reúne todos los registros de metadatos en un único archivo XML (ver Fig.3) y llama a los *scripts*:

- **formateo.sh**, para la transformación XSL(ver Fig.4) y la reconversión de datos a una base de datos auxiliar, con el formato estándar de las bases de datos de Memoria y la granularidad de datos necesaria;
- **chequeo.sh**, para el análisis de correspondencias entre registros existentes ya en Memoria Académica y la generación de la base **conicet**, conteniendo todos los registros cosechados con sus correspondencias (ver Fig.5). El análisis de correspondencias incluye la existencia de URLs de CONICET Digital en registros de Memoria, y la correspondencia exacta por tipo de documento, primer autor, título completo y año de publicación. El *script* chequea también los registros y/o documentos que deben ignorarse al momento del procesamiento de datos.

En base a estos chequeos, identifica los registros de Memoria Académica sin texto completo que están presentes en CONICET Digital y en acceso abierto, y descarga los documentos faltantes. Finalmente, imprime informes de resultados y los almacena, junto con la base **conicet** y los documentos faltantes, en el servidor de Procesos Digitales.

Sus archivos de entrada son:

- **xoai.xsl**, el archivo XSLT que analiza **all.xml**, un archivo que reúne todos los registros de metadatos cosechados del set, dispuestos originalmente por el programa cosechador en un conjunto de directorios.

- **gidioma.txt**, **gsubject.txt**, **gtat.txt**: archivos de cambios globales de caracteres para idiomas, temas y uniformidad de caracteres, esto último para la búsqueda de correspondencias por tipo, autor, título y año.
- **bases2linux.bat**, un ejecutable para la conversión de bases ISIS a Linux, a fin de realizar el análisis de correspondencias. En el caso de los registros en Koha, se trabaja con una base auxiliar que reúne los registros MARC21 en formato ISIS.
- **xdocs.txt**, un archivo conteniendo las URL en CONICET Digital de los recursos que deben ignorarse al momento de su inclusión en Memoria Académica, y sus motivos; por ejemplo, PDFs consistentes sólo en tablas de contenido o resumen, que se descartarían porque en Memoria Académica sólo se ingresan documentos a texto completo.
- **.aCDMA\_tecno.sh**, un ejecutable para correr la aplicación desde el escritorio.

Sus salidas son:

- **conicet.mst** y **conicet.xrf**, una base de datos ISIS conteniendo todos los registros reformateados del set IdIHCS, más el marcado de correspondencias según tipo y los registros a ignorar en la segunda aplicación.
- En correspondencias por tipo, autor, título y año, documentos en acceso abierto faltantes en registros de Memoria y descargados para su ingreso.
- Impresión de informes de resultados: 1) listado completo de registros cosechados, con sus correspondencias con Memoria Académica; 2) en correspondencias tipo, autor, título y año, listado de URLs de CONICET Digital e identificadores de Memoria Académica para la posterior incorporación automática de dichas URLs a los registros de Memoria; 3) listado de documentos descargados, faltantes en registros de Memoria Académica.
- Un archivo de alerta de existencia de nuevos tipos documentales en el set IdIHCS, que requerirán una adaptación de los *scripts*.

La base de salida **conicet** es la fuente de información principal del programa **CDMA\_procesos.sh**, a utilizar por el área de Procesos Digitales. Contiene todos los datos identificados como relevantes para el procesamiento de recursos en Memoria Académica, más algunos campos de control, estructurados del siguiente modo:

Campo	Subcampo	Dato relevante
1		Identificador del recurso en CONICET Digital
2		Fecha de cosecha
3		Tipo de documento
4		Tipo de acceso
5	^* tipo documento ^i ID ^d documento	Correspondencia por URL con Memoria Académica
6	^* tipo documento ^i ID ^d documento	Correspondencia por tipo   autor   título   año con Memoria
7		Recursos a ignorar en futuros procesamientos
10		Título
11		Título en inglés
12	^* título ^l lengua	Título en otro idioma
13	^a apellido ^n nombre ^f afiliación	Autores
14	^a apellido ^n nombre	Colaboradores
15		Resumen en castellano
16		Abstract
17	^* resumen * l lengua	Resumen en otro idioma
18		Palabras claves
19		Idioma
20		Versión del documento
21		Notas
22		Editor

23		Título fuente (en capítulos y doc. de conferencia)
24	^r revista ^v volumen ^n número ^f año	Datos de la revista
25	^n nombre ^l lugar ^f fecha	Datos de la reunión
26		Fecha del recurso (defensa en tesis, creación en conjunto de datos, publicación en los restantes casos)
27	^x paginación ^l lugar ^p país	Restantes datos de edición
28	^t tipo ^n número	ISSN / ISBN
29		DOI
30		Recursos relacionados
31	^a organismo ^b ID proyecto	Identificación del proyecto
32		Título otorgado (en tesis)
33		Universidad otorgante (en tesis)
34		Formato
35	^* tipo ^u URL	Licencia CC
36		Fecha de finalización del embargo
37		Acceso al documento

Tabla 2. Definición de la base de datos *conicet*, conteniendo los datos cosechados.



```

mfn=
1
1 "140283"
2 "2024-03-08"
3 "articulo"
4 "openAccess"
10 "Styles of academic production in the Argentine social sciences"iHeterogeneity and heterodoxy"
13 ""aPiovani^nJuan Ignacio^fConsejo Nacional de Investigaciones Cientificas y Tecnicas. Centro Cientifico Tecnológico Conicet - La Plata. Instituto de Investigaciones en Humanidades y Ciencias Sociales. Universidad Nacional de La Plata. Facultad de Humanidades y Ciencias de la Educación. Instituto de Investigaciones en Humanidades y Ciencias Sociales; Argentina"
16 "Argentina is an appealing case for analyzing the social science system. In recent years (until 2015) there has been a robust increase in public funding, giving way to the expansion of research, the recruitment of hundreds of new full-time researchers and the consolidation of scholarships for PhD students. All this, in turn, has resulted in a remarkable increase in publications. Although these processes have occurred in the midst of professionalization (which implies higher levels of adherence to international academic standards), recent studies have shown that two contesting models within the social sciences continue to prevail in Argentina: one that conforms to international standards and practices, and another of a more endogenous nature, with its own logic for knowledge production, evaluation and circulation. In order to examine the impact of international standards in the Argentine social sciences, in this paper I analyze the styles of academic production. This implies the study of three closely related dimensions: research processes and models (theoretical foundations, methods, techniques, etc.); writing formats (structure and organization of academic texts); and publication logics (types of publications, profiles of the journals where the articles are published, etc.). The analysis is based on a large sample of publications selected by other researchers in order to carry out a comprehensive review of Argentine literature with regard to six key themes of the social sciences. These publications were also used to produce a dataset with several variables related to the three above-mentioned dimensions. In particular, this paper focuses on the publications grounded in empirical research, and compares qualitative with quantitative research and the various types of publications (journal articles, book chapters and conference papers). The results show that regardless of the recent process of professionalization, "heterodox models" of academic production are still largely pervasive within the Argentine social sciences."
18 "Argentina"
18 "Ciencias sociales"
18 "Estilos de producción"
19 "en"
20 "Publicado"
22 "Karl Franzens Universität Graz"
24 ""rSerendipities^v4^nl-2^f2020"
26 "2020-12-07"
27 ""x27-48^lGraz^pAustria"
28 ""tISSN^nl0947"
29 "10.7146/serendipities.v4i1.123022"
30 "https://tidsskrift.dk/Serendipities/article/view/123022"
30 "https://ri.conicet.gov.ar/handle/11336/140283"
34 "application/pdf"
35 "2.5 Argentina"uhttps://creativecommons.org/licenses/by-nc-nd/2.5/ar/"
37 "https://ri.conicet.gov.ar/bitstream/11336/140283/2/CONICET_Digital_Nro.9d79e398-b554-4493-9ad0-clb66ee65459_A.pdf"

```

Figura 5. Registro de ejemplo de la base de salida *conicet*.

La documentación con el código fuente de **CDMA\_tecno.sh** se incluye en el Anexo B.

### Descripción de **CDMA\_procesos.sh**

Permite su ejecución periódica por parte del personal del área de Procesos Digitales. Su fuente de información principal es la base **conicet**, conteniendo los resultados de la cosecha y los datos de correspondencias. Trabaja sobre un tipo documental a la vez y opcionalmente por unidad de investigación, seleccionados por el usuario (ver Fig.6 y Fig.7), mostrando secuencialmente los registros a chequear y/o ingresar, y haciendo las operaciones de ingreso de registros a las bases de datos (ver Fig.8 y Fig.9) y de documentos al servidor de trabajo.

En el caso de la base de libros y capítulos almacenada en Koha, descarga un archivo MARC (.mrc) para su ingreso manual en el sistema. Para tener un mejor control sobre la importación de los datos, se optó por la importación manual de cada registro en lugar de una importación de múltiples registros.

Sus archivos de entrada son:

- **conicet.mst** y **conicet.xrf**, los archivos que conforman la base de datos ISIS resultante de la ejecución de **CDMA\_tecno.sh**

- **centros.txt**, un listado de las unidades de investigación del IdHICS para elaborar la interfaz de selección manual por unidad.
- **basesMA.bat**, **koha.sh** y **koha.xsl**, ejecutables y transformadores para el chequeo de correspondencias en las bases de Memoria (ISIS y MySQL) por URL CONICET previo a cada ejecución del programa, a fin de mantener actualizada la presentación de registros a procesar. **basesMA.bat**, además, imprime datos de becarios e investigadores, a fin de cruzar información de pertenencia con los recursos CONICET y así elaborar la interfaz de selección manual por unidad de investigación.
- **arti.prc**, **evento.prc**, **tesis.prc** y **koha.pft**, archivos de conversión de datos para su ingreso a sistema ISIS y Koha.
- **intro.bat**, un ejecutable que agrega los registros indicados a las bases ISIS.
- **.aCDMA\_procesos.sh**, ejecutable para correr la aplicación desde el escritorio.

Sus salidas son:

- nuevos registros bibliográficos agregados en bases ISIS de Memoria.
- archivos MARC (.mrc) para su importación manual a Koha.
- **autores.txt**, para completar autoridades de nombre en registros de libros, capítulos y datasets, ya que no se incluyen en el archivo de importación a Koha.
- Documentos descargados de los recursos ingresados y almacenados en el servidor de trabajo, bajo un identificador de Memoria previamente solicitado.
- **ingresos.log**, un archivo de texto que guarda registro de los ingresos hechos por el programa, conteniendo los siguientes datos: fecha, *hostname*, recurso ingresado a Memoria y descarga o no de texto completo

La documentación con el código fuente de **CDMA\_procesos.sh** se incluye en el Anexo C.

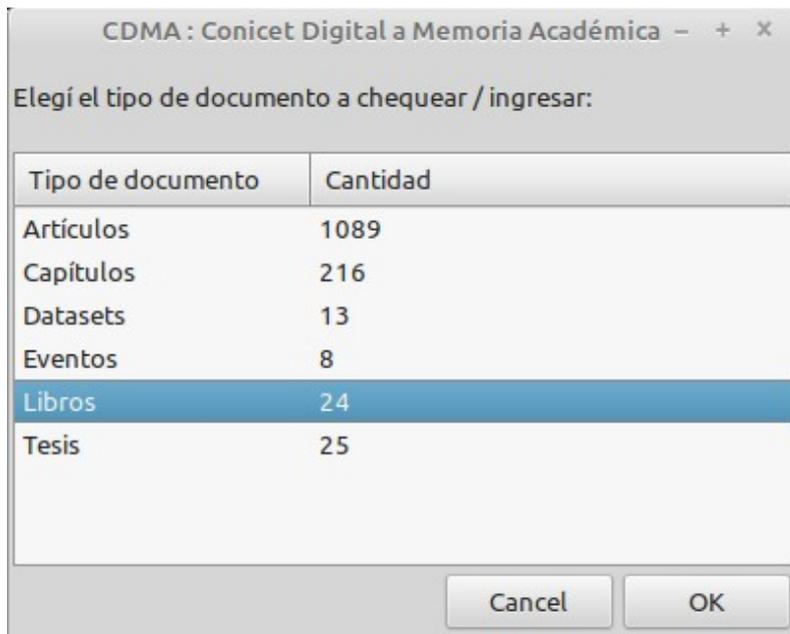


Figura 6. Interfaz de selección de tipo documental.

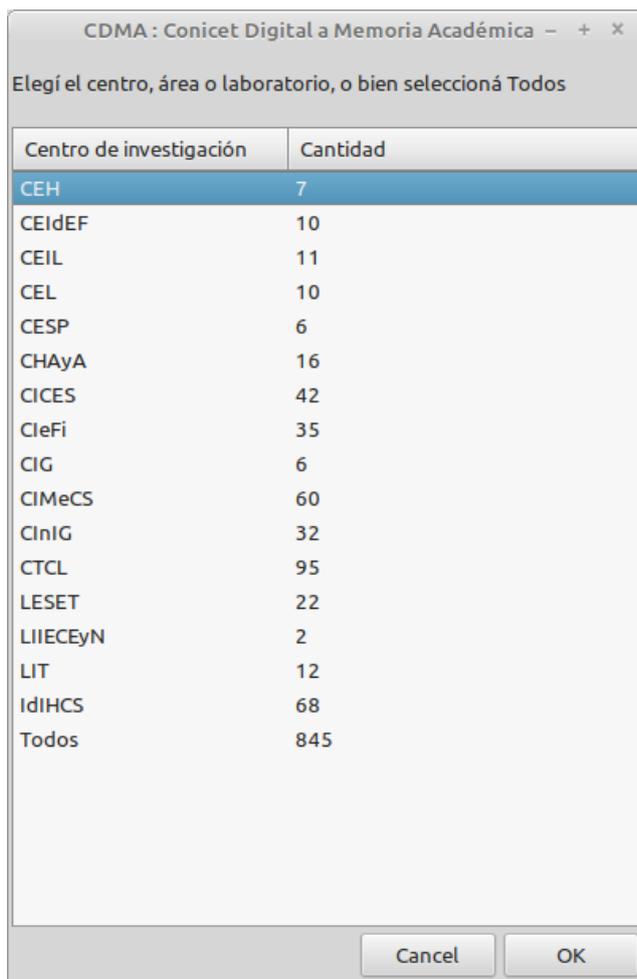


Figura 7. Interfaz de selección de unidad de investigación.

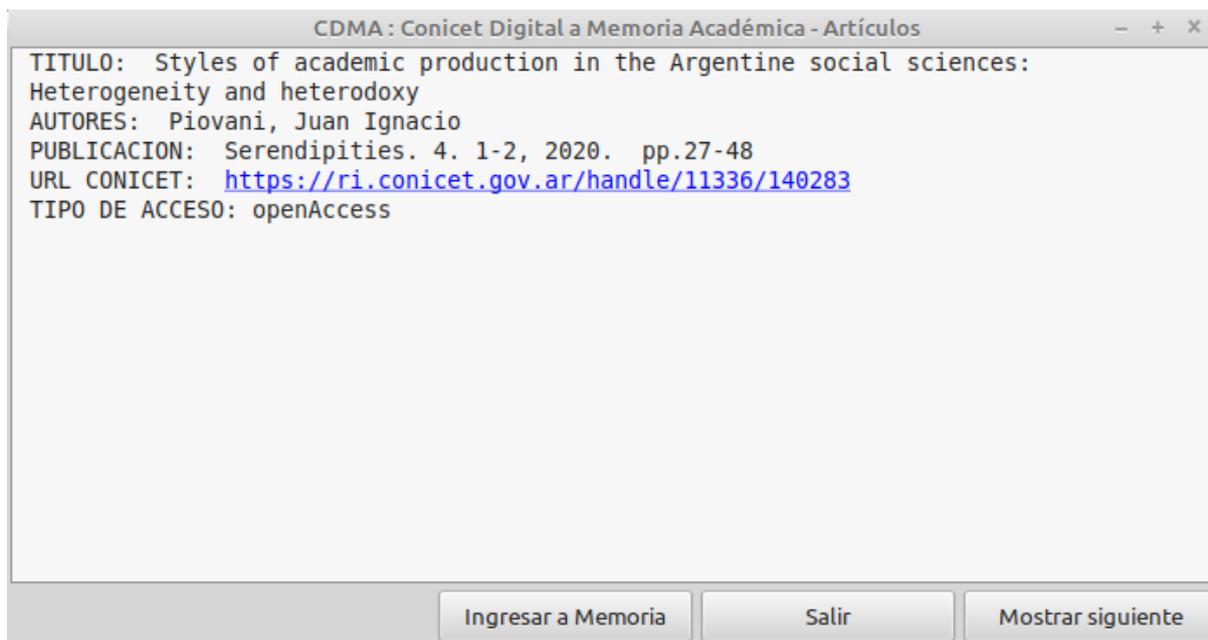


Figura 8. Interfaz de muestra secuencial de recursos.

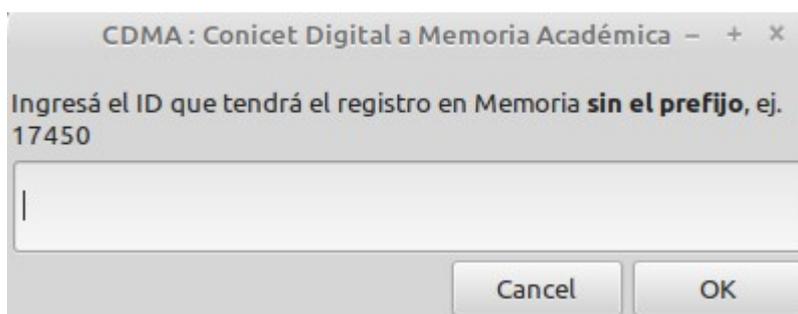


Figura 9. Interfaz de ingreso manual de datos.

### Limitaciones a considerar

Los programas desarrollados tienen algunas limitaciones, a saber:

- Los registros requieren un chequeo manual antes de su ingreso a Memoria Académica, ya que el chequeo de correspondencias por tipo de documento, autor, título y año no es infalible. En una revisión manual, se observaron diferencias menores en los títulos, que no pueden ser recogidas mediante esta metodología de contraste.

- Una vez ingresados los registros, también deben ser chequeados para efectuar correcciones o para completar datos, si es necesario, ya que el formateo no pudo resolver todos los requerimientos. Es el caso de los documentos de conferencia, en los cuales no pudieron separarse u obtenerse los datos de la reunión con la granularidad que requieren los registros de Memoria, así como también el de los libros y capítulos, en los que las autoridades de nombre no pudieron incluirse en el archivo de importación a Koha.
- Puede haber documentos que no se descarguen, es decir, cuya URL no coincida con la provista en el registro de metadatos. En esos casos, se deberá ingresar a la página del recurso en CONICET Digital y descargarlo manualmente; en estos casos, el acceso suele estar a pie de página.

## RESULTADOS OBTENIDOS

Ambos programas se dejaron en funcionamiento en marzo de 2024. **CDMA\_tecno** se corrió por última vez el 7 de marzo de 2024, cosechándose un total de 3043 recursos distribuidos del siguiente modo, según tipo de documento y de acceso:

Tipo documento	Acceso abierto	A. embargado	A. restringido	Acceso cerrado	Total
Artículos	2394	1	101	6	2502
Capítulos de libros	275	0	158	16	449
Datasets	12	1	1	0	14
Doc. de conferencia	11	0	0	0	11
Libros	14	0	13	15	42
Tesis doctorales	16	2	7	0	25
<b>Total cosechado</b>	<b>2722</b>	<b>4</b>	<b>280</b>	<b>37</b>	<b>3043</b>

Tabla 3. Distribución de los recursos cosechados de CONICET Digital según tipo de documento y de acceso.

Del chequeo de correspondencias de lo cosechado con recursos existentes en Memoria Académica, se obtuvo un total de 1468 recursos por chequear / ingresar a Memoria, es decir, algo menos de la mitad de lo cosechado.

Tipo documento	Corresp. por URL	Corresp. por autor título año	PDFs faltantes	Restan procesar
Artículos	359	1026	6	1117
Capítulos de libros	1	168	6	280
Datasets	0	0	0	14
Doc. de conferencia	0	3	0	8
Libros	0	18	0	24
Tesis doctorales	0	0	0	25
<b>Total</b>	<b>360</b>	<b>1215</b>	<b>12</b>	<b>1468</b>

Tabla 4. Distribución de los recursos cosechados de CONICET Digital según tipo de correspondencia con recursos de Memoria Académica.

Por su parte, el programa **CDMA\_procesos** ya comenzó a ser usado por el área de Procesos Digitales en marzo de 2024, mostrando un buen funcionamiento y facilidad de uso. Al momento de la elaboración final de este informe (abril de 2024), ya se han procesado 305 recursos, y se han incorporado a Memoria 157 registros de metadatos y 133 documentos a texto completo.

## **ADAPTACIONES FUTURAS**

En el caso de la base de libros y capítulos almacenada en Koha, el programa **CDMA\_procesos** descarga archivos MARC (.mrc) individuales para su ingreso manual en el sistema. A sugerencia de las autoridades del área, se optó por la importación manual de cada registro en lugar de una importación de múltiples registros (por ejemplo, los marcados para su ingreso en un día o una secuencia de trabajo), para tener mayor control sobre la importación de los datos.

Sin embargo, esto puede adaptarse a futuro si resultara más conveniente importar tandas diarias de registros, en especial en la medida en que se vaya cumpliendo el objetivo de la Biblioteca de migrar todas sus bases bibliográficas al sistema Koha. Se podrá entonces adaptar el programa **CDMA\_procesos** incorporando una importación programada al final del día, utilizando para ello el ejecutable de importación de Koha, de modo tal que los recursos procesados en una jornada sean importados y estén disponibles al día siguiente para que el personal de Procesos Digitales complete su procesamiento.

Una adaptación general de ambos programas al repositorio SEDICI está también contemplada, a fin de automatizar todos los procesos de sincronización de recursos FaHCE/IdIHCS de ambos repositorios, hoy realizados en forma semiautomática.

## BIBLIOGRAFÍA Y FUENTES

Alonso-Arévalo, J., Subirats-Coll, I. y Martínez-Conde, M. (2008). *Informe APEI sobre acceso abierto*. Gijón: APEI, Asociación Profesional de Especialistas en Información.

<http://eprints.rclis.org/12507/1/informeapeiaccesoabierto.pdf>

Argentina. Ministerio de Ciencia, Tecnología e Innovación Productiva. (2011). *Informe sobre la situación de los repositorios digitales argentinos*. Buenos Aires: Subsecretaría de Coordinación Institucional. Secretaría de Articulación Científico-Tecnológica. MinCyT.

[http://www2.biblio.unlp.edu.ar/jubiuna/grupos-de-trabajos/grupo-repositorios-institucionales/sistema-nacional-de-repositorios-digitales-1/InformeRepositoriosFinal\\_25feb2011.pdf](http://www2.biblio.unlp.edu.ar/jubiuna/grupos-de-trabajos/grupo-repositorios-institucionales/sistema-nacional-de-repositorios-digitales-1/InformeRepositoriosFinal_25feb2011.pdf)

Babini, D. y Rovelli, L. (2020). *Tendencias recientes en las políticas científicas de ciencia abierta y acceso abierto en Iberoamérica*. Fundación Carolina; CLACSO. <https://biblioteca-repositorio.clacso.edu.ar/handle/CLACSO/15699>

Bash Reference Manual. <https://www.gnu.org/savannah-checkouts/gnu/bash/manual/bash.html>

Barrueco, J. (2013). RePEc o cómo proveer servicios de valor añadido a una comunidad de investigadores. *Métodos de información (MEI)*, II Época, 4(6): 5-15. DOI: 10.5557/IIMEI4-N6-005015.

Barton, M.R. y Waters, M.W. (2004). *Cómo crear un repositorio institucional: manual LEADIRS II*. MIT Press. <https://docplayer.es/36849-Como-crear-un-repositorio-institucional-manual-leadirs-ii.html>

BIREME/PAHO/WHO. (2005-2007). *CISIS Utilities. Reference Manual Version 5.2*. Sao Paulo: Latin American and Caribbean Center on Health Sciences Information.

<https://abcd-community.org/download/cisis-utilities-reference-manual-version-5-2/>

Bongiovani, P. (2013). *Crecimiento y sostenibilidad de los repositorios institucionales*. Curso de Capacitación Repositorios digitales institucionales: Diseño, implementación y optimización de un recurso estratégico para las Universidades - PICT-O CIN 2010-149. Módulo 5.

Castañeda de León, L. (2004). Interoperabilidad; estándares. *Revista Digital Universitaria*, 2004 5(10).

<https://www.ru.tic.unam.mx/handle/123456789/811>

COAR (2011). *El caso de interoperabilidad para repositorios de acceso abierto*. Grupo de trabajo 2: Interoperabilidad del repositorio. Versión 1.0, julio 2011. Confederation of Open Access Repositories.

[https://www.coar-repositories.org/files/de\\_la\\_investigacion-a-través-de-redes-globales-de-Repositorios-de-Acceso-Abierto-final-version.pdf](https://www.coar-repositories.org/files/de_la_investigacion-a-través-de-redes-globales-de-Repositorios-de-Acceso-Abierto-final-version.pdf)

CONICET (2023). Boletín Nro. 4. Indicadores de la Producción Científico Tecnológica. Oficina Técnica del Repositorio CONICET Digital, Gerencia de Organización y Sistemas, 2023.

<https://ri.conicet.gov.ar/wp/wp-content/uploads/2023/11/Boletin-4-02112023-v2.pdf>

Fushimi, M., Mallo, J., Pené, M., Molfino, R. y Pichinini, M. (2006). *Presentación de la Memoria académica de la Facultad de Humanidades y Ciencias de la Educación de la UNLP*. <https://www.memoria.fahce.unlp.edu.ar/libros/pm.4764/pm.4764.pdf>

Genovés, P. (2017). Perfiles de autor en repositorios institucionales. *Palabra Clave*, 7(1). DOI: 10.24215/18539912e033

Gómez, L. (2007). Interoperabilidad en los Sistemas de Información Documental (SID): la información debe fluir. *Revista Códice*, 3(1): 23-39. <https://www.redalyc.org/pdf/953/95330103.pdf>

Hammer, S., Dickmeiss, A., Taylor, M., Levanto, H. y Schafroth, D. (2003). *YAZ User's Guide and Reference*. Index Data. <https://software.indexdata.com/yaz/doc/yaz.pdf>

IEEE (1990). IEEE Standard Glossary of Software Engineering Terminology. IEEE Std 610.12-1990. Institute of Electrical and Electronics Engineers.

Lagoze, C.; Van de Sompel, H. (2008). *The Open Archives Initiative Protocol for Metadata Harvesting*. Version 2.0. <http://www.openarchives.org/OAI/openarchivesprotocol.html>

Lapiedra Alcamí, R., Forés Julián, B., Puig Denia, A. y Martínez Cháfer, L. (2021). Introducción a la gestión de sistemas de información en las empresas. Universitat Jaume I. DOI: 10.6035/Sapientia178

Ley 26.899 de 2013, Repositorios digitales institucionales de acceso abierto. *Boletín Oficial de la República Argentina*, 32.781, Primera Sección, 3-4. <http://servicios.infoleg.gob.ar/infolegInternet/anexos/220000-224999/223459/norma.htm>

Mallo, J., Fushimi, M., Pichinini, M. y Pené, M. (2005). *Memoria Académica de la Facultad de Humanidades y Ciencias de la Educación de la UNLP*. Universidad Nacional de La Plata. Facultad de Humanidades y Ciencias de la Educación. Biblioteca de Humanidades Profesor Guillermo Obiols. <https://www.memoria.fahce.unlp.edu.ar/libros/pm.4751/pm.4751.pdf>

Míguez, J. (2021) La interoperabilidad de datos como desafío social y político. *XV Simposio de Informática en el Estado (SIE 2021) - JAIIO 50 (Modalidad virtual)*. Sociedad Argentina de Informática e Investigación Operativa. pp. 206-212. <https://50jaiio.sadio.org.ar/pdfs/sie/SIE-18.pdf>

Moreiro González, J. (1995). Los sistemas de información documental: consideraciones sobre sus características, concepto y funciones. *Boletín Millares Carlo*, 1995(14): 137-150.

Naranjo Vélez, E. (2014). Didáctica de los sistemas de información documental: Transformación de la información en conocimiento. UNAM. Instituto de Investigaciones Bibliotecológicas y de la Información. [https://ru.iibi.unam.mx/jspui/handle/IIBI\\_UNAM/L73](https://ru.iibi.unam.mx/jspui/handle/IIBI_UNAM/L73)

OAI. (2002). *The Open Archives Initiative Protocol for Metadata Harvesting Protocol*. Version 2.0 of 2002-06-14. <http://www.openarchives.org/OAI/openarchivesprotocol.html>

Ponjuán, G. (2004). *Sistemas de información: Principios y Aplicación*. Félix Varela.

Resolución Nro. 527/2021. *Política de acceso abierto al conocimiento científico de la Facultad de Humanidades y Ciencias de la Educación y el Instituto de Investigaciones en Humanidades y Ciencias Sociales (FaHCE-IdIHCS) de la Universidad Nacional de La Plata*. Universidad Nacional de La Plata. Facultad de Humanidades y Ciencias de la Educación. Consejo Directivo, sesión 31-03-2021. <https://www.memoria.fahce.unlp.edu.ar/normativa/no.154/no.154.pdf>

Resolución 753 de 2016. Ministerio de Ciencia, Tecnología e Innovación Productiva. *Reglamentación de la Ley 26.899 sobre Creación de Repositorios Digitales Institucionales de Acceso Abierto*. <http://www.argentina.gob.ar/normativa/nacional/resoluci%C3%B3n-753-2016-267833/texto>

Rueda Vildoso, H. (2006). *CDS/ISIS de UNESCO*. 10 de diciembre 2006. <https://hrueda-isis.blogspot.com/2006/>

Senn, J. (1992). *Análisis y diseño de sistemas de información*. McGraw-Hill.

SNRD. (2015). *Directrices para proveedores de contenido del Sistema Nacional de Repositorios Digitales*. Sistema Nacional de Repositorios Digitales. Ministerio de Ciencia, Tecnología e Innovación.

Suber, P. (2015). *Acceso abierto*. Traducción de Remedios Melero. Toluca: Universidad Autónoma del Estado de México. <http://ri.uaemex.mx/handle/20.500.11799/21710>

Tackett, J., Gunter, D. y Brown, L. (1995). *Linux*. Prentice-Hall Hispanoamericana.

Testa, P. y Degiorgi, E.H. (2013). *Esquemas de metadatos para los repositorios institucionales de las universidades nacionales argentinas*. Tesis de Licenciatura en Documentación y Gestión de la Información, Facultad de Educación Elemental y Especial, Universidad Nacional de Cuyo. [http://bdigital.uncu.edu.ar/objetos\\_digitales/5881/tesisdegiorgitesta.pdf](http://bdigital.uncu.edu.ar/objetos_digitales/5881/tesisdegiorgitesta.pdf)

Valentim, M. (2017). Gestión de la información: dilemas y perspectivas. *Palabra Clave*, 7(1). DOI: 10.24215/18539912e026

## FUENTES

CONICET Digital. <https://www.conicet.gov.ar/conicet-digital/>

CONICET en cifras. <https://cifras.conicet.gov.ar/publica/>

Instituto de Investigaciones en Humanidades y Ciencias Sociales (CONICET-UNLP).  
<https://idihcs.fahce.unlp.edu.ar/>

LA Referencia. <https://www.lareferencia.info/es/>

Memoria Académica. <https://www.memoria.fahce.unlp.edu.ar>

OpenDOAR. <https://v2.sherpa.ac.uk/opensoar/>

Registry of Open Access Repositories. <http://roar.eprints.org/>

Sistema Nacional de Repositorios Digitales. <https://repositoriosdigitales.mincyt.gob.ar/vufind/>

## ANEXO A. Datos y metadatos relevantes según formato de cosecha

Tipo de dato / Formato	oai_dc	didl	dim	etdms	marc	mets
Identificador del recurso	<dc:identifier>	<doi:Identifier>	<dim:field mdschema="dc" element="identifier" qualifier="uri">	<identifier>	<datafield tag="024" ind2=" " ind1="8" ><subfield code="a">	<mods:identifier type="uri">
Tipo de documento	<dc:type>	<dc:type>	<dim:field mdschema="dc" element="type">	<type>	<leader>	<mods:genre>
Versión del documento	<dc:type>	<dc:type>	<dim:field mdschema="dc" element="type">	<type>		<mods:genre>
Idioma	<dc:language>	<dc:language>	<dim:field mdschema="dc" element="language">	<language>		<mods:language > <mods:language Term authority="rfc306 6">
Título	<dc:title>	<dc:title>	<dim:field mdschema="dc" element="title">	<title>	<datafield tag="245" ind1="0" ind2="0" ><subfield code="a">	<mods:titleInfo > <mods:title>
Autor	<dc:creator>	<dc:creator>	<dim:field mdschema="dc" element="contributor" qualifier="author">	<creator>	<datafield tag="720" ind1=" " ind2=" " ><subfield code="a" ><subfield code="e"=author >	<mods:roleTerm type="text">author > <mods:namePart >
Filiación del autor	<dc:description>		<dim:field mdschema="dc" element="description" qualifier="fil">			
Colaborador	<dc:contributor>	<dc:contributor>	<dim:field mdschema="dc" element="contributor" qualifier="other">	<contributor>		<mods:role> <mods:roleTerm type="text">other > <mods:namePart >
Fecha de publicación	<dc:date>	<dc:date>	<dim:field mdschema="dc" element="date" qualifier="issued">	<date>	<datafield tag="260" ind1=" " ind2=" " ><subfield code="c">	<mods:originInfo > <mods:dateIssued encoding="iso86 01">
Fecha de embargo	<dc:date>		<dim:field mdschema="dc" element="rights" qualifier="embargoDate">	<rights>		
Título de revista	<dc:identifier>	<dc:identifier>	<dim:field mdschema="dc" element="identifier" qualifier="citation">	<identifier>	<datafield tag="024" ind2=" " ind1="8" ><subfield code="a">	<mods:identifier type="citation">
Designación de la revista	<dc:identifier>	<dc:identifier>	<dim:field mdschema="dc" element="identifier" qualifier="citation">	<identifier>	<datafield tag="024" ind2=" " ind1="8" ><subfield code="a">	<mods:identifier type="citation">
Nombre de la reunión	<dc:identifier>	<dc:identifier>	<dim:field mdschema="dc" element="identifier" qualifier="citation">	<identifier>	<datafield tag="024" ind2=" " ind1="8" ><subfield code="a">	<mods:identifier type="citation">
Datos de la reunión	<dc:identifier>	<dc:identifier>	<dim:field mdschema="dc" element="identifier" qualifier="citation">	<identifier>	<datafield tag="024" ind2=" " ind1="8" ><subfield code="a">	<mods:identifier type="citation">
Título otorgado (tesis)						
Universidad otorgante						
Editor	<dc:publisher>	<dc:publisher>	<dim:field mdschema="dc" element="publisher">	<publisher>		

Tipo de dato / Formato	oai_dc	didl	dim	etdms	marc	mets
Fuente		<dc:source>				
Paginación	<dc:identifier>	<dc:identifier>	<dim:field mdschema="dc" element="identifier" qualifier="citation">	<identifier>	<datafield tag="024" ind2=" " ind1="8"><subfield code="a">	<mods:identifier type="citation">
Resumen	<dc:description>	<dc:description>	<dim:field mdschema="dc" element="description" qualifier="abstract">	<description>	<datafield tag="520" ind1=" " ind2=" " "><subfield code="a">	<mods:abstract>
Temas	<dc:subject>	<dc:subject>	<dim:field mdschema="dc" element="subject">	<subject>	<datafield ind1=" " ind2=" " " tag="653"><subfield code="a">	<mods:subject> <mods:topic>
ISSN/ISBN	<dc:identifier>	<dc:identifier>	<dim:field mdschema="dc" element="identifier" qualifier="isbn">	<identifier>	<datafield tag="024" ind2=" " ind1="8"><subfield code="a">	<mods:identifier type="issn"/> / <mods:identifier type="isbn">
DOI	<dc:relation>	<dc:relation>	<dim:field mdschema="dc" element="relation" qualifier="alternativeid"/> / <dim:field mdschema="dc" element="relation" qualifier="alternativeid">	<relation>		
URLs alternativas	<dc:relation>	<dc:relation>	<dim:field mdschema="dc" element="relation" qualifier="alternativeid">	<relation>		
Formato	<dc:format>		<dim:field mdschema="dc" element="format">			<premis:formatName>
Tipo de acceso	<dc:rights>	<dc:rights>	<dim:field mdschema="dc" element="rights">	<rights>		<mods:accessCondition type="useAndReproduction">
Licencia CC	<dc:rights>	<dc:rights>	<dim:field mdschema="dc" element="rights" qualifier="uri">	<rights>		<mods:accessCondition type="useAndReproduction">
Acceso al recurso		<d:Resource>				<premis:objectIdentifierValue>

Tipo de dato / Formato	mods	ore	qdc	rdf	uketd_dc	xoai
Identificador del recurso	<mods:identifier type="uri">	<atom:id>	<dc:identifier>	<dc:identifier>	<dcterms:isReferencedBy>	<element name="others"> <field name="handle">
Tipo de documento	<mods:genre>		<dc:type>	<dc:type>	<dc:type>	<element name="dc"> <element name="type"> <element name="none"> <field name="value">
Versión del documento	<mods:genre>		<dc:type>	<dc:type>	<dc:type>	<element name="dc"> <element name="type"> <element name="none"> <field name="value">
Idioma	<mods:language > <mods:language Term>		<dc:language>	<dc:language>	<dc:language>	<element name="dc"> <element name="language"> <element name="iso"> <element name="es"> <field name="value">
Título	<mods:titleInfo> <mods:title>	<atom:title>	<dc:title>	<dc:title>	<dc:title>	<element name="dc"> <element name="title"> <element name="es"> <field name="value">
Autor	<mods:name> <mods:namePart >	<atom:author> <atom:name>	<dc:creator>	<dc:creator>	<dc:creator>	<element name="dc"> <element name="contributor"> <element name="author"> <element name="none"> <field name="value">
Filiación del autor						<element name="dc"> <element name="fil"> <element name="none"> <field name="value">
Colaborador			<dc:contributor>	<dc:contributor>		<element name="dc"> <element name="contributor"> <element name="other"> <element name="none"> <field name="value">
Fecha de publicación	<mods:originInfo > <mods:dateIssued encoding="iso8601">		<dcterms:issued>	<dc:date>	<dcterms:issued >	<element name="dc"> <element name="date"> <element name="issued"> <element name="none"> <field name="value">
Fecha de embargo						<element name="dc"> <element name="embargo Date"> <element name="*"> <field name="value">

Tipo de dato / Formato	mods	ore	qdc	rdf	uketd_dc	xoai
<b>Título de revista</b>	<mods:identifier type="citation">		<dc:identifier>	<dc:identifier>		<element name="dc"> <element name="journal"> <element name="title"> <element name="es"> <field name="value">
<b>Designación de la revista</b>	<mods:identifier type="citation">		<dc:identifier>	<dc:identifier>		<element name="dc"> <element name="journal"> <element name="volume"> <element name="es"> <field name="value"> / // <element name="number"> > <element name="es"> <field name="value"> <element name="es"> <field name="value">
<b>Nombre de la reunión</b>	<mods:identifier type="citation">		<dc:identifier>	<dc:identifier>		<element name="dc"> <element name="identifier"> > <element name="citation"> <element name="es"> <field name="value">
<b>Datos de la reunión</b>	<mods:identifier type="citation">		<dc:identifier>	<dc:identifier>		<element name="dc"> <element name="identifier"> > <element name="citation"> <element name="es"> <field name="value">
<b>Título otorgado (tesis)</b>						<element name="conicet"> <element name="titulo"> <element name="es"> <field name="value">
<b>Universidad otorgante</b>						<element name="conicet"> <element name="otorgante"> > <element name="es"> <field name="value">
<b>Editor</b>			<dc:publisher>	<dc:publisher>		<element name="dc"> <element name="publisher"> > <element name="es"> <field name="value">

Tipo de dato / Formato	mods	ore	qdc	rdf	uketd_dc	xoai
<b>Fuente</b>			<dc:source>	<dc:source>		<element name="dc"> <element name="source"> <element name="titulo"> <element name="es"> <field name="value">
<b>Paginación</b>	<mods:identifier type="citation">		<dc:identifier>	<dc:identifier>		<element name="dc"> <element name="journal"> <element name="paginatio n"> <element name="es"> <field name="value">
<b>Resumen</b>	<mods:abstract>		<dcterms:abstract>	<dc:description>	<dcterms:abstract>	<element name="dc"> <element name="description"> <element name="abstract"> <element name="es"> <field name="value">
<b>Temas</b>	<mods:subject> <mods:topic>		<dc:subject>	<dc:subject>	<dc:subject>	<element name="dc"> <element name="subject"> <element name="es"> <field name="value">
<b>ISSN/ISBN</b>	<mods:identifier type="issn"> / <mods:identifier type="isbn">		<dc:identifier>	<dc:identifier>		<element name="dc"> <element name="issn"> <element name="es"> <field name="value"> / // <element name="eissn"> <element name="es"> <field name="value"> / // <element name="isbn"> <element name="es"> <field name="value">
<b>DOI</b>			<dc:relation>	<dc:relation>		<element name="dc"> <element name="relation"> <element name="alternativ eid"> <element name="none"> <field name="value">
<b>URLs alternativas</b>			<dc:relation>	<dc:relation>		<element name="dc"> <element name="relation"> <element name="alternativ eid"> <element name="none"> <field name="value">

Tipo de dato / Formato	mods	ore	qdc	rdf	uketd_dc	xoai
Formato		<atom:link type=				<element name="dc"> <element name="format"> <element name="es"> <field name="value">
Tipo de acceso	<mods:accessCondition type="useAndReproduction">		<dc:rights>	<dc:rights>	<dc:rights>	<element name="dc"> <element name="rights"> <element name="*"> <field name="value">
Licencia CC	<mods:accessCondition type="useAndReproduction">		<dc:rights>	<dc:rights>	<dc:rights>	<element name="dc"> <element name="uri"> <element name="*"> <field name="value">
Acceso al recurso		<atom:link href=			<dc:identifier xsi:type="dcterms:URI">	<element name="bundles"> > <element name="bundle"> <field name="name">ORIGINAL</field> <element name="bitstreams"> <element name="bitstream"> <field name="name">

## Anexo B. Código fuente de CDMA\_tecno

```
CDMA_tecno.sh x
1 #!/bin/bash
2 # CDMA_TECNO: cosecha el set IdIHCS de Conicet Digital, transforma a base Isis, chequea correspondencias con Memoria, descarga PDFs faltantes e imprime informes. El proceso de cosecha elimina registros de cosecha anterior; realiza una nueva cosecha; comprueba que no haya habido errores; reúne todos los archivos XML cosechados en un sólo archivo XML.
3 # Llama a los siguientes scripts:
4 # formateo.sh: aplica una transformación XSL; crea una base de datos ISIS auxiliar; procesa los campos y subcampos para adecuarlos a los requerimientos de Memoria.
5 # chequeo.sh : marca los registros y documentos de CONICET Digital ya existentes en Memoria; descarga los PDFs en abierto faltantes en Memoria; comprueba la existencia de nuevos tipos documentales.
6 # Patricia Genovés, marzo/2024.
7
8 COSECHA="/home/pgenoves/cosechas/conicet_xoai"
9 OAITMP="/home/pgenoves/tmp/oaiharvester"
10 DT="/home/pgenoves/CDMA_tecno"
11 PROCESO="/home/pgenoves/CDMA_tecno/proceso"
12 SALIDAS="/home/pgenoves/CDMA_tecno/salidas"
13 ALMACEN="/SERVIDOR/procesamiento/CDMA"
14 FECHA=`date +%Y-%m-%d`
15
16 function cancelar() {
17     if [ $1 -eq 1 ]; then
18         exit
19     fi
20 }
21
22 cd ${DT}
23
24 # eliminación de archivos de cosecha anterior
25 rm -rf ${COSECHA}/{*.*}
26 rm -r ${OAITMP}/*
27
28 # cosecha xoai
29 oaiharvester -r conicet_xoai | zenity --title="CDMA_tecno" --text="Cosechando set IdIHCS de Conicet Digital ..." --progress --pulsate --auto-close --width=600 --height=100
30     cancelar $?
31 # verificación cosecha
32 find ${OAITMP} -type d | sed -e 'ld' >error.txt
33 if [ -s /error.txt ]
34 then
35     zenity --info --title=" Error de cosecha " --width="100" --text=" La cosecha dió error. Volver a intentar. "
36 else
37     exit
38 # reunión de archivos xml y llamada a los siguientes scripts
39
40 touch ${COSECHA}/all.xml
41 for i in $(ls ${COSECHA}/*/*); do cat $i >>${COSECHA}/all.xml; done
42 cp ${COSECHA}/all.xml ${PROCESO}/all_xoai.xml
43 # llamada a formateo.sh
44 bash formateo.sh | zenity --title="CDMA_tecno" --text="Formateando datos ..." --progress --pulsate --auto-close --width=600 --height=100
45 cancelar $?
46 # llamada a chequeo.sh
47 bash chequeo.sh | zenity --title="CDMA_tecno" --text="Chequeando correspondencias ..." --progress --pulsate --auto-close --width=600 --height=100
48 cancelar $?
49 fi
50 # Impresión de informes
51 cd ${SALIDAS}
52
53 # Listado de lo cosechado y correspondencias con registros de Memoria
54 echo "LISTADO DE LO COSECHADO - ${FECHA}" >${SALIDAS}/${FECHA}_informe_1.txt"
55 echo "MFN|URL Conicet|Tipo de documento|Formato|Acceso|Correspondencia por URL|Correspondencia por Tipo-Autor-Titulo|A ignorar" >>${SALIDAS}/${FECHA}_informe_1.txt"
56 mx conicet lw=10000 "pft=mfn(5)|https://ri.conicet.gov.ar/handle/11336/'v1','v3','|',(v34+| - |),'|',v4'|',v5'|',(v6+| - |),'|',v7/'" >>${SALIDAS}/${FECHA}_informe_1.txt" now
57 # Listado de correspondencias Tipo-Autor-Titulo (para inclusión posterior de URL Conicet)
58 echo "LISTADO DE CORRESPONDENCIAS TIPO-AUTOR-TITULO - ${FECHA}" >${SALIDAS}/${FECHA}_informe_2.txt"
59 echo "ID en Memoria|URL Conicet" >>${SALIDAS}/${FECHA}_informe_2.txt"
60 mx conicet lw=10000 "pft=if p(v6) then (v6^i|https://ri.conicet.gov.ar/handle/11336/'v1[1]/),fi" >>${SALIDAS}/${FECHA}_informe_2.txt" now
61 # Listado de documentos abiertos en Conicet y restringidos o faltantes en Memoria
62 echo "LISTADO DE DOCUMENTOS ABIERTOS EN CONICET Y RESTRINGIDOS O FALTANTES EN MEMORIA - ${FECHA}" >${SALIDAS}/${FECHA}_informe_3.txt"
63 echo "ID en Memoria|URL Conicet" >>${SALIDAS}/${FECHA}_informe_3.txt"
64 mx conicet lw=10000 "pft=if p(v6) then ,(if v6:'REST' or a(v6^d) then ,if v4[1]:'open' then v6^i|https://ri.conicet.gov.ar/handle/11336/'v1[1],fi,fi,/,),fi" >>${SALIDAS}/${FECHA}_informe_3.txt" now
65
66 # Almacenamiento en servidor de procesos técnicos
67 cp "conicet.mst" "${ALMACEN}/datos_cosecha/conicet.mst"
68 cp "conicet.xrf" "${ALMACEN}/datos_cosecha/conicet.xrf"
69 cp "${SALIDAS}/${FECHA}_informe_1.txt" "${ALMACEN}/datos_cosecha/${FECHA}_informe_1.txt"
70 cp "${SALIDAS}/${FECHA}_informe_2.txt" "${ALMACEN}/datos_cosecha/${FECHA}_informe_2.txt"
71 cp "${SALIDAS}/${FECHA}_informe_3.txt" "${ALMACEN}/datos_cosecha/${FECHA}_informe_3.txt"
72
73 zenity --info --title=" CDMA_tecno " --width="100" --text=" Todos los procesos han terminado "
74
```



```

65 mx bx2 "proc=if p(v104) then ,if v104[1]:' che ' then 'a17#v104[1]'^lit# ' fi,fi" "proc=if p(v104) then ,if v104[2]:' che '
  then 'a17#v104[2]'^lit# ' fi,fi" "proc=if p(v104) then ,if v104[3]:' che ' then 'a17#v104[3]'^lit# ' fi,fi" copy=bx2 -a
  ll now
66 mx bx2 "proc=if p(v104) then ,if v104[1]:' er ' then 'a17#v104[1]'^lde# ' fi,fi" "proc=if p(v104) then ,if v104[2]:' er '
  then 'a17#v104[2]'^lde# ' fi,fi" "proc=if p(v104) then ,if v104[3]:' er ' then 'a17#v104[3]'^lde# ' fi,fi" copy=bx2 -all n
  ow
67 # v18 <subject> = palabras claves
68 mx bx2 "pft=if p(v119) then (v119|'mfn/),fi now |tr '[:upper:]' '[:lower:]' >subject.txt
69 mx seq=subject.txt create=subject -all now
70 mx seq=gsubject.txt create=gsubject -all now
71 mx subject gizmo=gsubject,1 copy=subject -all now
72 mx subject "proc='d1a1#mhu,v1.1,mhl,v1*1#'" copy=subject -all now
73 mx subject "fst=1 0 v2/" fullinv=subject -all now
74 mx bx2 "join=subject,18:1=mhu,mfn/" create=bx3 -all now
75 mx bx3 "proc='d32001'" copy=bx3 -all now
76 mx bx3 "proc=if p(v18) then 'd18',(if v18: '/' then '' else 'a18#v18#' fi) fi" copy=bx3 -all now
77 # v19 <language> = idioma (ARMAR GIZMO) - gidionma.txt puede contener todos los codigos
78 mx bx3 "proc=if p(v111) then 'a19#v111#' fi)," create=bx4 -all now
79 mx seq=gidioma.txt create=gidioma -all now
80 mx bx4 gizmo=gidioma,19 copy=bx4 -all now
81 # v20 <type> = version del documento
82 mx bx4 "proc=if p(v106) then ,(if v106:'Version' then 'a20#v106#' fi),fi" copy=bx4 -all now
83 mx bx4 "proc=if p(v20) then 'd20a20#',if v20:'published' then 'Publicado' else ,if v20:'Accepted' then 'Aceptado' else ,i-
  f v20:'Up' then 'Actualizado' fi,fi,fi,'#',fi" copy=bx4 -all now
84 mx bx4 "proc=if v3:'dataset' then 'a20#Inédito#' fi" copy=bx4 -all now
85 # v21 = notas (versión del dataset)
86 mx bx4 "proc=if p(v112) then 'a21#Versión 'v112#' fi" copy=bx4 -all now
87 # v22 <publisher> = editor/organizador
88 mx bx4 "proc=if p(v113) then 'a22#v113^*#' fi" copy=bx4 -all now
89 # v23 <source> = titulo fuente principal
90 mx bx4 "proc=if p(v114) then 'a23#v114^*#' fi" copy=bx4 -all now
91 # v24 <journal> = datos revista
92 mx bx4 "proc=if p(v115) then 'a24#v115#' fi" copy=bx4 -all now
93 mx bx4 "proc=if v3:'articulo' then 'd24a24#v24^f/v105.4#' fi" copy=bx4 -all now
94 # v25 <identifiaer> = datos de reunión
95 mx bx4 "proc=if v3:'conferencia' then 'a25#v123#' fi" create=bx5 -all now
96 mx bx5 "proc=if p(v25) then 'd25a25#replace(v25,',' ,'%a')#' fi" copy=bx5 -all now
97 mx bx5 lw=10000 "pft=if p(v10) then (v13^a, ' ,v13^n'|/),'et al.'|/,'v10'|/,'fi" >gfuente.txt now

98 mx bx5 lw=10000 "pft=if p(v10) then ,if size(v105)=10 then v105*8.2,'-',v105*5.2,'-',v105.4'|/,'v105*9.1,'-',v105*5.2,'-',v-
  105.4'|/,'v105*8.2,'-',v105*6.1,'-',v105.4'|/,'v105*9.1,'-',v105*6.1,'-',v105.4'|/,'fi,if size(v105)=7 then v105*5.2,'-',v1-
  05.4'|/,'v105*6.1,'-',v105.4'|/,'fi,fi" >>gfuente.txt now
99 mx seq=gfuente.txt create=gfuente -all now
100 mx bx5 gizmo=gfuente,25 copy=bx5 -all now
101 mxcp bx5 create=bx6 repeat=%,25
102 mx bx6 "proc=if p(v25) then 'd25a25#n',if v25[1].2:'^a' then v25[1]^a,'^l^v25[2]^a,if p(v25[5]) then ', ,v25[3]^a,'^f^v-
  25[4]^a else '^f^v25[3]^a,fi, else v25[2]^a,'^l^v25[3]^a,if p(v25[6]) then ', ,v25[4]^a,'^f^v25[5]^a else '^f^v25[4]^a,f-
  i,fi,'#' fi" copy=bx6 -all now
103 # v26 <dc:date> = fecha del recurso / fecha de publicación
104 mx bx6 "proc=if p(v105) then 'a26#v105#' fi" copy=bx6 -all now
105 # v27 <journal.pagination> <journal.ciudad> <journal.país> = paginación/ciudad/país
106 mx bx6 "proc=if p(v116) then 'a27#v116#' fi" copy=bx6 -all now
107 # v28 <identifiaer> = issn, eissn, isbn
108 mx bx6 "proc=if p(v108) then ('a28#replace(v108,' - ','-'),'#') fi" copy=bx6 -all now
109 mx bx6 "proc=if p(v109) then ('a28#replace(v109,' ','-'),'#') fi" copy=bx6 -all now
110 # v29 <relation> = DOI
111 mx bx6 "proc=(if v120:'altIdentifiaer/doi/' or v120:'doi.org' or v120:'altIdentifiaer/url/10.' then 'a29#replace(v120,' ','-
  '),'#' fi)" copy=bx6 -all now
112 mx bx6 "proc='d29',(if p(v29) then 'a29#replace(v29,'info:eu-repo/semantics/altIdentifiaer/doi/',),' ,'#' fi)" copy=bx6 -al-
  l now
113 mx bx6 "proc='d29',(if p(v29) then 'a29#replace(replace(replace(replace(replace(replace(v29,'info:eu-rep-
  o/semantics/altIdentifiaer/url/',),' ','https','http'),'http://doi.org/',),' ','http://dx.doi.org/',),' ,DOI:',),'http://www-
  .dx.doi.org/',),' ','http://doi.org/',),' ,www.dx.doi.org/',),' ,'#' fi)" copy=bx6 -all now
114 # v30 <relation> = otras URLs externas
115 mx bx6 "proc=(if v120:'altIdentifiaer/doi/' or v120:'doi.org' then '' else ,if v120:'altIdentifiaer/url/' then 'a30#v120#'-
  fi,fi)" copy=bx6 -all now
116 mx bx6 "proc='d30',(if p(v30) then 'a30#replace(v30,'info:eu-repo/semantics/altIdentifiaer/url/',),' ,'#' fi)" copy=bx6 -al-
  l now
117 mx bx6 "proc='d30',(if p(v30) then 'a30#replace(replace(v30,'www.http','http'),'chrome-extension://efaidnbmninnbpcjagcl-
  clefindmkaj/',),' ,'#' fi)," copy=bx6 -all now
118 mx bx6 "proc=('a30#https://ri.conicet.gov.ar/handle/11336/'v1[1]#)" copy=bx6 -all now
119 # v31 <FunderName> <FundingReference> = financiamiento
120 mx bx6 "proc=if p(v121) then 'a31#^a^v121,if p(v122) then '^b^v122[1],fi,'#' fi" copy=bx6 -all now
121 # v32 <titulo> = titulo académico otorgado (tesis)
122 mx bx6 "proc=if p(v124) then 'a32#v124#' fi" copy=bx6 -all now
123 # v33 <otorgante> = universidad otorgante (tesis)
124 mx bx6 "proc=if p(v125) then 'a33#v125#' fi" copy=bx6 -all now
125

126 # C3. Datos del documento
127 # v34 <format> = formato
128 mx bx6 "proc=if p(v110) then ('a34#v110#' fi) copy=bx6 -all now
129 # v35 <dc:rights> licencia CC
130 mx bx6 "proc=if p(v117^u) then 'a35#v117^u#' fi" copy=bx6 -all now
131 mx bx6 "proc=if p(v35) then 'd35a35#2.5 Argentina^u^v35#' fi" copy=bx6 -all now
132 # v36 <rights> = fecha de embargo
133 mx bx6 "proc=if p(v118) then 'a36#v118#' fi" copy=bx6 -all now
134 # v37 <bundle> = acceso al recurso
135 mx bx6 "proc=if p(v126) then ('a37#replace(v126,'http://localhost:8787','https://ri.conicet.gov.ar'),' ,'#' fi) copy=bx6 -a
  ll now
136
137 # D. Base final auxiliar
138 mx bx6 "proc='d100d101d02d103d104d105d106d107d108d109d110d111d112d113d114d115d116d117d118d119d120d121d122d123d124d125d12-
  6'" copy=bx6 -all now
139 mxcp bx6 create=bx7 clean
140

```



```

62 # Marcado de documentos a ignorar
63 mx seq=${ALMACEN}/xdocs.txt create=xdocs -all now
64 mx xdocs "proc='a3#v1'^m'v2'#'" copy=xdocs -all now
65 mx xdocs "fst=1 0 v1/" fullinv=xdocs -all now
66 mx bx9 "join=xdocs,7:3=mhu,(v30/)" create=bx10 -all now
67 mx bx10 "proc=if p(v7) then 'd7a7#v7^m#' fi" copy=bx10 -all now
68 mx bx10 "proc='d32001'" copy=bx10 -all now
69 mxcp bx10 create=conicet clean
70
71 # Descarga de archivos PDF de Conicet Digital, restringidos o faltantes en Memoria
72 mx conicet lw=10000 "pft=if a(v7) then ,if p(v6) then ,if v6:'REST' or a(v6^d) then ,if v4:'open' then (v6^i[1]:'v37'|ht-
ps://ri.conicet.gov.ar/11336/'v1[1]/),fi,fi,fi,fi" >descargas.txt now
73 for i in $(cat descargas.txt)
74 do
75 wget --user-agent="Mozilla" -P ${ALMACEN}/para_procesar/`echo $i |awk -F "|" '{print $1}'`/ `echo $i |awk -F "|" '{print $2-
}'`
76 echo $i |awk -F "|" '{print $3}' > ${ALMACEN}/para_procesar/`echo $i |awk -F "|" '{print $1}'`/urlConicet.txt
77 done
78
79 # Verificación de nuevos tipos documentales
80 mx seq=${SALIDAS}/tiposDocumentales.txt create=TD -all now
81 mx seq=tiposDocumentales_actual.txt create=TDA -all now
82 mx TD "fst=1 0 v1/" fullinv=TD -all now
83 mx TDA "join=TD,10:1=mhu,v1/" copy=TDA -all now
84 mx TDA "pft=if a(v10) then v1/ fi" >tiposDocumentales_new.txt now
85 if [ -s tiposDocumentales_new.txt ]
86 then
87 mv tiposDocumentales_new.txt ${SALIDAS}/tiposDocumentales_new.txt
88 else
89 rm tiposDocumentales_new.txt
90 fi
91
92 mv "conicet.mst" "${SALIDAS}/conicet.mst"
93 mv "conicet.xrf" "${SALIDAS}/conicet.xrf"
94

```

## Anexo C. Código fuente de CDMA\_procesos

```
CDMA_procesos.sh x
1 #!/bin/bash
2
3 # CDMA_PROCESOS:
4 # Lee la base de cosecha generada por CDMA_tecno y da acceso secuencial a los registros sin correspondencias en Memoria, según tipo documental y unidad de investigación.
5 # En base a elecciones del usuario, ingresa registros y PDFs a las bases de Memoria y al servidor de procesos técnicos, respectivamente, o bien continúa con el acceso secuencial a los registros.
6 # Guarda registro de los ingresos hechos en ingresos.log
7 # Patricia Genovés, marzo/2024.
8
9 DT="${HOME}/CDMA_procesos"
10 PROCESO="${HOME}/CDMA_procesos/proceso"
11 DESCARGAS="/SERVIDOR/procesamiento/CDMA/en_proceso"
12 BCONICET="/SERVIDOR/procesamiento/CDMA/datos_cosecha"
13 LOG="/SERVIDOR/procesamiento/CDMA/ingresos.log"
14 FECHA="date +%Y-%m-%d"
15
16 function registro() {
17 IDCD=`mx ${PROCESO}/regs "pft=v1/" from=${MFN} count=1 now`
18 mx ${PROCESO}/regs "pft=TITULO: ',replace(v10,'^i','^i')/,',AUTORES: ',(v13^a', ',v13^n+ ; |),/,if v3:'tesis' then 'T-
19 ITULO: ',v32/,',UNIVERSIDAD OTORGANTE: ',v33/,',FECHA DE DEFENSA: ',v26 ,else 'PUBLICACION: ',if p(v24) then mdl,v24,mpl,i-
20 f p(v27^x) then 'pp.',v27^x,fi, else ,if p(v25) then mdl,v25, else ,if p(v23) then ,if p(v14) then (v14^n.i. |,v14^a+, |
21 ),',',fi,v23,if p(v27^l) then ',',v27^l,fi,',',',v22,', ',v26, else ,v23^l', ',v22,', ',v26,fi,fi,fi,fi,/,URL CONICET:
22 ',mpl,',https://ri.conicet.gov.ar/handle/11336/,v1/,',TIPO DE ACCESO: ',v4/' lw=100000 from=${MFN} count=1 now >"${IDCD}.t-
23 xt"
24 }
25
26 function mostrar() {
27 yad --title="CDMA : Conicet Digital a Memoria Académica - ${TIPODOC2}" --text-info < ${IDCD}.txt --show-uri --margins=10 -
28 -wrap --button="Ingresar a Memoria:2" --button="Salir:1" --button="Mostrar siguiente:3" --width=700 --height=350
29
30 accion=?
31 cancelar $?
32 }
33
34 function cancelar() {
35 if [ $accion -eq 1 ]; then
36 exit
37 fi
38 if [ $1 -eq 1 ]; then
39 exit
40 fi
41 }
42
43 cd ${PROCESO}
44
45 # Lectura de URLs Conicet de las bases de Memoria para operar sobre registros sin correspondencia actual
46 touch urlma.txt
47 wine cmd /c "basesMA.bat" | zenity --title="CDMA : Conicet Digital a Memoria Académica" --text="Leyendo URLs Conicet en ba-
48 ses de Memoria ..." --progress --pulsate --auto-close --width=600 --height=50
49
50 accion=?
51 cancelar $?
52 bash koha.sh | zenity --title="CDMA : Conicet Digital a Memoria Académica" --text="Leyendo URLs Conicet en Koha ..." --pr-
53 ogress --pulsate --auto-close --width=600 --height=50
54
55 accion=?
56 cancelar $?
57 cat urlma.txt | sed -e 's/http://\hdl.handle.net/11336/???g' | sed -e 's/https://\ri.conicet.gov.ar/handle/11336/
58 ???g' | sed -e 's/https://\ri.conicet.gov.ar/11336/???g' >urlma2.txt
59
60 mx seq=urlma2.txt create=urlma -all now
61 mx urlma "fst=1 0 v1/" fullin=urlma -all now
62 mx ${BCONICET}/conicet "join=urlma,999:l=mhu,v1/" create=regs -all now
63 mx regs "proc=if p(v999) or p(v5) or p(v6) or p(v7) then 'd*' fi" copy=regs -all now
64 mx regs "proc=if v3:'dataset' then ,if v4:'open' then '' else 'd*' fi,fi" copy=regs -all now
65 mx regs "proc='d32001'" copy=regs -all now
66 mx iso=regs.iso -all now
67 mx iso=regs.iso create=regs -all now
68
69 # Conteo de registros por tipo
70 CARTI=`mx regs "pft=if v3:'articulo' then v3/ fi" now |wc -l`
71 CCAP=`mx regs "pft=if v3:'capitulo' then v3/ fi" now |wc -l`
72 CDAT=`mx regs "pft=if v3:'dataset' then v3/ fi" now |wc -l`
73 CEVEN=`mx regs "pft=if v3:'conferencia' then v3/ fi" now |wc -l`
74 CLIB=`mx regs "pft=if v3:'libro' then v3/ fi" now |wc -l`
75 CTES=`mx regs "pft=if v3:'tesis' then v3/ fi" now |wc -l`
76
77 rm tipos.txt
78 touch tipos.txt
79 if [[ ${CARTI} -ge 1 ]] ; then echo "Artículos ${CARTI}" >> tipos.txt ;fi
80 if [[ ${CCAP} -ge 1 ]] ; then echo "Capítulos ${CCAP}" >> tipos.txt ; fi
81 if [[ ${CDAT} -ge 1 ]] ; then echo "Datsets ${CDAT}" >> tipos.txt ; fi
82 if [[ ${CEVEN} -ge 1 ]] ; then echo "Eventos ${CEVEN}" >> tipos.txt ; fi
83 if [[ ${CLIB} -ge 1 ]] ; then echo "Libros ${CLIB}" >> tipos.txt ; fi
84 if [[ ${CTES} -ge 1 ]] ; then echo "Tesis ${CTES}" >> tipos.txt ; fi
85
86 cat tipos.txt | tr -d "\n" >tiposok.txt
87 TIPOS=`cat tiposok.txt`
```

```

73
74 # Selección de registros a ingresar según tipo
75
76 if [[ -s tiposok.txt ]] ; then
77     TIPODOC=yad --list --center --title="CDMA : Conicet Digital a Memoria Académica" --text="\nElegí el tipo de document-
78     o a chequear / ingresar:\n" --column="Tipo de documento" --column="Cantidad" ${TIPOS} --width=400 --height=300 `
79     accion=$?
80     cancelar $?
81 else
82     yad --title="CDMA : Conicet Digital a Memoria Académica" --text-align=center --text="\nNo hay más registros de CONICE-
83     T Digital para ingresar." --image "dialog-info" --width=400 --button="Cerrar:1"
84     accion=$?
85     cancelar $?
86     exit
87 fi
88 TIPODOC2=`echo ${TIPODOC} |awk -F "|" '{print $1}'`
89 echo ${TIPODOC} |awk -F "|" '{print $1}' |sed -e '/\ /d' >tipodoc.txt
90
91 if grep "Articulos" tipodoc.txt ; then
92     mx regs "proc=if v3:'articulo' then '' else 'd*' fi" copy=regs -all now
93     BASE=arti
94     PREID=pr.
95 fi
96 if grep "Capítulos" tipodoc.txt ; then
97     mx regs "proc=if v3:'cap' then '' else 'd*' fi" copy=regs -all now
98     BASE=biblok
99     PREID=pm.
100 fi
101 if grep "Datasets" tipodoc.txt ; then
102     mx regs "proc=if v3:'dataset' then '' else 'd*' fi" copy=regs -all now
103     BASE=biblok
104     PREID=pm.
105 fi
106 if grep "Eventos" tipodoc.txt ; then
107     mx regs "proc=if v3:'conferencia' then '' else 'd*' fi" copy=regs -all now
108     BASE=evento
109     PREID=ev.
110 fi
111
112 if grep "Libros" tipodoc.txt ; then
113     mx regs "proc=if v3:'libro' then '' else 'd*' fi" copy=regs -all now
114     BASE=biblok
115     PREID=pm.
116 fi
117 if grep "Tesis" tipodoc.txt ; then
118     mx regs "proc=if v3:'tesis' then '' else 'd*' fi" copy=regs -all now
119     BASE=tesis
120     PREID=te.
121 fi
122
123 mx regs iso=regs2.iso -all now
124 mx iso=regs2.iso create=regs -all now
125
126 # Marcación de registros por centro
127 mx seq=autores.txt create=autores -all now
128 mx autores "fst=1 0 v1/" fullinv=autores -all now
129 mx regs "proc=if p(v13) then ('a100#v13^a', ',v13^n#'),fi" create=regs1 -all now
130 mx regs1 "join=autores,101:2=mhu,(v100/)" create=regs2 -all now
131 mx regs2 "pft=if p(v101) then (mfn(6)'|v101/),fi" >centros.txt now
132 cat centros.txt |sort |uniq >centros1.txt
133 mx seq=centros1.txt create=centros -all now
134 mx centros "fst=1 0 v1/" fullinv=centros -all now
135 mx regs2 "join=centros,102:2=mhu,mfn(6/)" create=regs3 -all now
136 mx regs3 "proc='d101d32001'" copy=regs3 -all now
137
138 # Conteo de registros por centro
139 rm centros1.txt
140 touch centros1.txt
141 for i in $(cat ../centros.txt)
142 do
143     CC=`mx regs3 "pft=if v102:'$i' then v102/ fi" now |wc -l`
144     if [[ ${CC} -ge 1 ]] ; then echo "$i ${CC}" " >> centros1.txt ;fi
145 done
146 CTODOS=`mx regs3 "pft=v1/" now |wc -l`
147 if [[ ${CTODOS} -ge 1 ]] ; then echo "Todos ${CTODOS}" " >> centros1.txt ;fi
148 cat centros1.txt | tr -d "\n" >centrosok.txt
149 CENTROS=`cat centrosok.txt`
150

```

```

148 # Selección de registros a ingresar según centro
149 if [[ -s centrosok.txt ]] ; then
150     CENTRO=`yad --list --center --title="CDMA : Conicet Digital a Memoria Académica" --text="\nElegí el centro, área o laboratorio, o bien seleccioná Todos\n" --column="Centro de investigación" --column="Cantidad" ${CENTROS} --width=400 --height=600`
151     accion=$?
152     cancelar $?
153 fi
154
155 CENTRO2=`echo ${CENTRO} |awk -F "|" '{print $1}'`
156 NOMBRECENTRO=`echo ${CENTRO} |awk -F "|" '{print $1}' |sed -e '/\ /d'`
157 if [[ $NOMBRECENTRO == "Todos" ]] ; then
158     mx regs3 iso=regs3.iso -all now
159     mx iso=regs3.iso create=regs -all now
160 else
161     mx regs3 "proc=if v102: '${NOMBRECENTRO}' then '' else 'd*' fi" copy=regs3 -all now
162     mx regs3 iso=regs3.iso -all now
163     mx iso=regs3.iso create=regs -all now
164 fi
165
166 # Operación sobre registros seleccionados
167 mx f0 regs create=regsl 0
168 REGISTROS=`mx regsl pft=v1020[2]^d/ now`
169 MFN=1
170 while [[ ${MFN} -le ${REGISTROS} ]]
171 do
172     registro
173     mostrar
174     if [ ${accion} -eq 2 ] ; then
175         mx regs from=${MFN} to=${MFN} iso=reg.iso -all now
176         mx iso=reg.iso create=reg -all now
177         NUMID=`yad --entry --title="CDMA : Conicet Digital a Memoria Académica" --text="\nIngresá el ID que tendrá el registro en Memoria <b>sin el prefijo</b>, ej. 17450" --width=400`
178         accion=$?
179         cancelar $?
180         IDMA=${PREID}${NUMID}
181         mx reg "proc='a920#${IDMA}#','a145#${FECHA}#' copy=reg -all now
182         mx reg lw=10000 "pft=if v4:'open' then (v920[1]'|v37/) fi">descarga.txt now
183         mx reg lw=10000 "pft='${FECHA}|${HOSTNAME}|${IDMA}|registro,if v4:'open' then '|PDF|https://ri.conicet.gov.ar/handle-11336/v1,fi/' >> "$LOG" now
184
185 # Genera archivo Marc de importación para Koha y descarga documentos
186 if [[ ${BASE} == biblok ]] ; then
187     mkdir ${DESCARGAS}/${IDMA}
188     mx reg lw=10000 "pft=@koha.pft now > "${IDMA}.txt"
189     yaz-marcdump -i line -o marc ${IDMA}.txt > "${DESCARGAS}/${IDMA}/${IDMA}.mrc"
190     mx reg lw=10000 "pft=if v3:'capitulo' then 'AUTORES'/(v13^a|,|v13^n/,v13^f/#)/ else ,if p(v13) then 'AUTORES'/(v13^a|,|v13^n/,v13^f/#)/,fi,if p(v14) then 'COLABORADORES'/(v14^a|,|v14^n/,|v14^f/#) fi,fi" > ${DESCARGAS}/${IDMA}/autores.txt now
191     if [[ -s descarga.txt ]] ; then
192         for i in $(cat descarga.txt)
193         do
194             wget --user-agent="Mozilla" -P ${DESCARGAS}/`echo $i |awk -F "|" '{print $1}'`/ `echo $i |awk -F "|" '{print $2}'` | zenity --title="CDMA : Conicet Digital a Memoria Académica" --text="Descargando PDF ..." --progress --pulsate --auto-close --width=600 --height=100
195             accion=$?
196             cancelar $?
197             done
198             yad --title="CDMA : Conicet Digital a Memoria Académica" --text="\n<b>Archivo Marc de importación y PDF descargados en <b>\n procesamiento/CDMA/en_proceso/<b>${IDMA}</b>" --text-align=center --button="Salir:1" --button="Mostrar siguiente:3" --width=500
199             accion=$?
200             cancelar $?
201             else
202             yad --title="CDMA : Conicet Digital a Memoria Académica" --text="\n<b>Archivo Marc de importación guardado en <b>\n procesamiento/CDMA/en_proceso/<b>${IDMA}</b>" --text-align=center --button="Salir:1" --button="Mostrar siguiente:3" --width=500
203             accion=$?
204             cancelar $?
205             fi
206         else
207         # Transforma el registro y lo ingresa a la base Isis de Memoria
208         # Previamente, elimina resúmenes que superan la capacidad del registro
209         mx reg lw=10000 "pft=v15/,v16/,v17" > resumen.txt now
210         RESUMEN=`cat resumen.txt |wc -c`
211         if [[ ${RESUMEN} -gt 7000 ]] ; then
212             mx reg "proc='d15d16d17'" copy=reg -all now
213             yad --title="CDMA : Conicet Digital a Memoria Académica" --text-align=center --text="\nSe eliminan los resúmenes por superar la capacidad del registro" --button="Cancelar:1" --button="Aceptar:0" --image "dialog-warning" --width=400
214             cancelar $?
215         fi

```

```

215     mx reg "proc=@${BASE}.prc" copy=reg -all now
216     crunchmf reg create=regdos
217     yad --title="CDMA : Conicet Digital a Memoria Académica" --text-align=center --text="\nRecordá que la base ${BASE}
debe estar cerrada" --button="Cancelar:1" --button="Aceptar:0" --image "dialog-warning" --width=400
218     cancelar $?
219     wine cmd /c "intro.bat ${BASE}" | zenity --title="CDMA : Conicet Digital a Memoria Académica" --text="Ingresando d
atos ..." --progress --pulsate --auto-close --width=600 --height=50
220     accion=$?
221     cancelar $?
222     # Si hay documentos asociados, los descarga en el servidor de Procesos
223     if [[ -s descarga.txt ]] ; then
224         for i in $(cat descarga.txt)
225         do
226             wget --user-agent="Mozilla" -P ${DESCARGAS}`echo $i |awk -F "|" '{print $1}'`/ `echo $i |awk -F "|" '{print $
2}` | zenity --title="CDMA : Conicet Digital a Memoria Académica" --text="Descargando PDF ..." --progress --pulsate --au-
to-close --width=600 --height=100
227     accion=$?
228     cancelar $?
229         done
230     yad --title="CDMA : Conicet Digital a Memoria Académica" --text="\n<b>Registro agregado a la base ${BASE} y PD
F descargado en </b>\nprocesamiento/CDMA/en_proceso/<b>${IDMA}</b>" --text-align=center --button="Salir:1" --button="Most-
rar siguiente:3" --width=500
231     accion=$?
232     cancelar $?
233     else
234     yad --title="CDMA : Conicet Digital a Memoria Académica" --text="\n<b>Registro agregado a la base ${BASE}</b>-
\n" --text-align=center --button="Salir:1" --button="Mostrar siguiente:3" --width=500
235     accion=$?
236     cancelar $?
237     fi
238     fi
239     fi
240     let MFN=MFN+1
241     done
242     yad --title="CDMA : Conicet Digital a Memoria Académica" --text-align=center --text="\nNo hay más ${TIPODOC2} para cheque-
ar/ingresar" --image "dialog-info" --width=400 --button="Salir:1"
243     accion=$?
244     cancelar $?
245

```

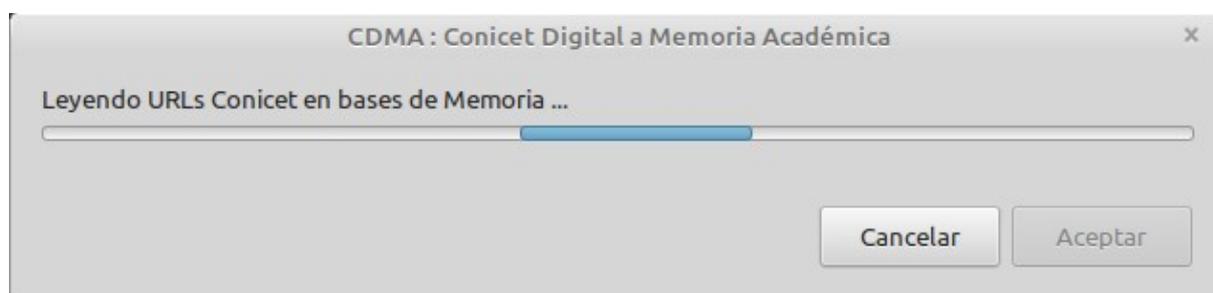
## Anexo D. Manuales de uso

### Manual de uso de CDMA\_procesos para el área de Procesos Digitales

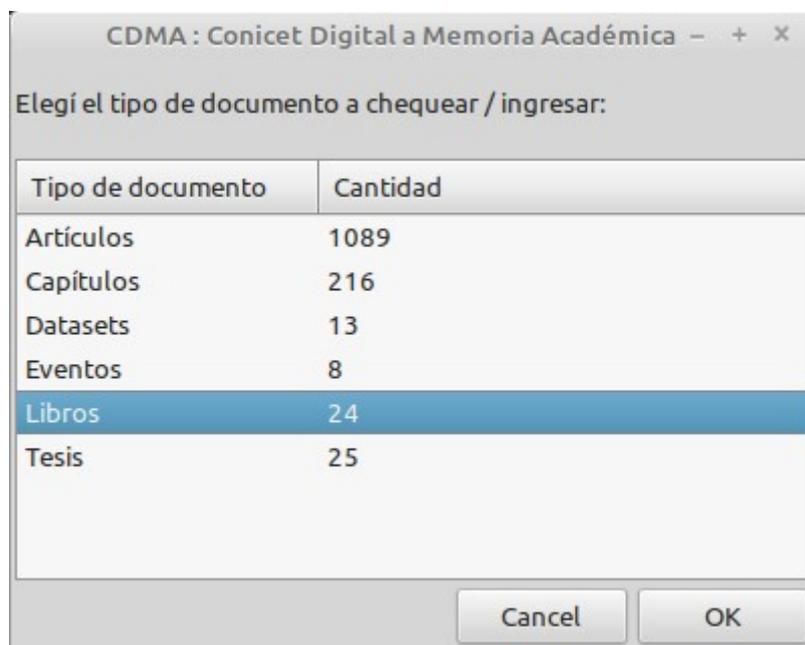
El programa se lanza desde el ícono CDMA ubicado en el Escritorio.

La secuencia de acciones que realiza, y las acciones que corresponden al usuario, son las siguientes:

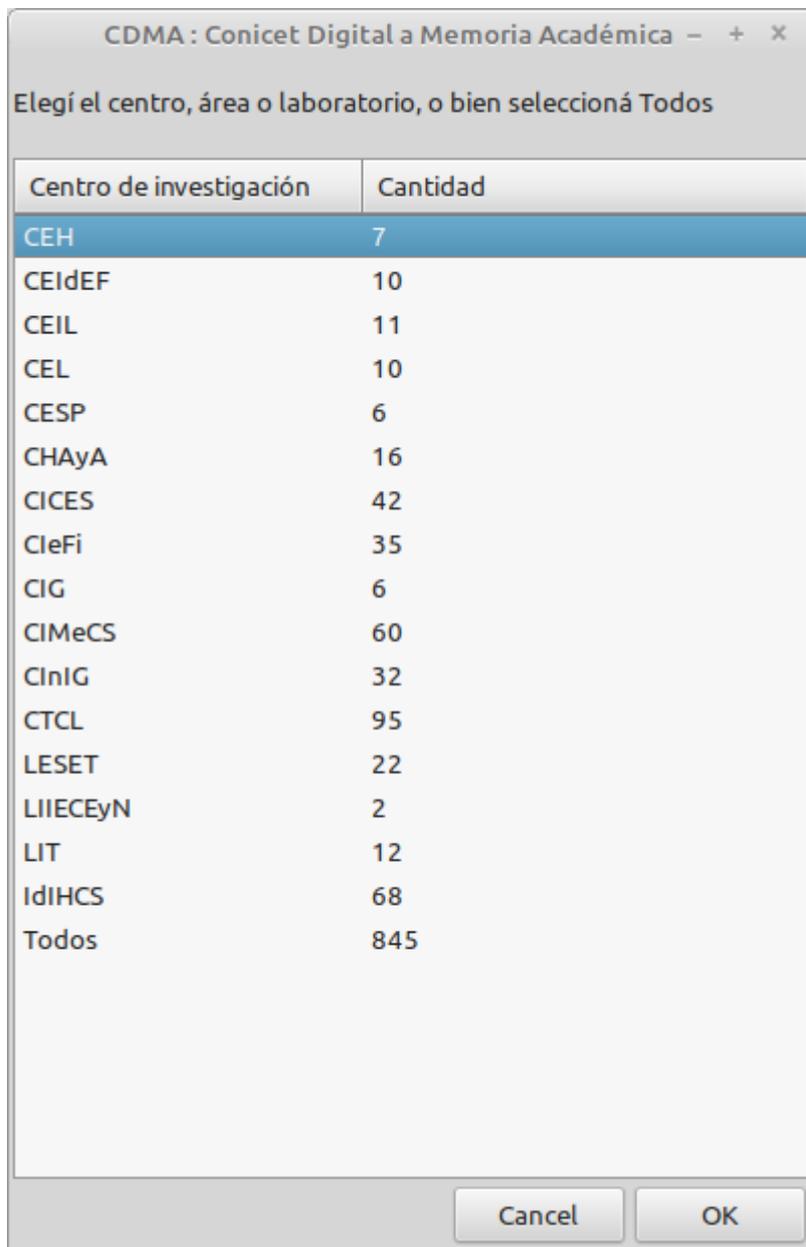
**1-** Comienza chequeando las URLs de CONICET Digital que puedan estar cargadas en las bases de Memoria, a fin de mostrar sólo registros CONICET sin correspondencia con los nuestros.



**2-** Da a elegir el tipo de documento sobre el que se va a trabajar, indicando la cantidad de registros por tipo a procesar para Memoria.

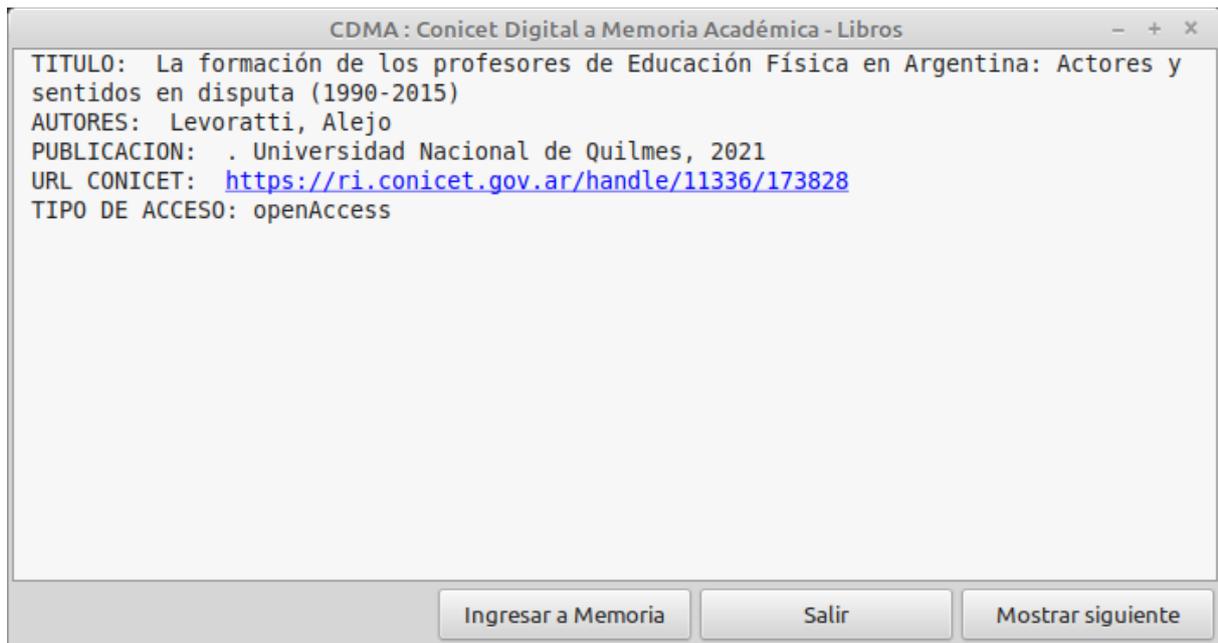


**3-** Permite seleccionar recursos de autores IdIHCS por centro, área o laboratorio de investigación, o bien seguir procesando todos los recursos del tipo documental seleccionado en el paso anterior.

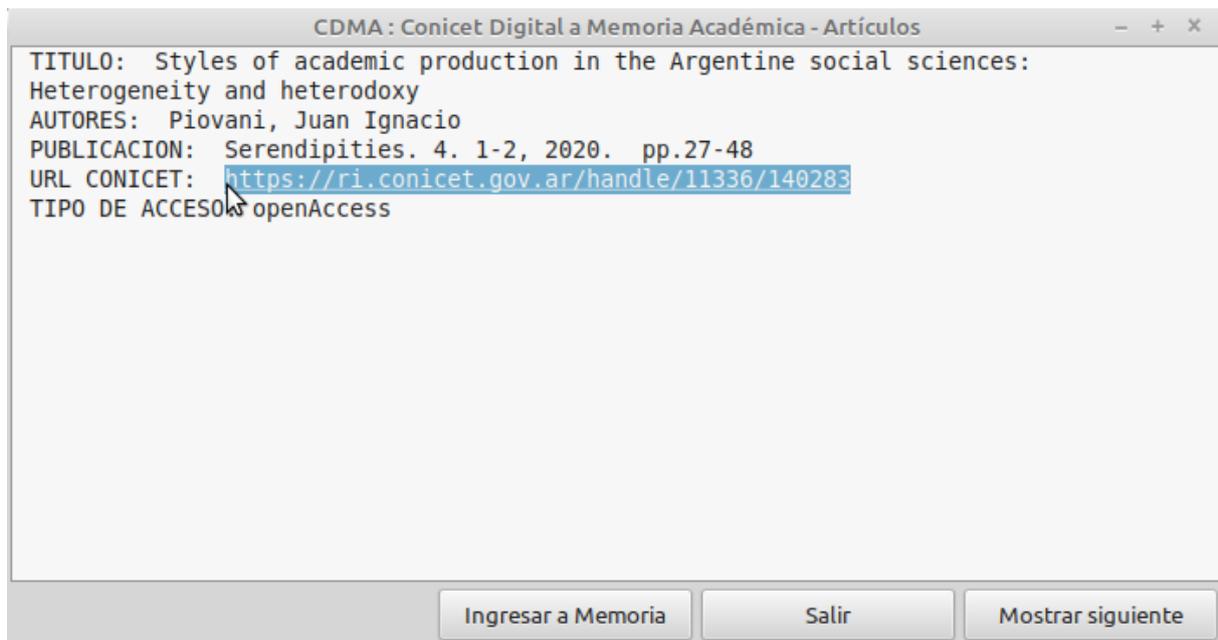


Centro de investigación	Cantidad
CEH	7
CEIdEF	10
CEIL	11
CEL	10
CESP	6
CHAyA	16
CICES	42
CleFi	35
CIG	6
CIMeCS	60
CInIG	32
CTCL	95
LESET	22
LIIECEyN	2
LIT	12
IdIHCS	68
Todos	845

**4-** Muestra el primero de los registros, y aquí se debe, antes que nada, CHEQUEAR que el registro no esté ya en las bases de datos de Memoria.

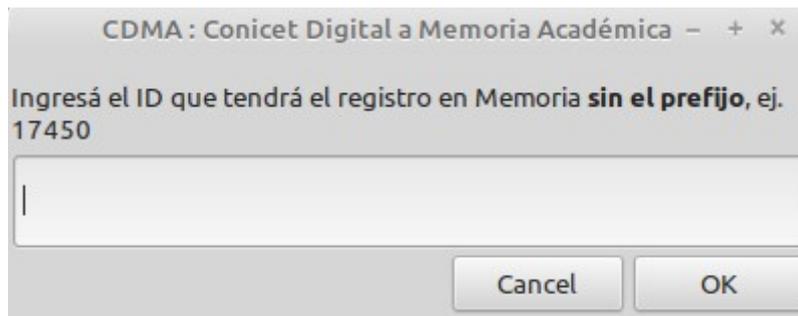


**4a-** Si el registro se encuentra ya cargado en Memoria, **copiar la URL CONICET Digital** y agregarla al registro de Memoria. Luego pulsar en **Mostrar siguiente**



**4b-** Si el registro no está en Memoria, pulsar en **Ingresar a Memoria**

El programa pide el ID que tendrá el recurso en Memoria, y luego recuerda que la base ISIS debe estar cerrada.



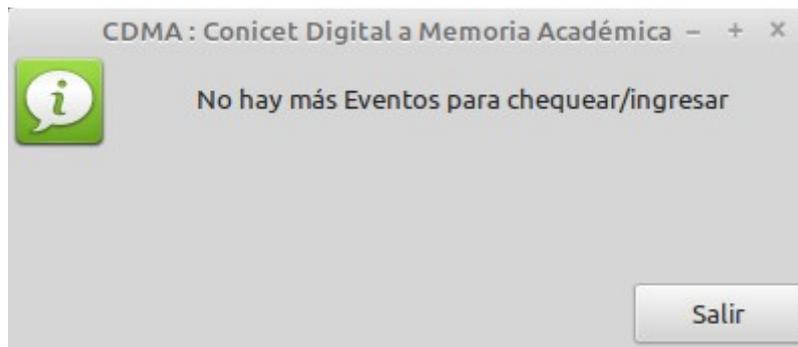
5- Finalizado el ingreso del registro y la descarga del documento (si lo hubiere), lo informa en una pantalla. Aquí se puede optar por continuar el chequeo pulsando en **Mostrar siguiente** o abandonar el programa y continuar más adelante.



En el caso de **libros, capítulos y datasets**, el programa informa que se ha generado y guardado el registro MARC de importación y/o descargado el PDF en el Servidor. También guarda allí un archivo llamado **autores.txt** con el dato de autores y colaboradores del documento, para completar el registro una vez importado a Koha el archivo MARC, ya que no se pueden incluir estos datos en la importación.



6- Cuando ya no hay más registros del tipo elegido para procesar, lo avisa en pantalla.



### Revisión de registros

Los registros ingresados a las bases, y los importados manualmente a Koha, deben ser revisados y completados, ya que puede haber datos faltantes o incompletos, o con errores gramaticales.

### Documentos faltantes

**CDMA\_tecno** es el programa pensado para correrse periódicamente, cada vez que resulte necesario, que cosecha los registros de metadatos correspondientes al IdIHCS en el repositorio CONICET Digital, los procesa y convierte en una base de datos. Esta base es la fuente de información del programa CDMA que corre Procesos Digitales.

Al momento de la cosecha, **CDMA\_tecno** guarda en *procesamiento/CDMA/ para\_procesar* los documentos de aquellos registros CONICET que están cargados en Memoria pero sin documento en abierto. Junto con los documentos se guarda también, en archivo aparte, la URL CONICET del recurso, la cual debe agregarse a los registros de Memoria.

En caso de ingresarlos a Memoria, marcar en el campo **Datos de depósito** lo siguiente:

[Fecha] ^p CR ^m CD

Por otra parte, puede haber documentos que no se descarguen desde la aplicación, es decir, cuya URL no coincida con la provista en el registro de metadatos. En esos casos, se deberá ingresar a la página del recurso en CONICET Digital y descargarlo manualmente; el acceso suele estar a pie de página. La URL CONICET se puede obtener del mismo registro, o bien del archivo **ingresos.log**

### Documentos a ignorar

Si en el transcurso el trabajo de procesamiento se encuentran registros que no deben incluirse en Memoria, o bien documentos faltantes descargados que tampoco van a incluirse por alguna razón, usar el archivo **xdocs.txt** para indicarlo, de modo que se ignoren en la próxima cosecha.

Indicarlo del siguiente modo: URL CONICET | motivo

Ejemplo: <https://ri.conicet.gov.ar/11336/140289>|El PDF contiene sólo un resumen.

## Instructivo para la importación de registros a KOHA

Cuando se procesan libros, capítulos o datasets, CDMA almacena en *procesamiento/CDMA/en\_proceso*, dentro de la carpeta identificada con el ID en Memoria, varios archivos:

- un archivo MARC (.mrc) de importación, para subir a Koha;
- un archivo autores.txt, con la información de autores y colaboradores que debe ingresarse luego manualmente, ya que no pueden importarse, y
- en recursos de acceso abierto, el o los documentos descargables del recurso.

Para la importación de registros MARC (.mrc) a Koha, seguir estos pasos:

### 1. Abrir Koha y pulsar en **Herramientas**.



### 2. Pulsar en **Subir archivo MARC para importación**

Circulación Usuarios Buscar ▼ 🛒 Carrito Más ▼ pgenoves | Bibhuma ▼ Idiomas ▼ Ayuda

koha **BIBHUMA** Ingrese palabras clave para la búsqueda:  
   
 Préstamo Devolución Renovar **Buscar en el catálogo**

> Inicio → Herramientas

**Usuarios y circulación**

**Listas de usuarios**  
Administrar listas de usuarios

**Clubes de usuarios**  
Administrar clubes de usuarios

**Comentarios**  
Moderar comentarios de usuarios

**Importar usuarios**  
Importar datos de usuario

**Avisos y recibos**  
Definir avisos (notificaciones impresas y por e-mail para retrasos, etc.)

**Avisos de retraso/disparadores de estado**  
Establecer avisos de retraso/disparadores de estado para ítem atrasados

**Creador de carnés de usuarios**  
Crear carnés de usuarios para imprimir

**Eliminación y anonimización de usuarios en lote**  
Eliminación en lote de usuarios e historial de circulación de usuarios

**Modificación de usuarios en lote**  
Modificar usuarios en lote

**Extensión de fechas de vencimiento en lote**  
Modificar los vencimientos de préstamos en lote

**Etiquetas**  
Moderar etiquetas de usuario

**Carga de imágenes de usuarios**  
Cargar imágenes de usuarios en lote o una a la vez

**Catálogo**

**Eliminar ítems en lote**  
Eliminar ítems en lotes

**Modificación de ítems en lote**  
Modificar ítems en un lote

**Eliminar registros en lote**  
Eliminar un lote de registros (bibliográficos o de autoridades)

**Modificación de registros en lote**  
Modificar un lote de registros (bibliográficos o autoridades)

**Modificaciones automáticas de ítems por edad**  
Definir reglas para modificar ítems por edad

**Exportar datos**  
Exportar registros bibliográficos, existencias, y autoridades

**Inventario**  
Realizar un inventario de su catálogo

**Creador de etiquetas**  
Crear etiquetas y códigos de barras para imprimir a partir de los datos del catálogo

**Creador rápido de marbete/tejeado**  
Introduzca un código de barras para generar un marbete para imprimir. Se utilizará en impresoras de etiquetas dedicadas

**Colecciones rotativas**  
Administrar colecciones rotativas

**Plantillas de modificación MARC**  
Administrar plantillas para modificar registros MARC durante la importación

**Subir archivos MARC para importación**  
Registros preparados MARC en el depósito

**Gestión de registros MARC para importación**  
Administrar registros MARC preparados, incluyendo completar y revertir importaciones

**Cargar imagen de cubierta local**  
Herramienta para cargar imágenes de cubiertas escaneadas para mostrar en el OPAC

**Herramientas adicionales**

**Calendario**  
Definir los días en los que la biblioteca permanece cerrada.

**Perfiles CSV**  
Administrar perfiles de exportación CSV

**Visor del log**  
Revisar registros del sistema

**Noticias**  
Escriba noticias para el OPAC y la interfaz administrativa

**Planificador de tareas**  
Tareas programadas para ser ejecutadas

**Edite frases para la funcionalidad FDD (Frase del Día)**  
Editor de frases para la característica Frase-del-Día del OPAC

**Subir**  
Subir cualquier tipo de archivo, administrar archivos

**Archivos de acceso**  
Archivos de acceso almacenados en el servidor, tal como archivos de log o informes

### 3. Pulsar en **Examinar** para seleccionar y subir el archivo .mrc

Circulación Usuarios Buscar ▼ 🛒 Carrito Más ▼ pgenoves | Bibhuma ▼ Idiomas ▼ Ayuda

koha **BIBHUMA** Ingrese palabras clave para la búsqueda:  
   
 Préstamo Devolución Renovar **Buscar en el catálogo**

> Inicio → Herramientas → Subir archivos MARC para importación

Herramientas

**Usuarios y circulación**

- ↳ Listas de usuarios
- ↳ Clubes de usuarios
- ↳ Comentarios
- ↳ Importar usuarios
- ↳ Avisos y recibos
- ↳ Avisos de retraso/disparadores de estado
- ↳ Creador de carnés de usuarios
- ↳ Eliminación y anonimización de usuarios en lote
- ↳ Modificación de usuarios en lote

#### Subir archivos MARC para importación

- Seleccione un archivo MARC para preparar en el depósito. Será procesado y cada registro válido será preparado para su posterior uso en el catálogo.
- Puede introducir un nombre para esta importación. Puede ser útil, cuando se crea un registro, para recordar de dónde vienen los datos de MARC sugeridos!

**Importar registros al depósito**

Seleccione el archivo a preparar:  pm.7000.mrc

#### 4. En la siguiente página, no modificar nada y sólo pulsar al final en **Seleccione el archivo a importar**

acerca de este archivo:

Tipo de registro: Bibliográfico

Codificación de caracteres: UTF-8 (Predeterminado)

Formato: MARC

Utilizar plantilla de modificación MARC:

Modificar el registro utilizando la siguiente plantilla: No utilizar.

¿Buscar los registros existentes en el catálogo?

Regla de coincidencia de registro: No busca de registros coincidentes

Acción en caso de encontrar registro coincidente: Reemplace un registro existente con uno nuevo

Acción si no se encontró coincidencia: Agregar registro ingresante

¿Verificar datos de ejemplar embebidos?

Sí

No (Si no comprueba los items mientras los prepara, no podrá cambiar esta opción más adelante)

Como procesar los items: Siempre agregar items

Seleccione el archivo a importar: Guardar perfil Nombre del perfil Eliminar el perfil

5. Koha no importa directamente el archivo dado, sino que lo deja preparado para poder verlo y asegurarse de que va a importarse como uno desea, o bien avisa si hubo errores de importación. Una vez importado sin errores, puede ser ingresado realmente (con lo que aparecerá en las búsquedas) o bien eliminado de este proceso de preparación.

En esta página se muestran los resultados de preparación del archivo MARC subido. Si hay algún error en el archivo, lo marcará en esta pantalla e indicará **1 registro no preparado debido a error MARC**.

También está la opción de incorporar más archivos .mrc y dejarlos preparados para administrarlos más tarde, de a uno por vez. Para incorporar nuevos archivos .mrc se debe pulsar en **+ Registros preparados MARC**.

Si no hubo errores de importación, pulsar en **Administrar registros preparados** para continuar.

- » Herramientas
- Usuarios y circulación**
- » Listas de usuarios
- » Clubes de usuarios
- » Comentarios
- » Importar usuarios
- » Avisos y recibos
- » Avisos de retraso/disparadores de estado
- » Creador de carnés de usuarios
- » Eliminación y anonimización de usuarios en lote
- » Modificación de usuarios en lote

### Subir archivos MARC para importación

Resultados de preparación MARC:

- Procesando registros bibliográficos
- 1 registros en el archivo
- 0 registros no preparados debido a error MARC
- 1 registros preparados
- No se comprobaron coincidencias con los registros existentes en el catálogo
- 1 registros procesados

6. En esta página lo **IMPORTANTE** es seleccionar la hoja de trabajo que tendrá el recurso:

- en libros y datasets = registro monográfico

- en capítulos = capítulo de libro

Puede cliquearse en el título del recurso y ver el registro MARC que se ha creado.

Para importarlo definitivamente, pulsar en **Importar este lote en el catálogo**

[Inicio](#) → [Herramientas](#) → [Administrar registros MARC preparados](#) → [Lote 1857](#)

**Administrar registros MARC preparados** » Lote 1857

**Nombre de archivo:** pm.7000.mrc

**Perfil:** (ninguno)

**Comentarios:** (ninguno)

**Tipo:** Registros bibliográficos

**Preparado:** 04/03/2024 12:30

**Estado:** Preparado

**Regla de coincidencia aplicada:**

**Acción en caso de registro coincidente encontrado:**

**Acción si no se encontró coincidencia:**

**Procesamiento del ítem:**

**Agregar un nuevo registro bibliográfico en esta hoja de trabajo**

Mostrando 1 a 1 de 1 entradas  entradas

#	Citación	Estado	Tipo de coincidencia	Detalles de coincidencia	Diferencia	Registro
223054	Tramas relacionales de las comunidades educativas (9874725354)	Preparado	Ninguna coincidencia			

entradas

7. Luego de la importación final, muestra los resultados y un cartel verde si no hubo problemas.

Ahora el registro es recuperable desde una búsqueda por ID, título, etc. Es **IMPORTANTE** buscarlo para verificar y completar datos, por ejemplo, el de autores y colaboradores.

Desde aquí también se puede deshacer la importación pulsando en **Deshacer importación en el catálogo**

> Inicio - Herramientas - Administrar registros MARC preparados - Lote 1857

Herramientas

Usuarios y circulación

- Listas de usuarios
- Clubes de usuarios
- Comentarios
- Importar usuarios
- Avisos y recibos
- Avisos de retrasos/paradores de estado
- Creador de carnés de usuarios
- Eliminación y anonimización de usuarios en lote
- Modificación de usuarios en lote
- Extensión de fechas de vencimiento en lote
- Moderación de etiquetas
- Carga de imágenes de usuarios

Catálogo

- Eliminar ítems en lote
- Modificación de ítems en lote
- Eliminar registros en lote
- Modificación de registros en lote
- Modificaciones automáticas de ítems por edad
- Exportar datos
- Inventario
- Creador de etiquetas
- Creador rápido de marbetes/tejuelo
- Colecciones rotativas

### Administrar registros MARC preparados > Lote 1857

Nombre de archivo: pm.7000.mrc  
Perfil: (ninguno)  
Comentarios: (ninguno)  
Tipo: Registros bibliográficos  
Preparado: 04/03/2024 12:30  
Estado: Importado  
Regla de coincidencia aplicada: No hay reglas de coincidencia en vigor  
Acción en caso de registro coincidente encontrado: Agregar registro entrante  
Acción si no se encontró coincidencia: Agregar registro ingresante  
Procesamiento del ítem: Siempre agregar ítems

Deshacer importación en el catálogo

Importación completa de registros

Número de registros agregados	1
Número de registros actualizados	0
Número de registros ignorados	0
Número de ítems agregados	1
Número de ítems sustituidos	0
Número de ítems ignorados debido a duplicaciones del código de barras	0

Mostrando 1 a 1 de 1 entradas  entradas [Primero](#) [Previo](#)  [Siguiente](#) [Último](#)

8. Dentro de **Herramientas**, la opción **Gestión de registros MARC para importación** permite:

- eliminar el registro ingresado, aún después de haberlo editado (una vez eliminado, ya no será recuperable en una búsqueda);
- terminar el proceso de importación si se sube el archivo pero se deja sin importar al catálogo;
- eliminar por completo el archivo de importación, es decir, eliminarlo también del proceso de preparación.

## **Anexo E. Contenido del conjunto de datos provisto**

### Contenido de **CDMA\_tecno.zip**:

- CDMA\_tecno.sh
- formateo.sh
- chequeo.sh
- .aCDMA\_tecno.sh
- proceso/
  - o bases2linux.bat
  - o gidioma.txt
  - o gsubject.txt
  - o gtat.txt
  - o txoai.xsl
- salidas/
  - o tiposDocumentales.txt

### Contenido de **CDMA\_procesos.zip**:

- CDMA\_procesos.sh
- .aCDMA\_procesos.sh
- centros.txt
- proceso/
  - o arti.prc
  - o basesMA.bat
  - o evento.prc
  - o intro.bat
  - o koha.pft
  - o koha.sh
  - o koha.xsl
  - o tesis.prc

*///*