



# Curso Taller sobre DROID – Herramienta de Identificación de Formatos de Objetos Digitales

Marisa De Giusti - Esteban Fernández  
PREBI-SEDICI-UNLP  
CESGI-CIC  
28 de Agosto de 2020



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



EDUCACIÓN  
PÚBLICA  
Y GRATUITA



UNIVERSIDAD  
NACIONAL  
DE LA PLATA



# Instalación de DROID

- Si utiliza Windows es posible descargar una versión que tiene incluido Java <https://github.com/digital-preservation/droid/releases/download/droid-6.5/droid-binary-6.5-bin-win32-with-jre.zip>
- Para instalarlo solo hay que descomprimir la carpeta y buscar el archivo droid.bat
- Por ejemplo se descomprime la carpeta en el Escritorio y ejecutamos como “Administrador” el archivo droid.bat

## **Ayuda:**

<https://docs.google.com/document/d/13qQIQGan2SxKvZWH5m39s9PSFE-3hvHdlulHEy1qFPc/edit?usp=sharing>






DROID collects a variety of information about your files and folders, including:

| <a href="#">Type</a>                       | <a href="#">File name</a>                        | <a href="#">File name extension</a> |
|--|--|-------------------------------------|
| <a href="#">Extension mismatch warning</a> | <a href="#">Location</a>                         | <a href="#">File size</a>           |
| <a href="#">Last modified date</a>         | <a href="#">Number of format identifications</a> | <a href="#">File formats</a>        |
| <a href="#">Identification method</a>      | <a href="#">Content hash</a>                     | <a href="#">Status</a>              |

## Type

[top](#)

DROID categorises the files and folders it profiles as being one of three types:

-  file
-  folder
-  archival file

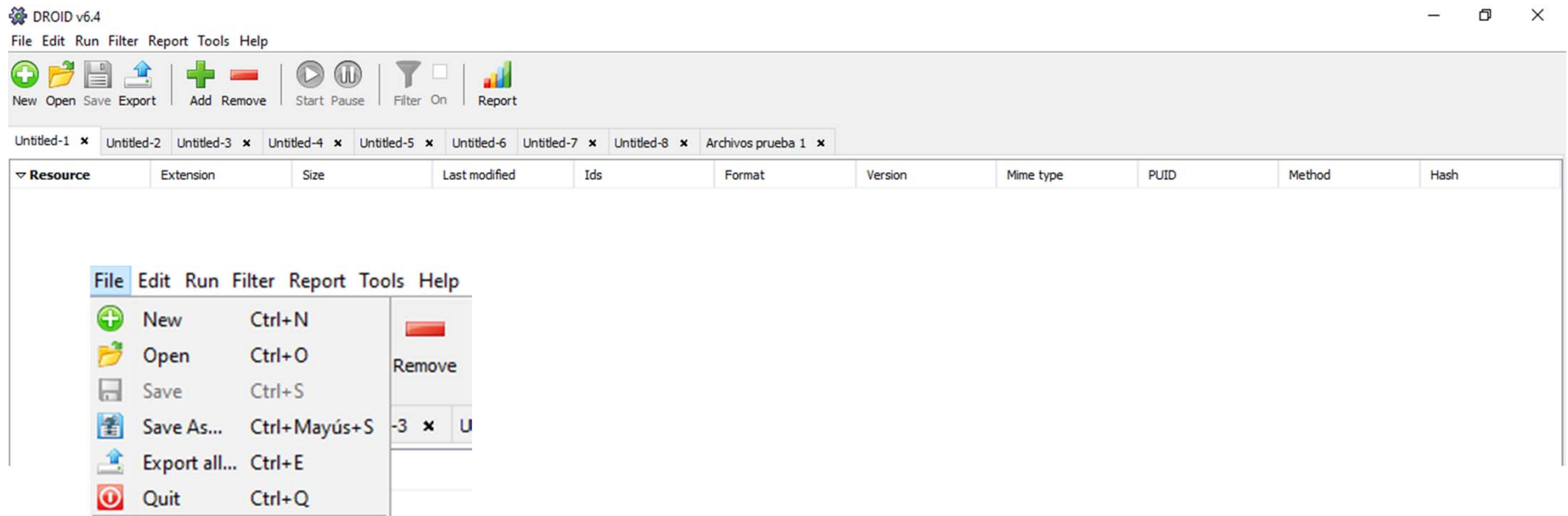
**Files** have format identifications, but do not have other files or folders inside them. **Folders** do not have any format identifications or sizes, but can contain other folders, files and archival files inside them. **Archival files** are like folders, in that they can contain other folders, files and archival files inside them, but they are also files, so they have format identifications and a file size. In this version, DROID can look inside zip, tar and gzip archival files. Archival files may have other archival files nested inside them. DROID will also profile inside these, and in any further nested archival files.

## File name

[top](#)

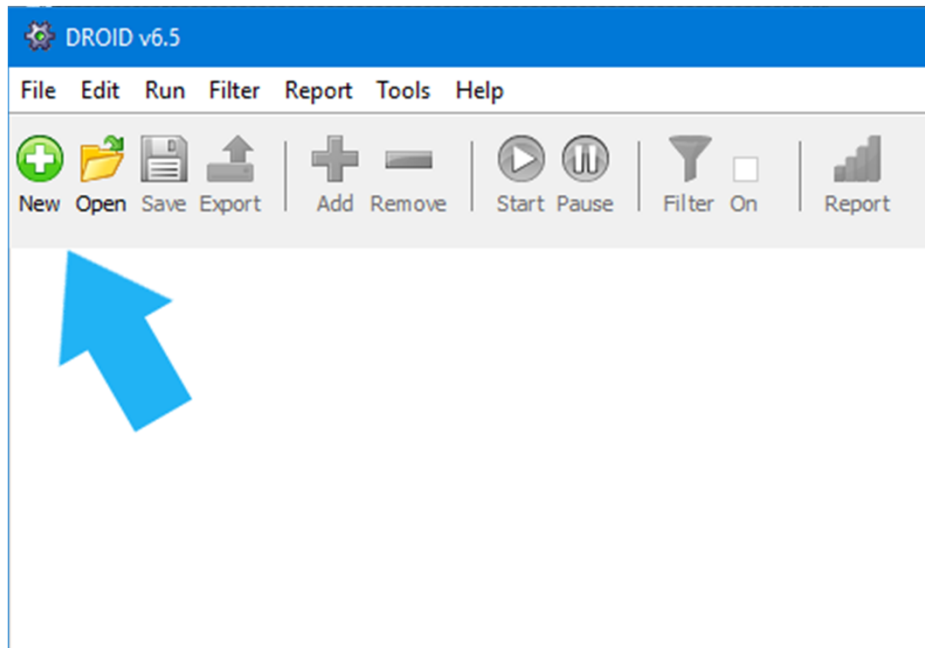
The name of a file, folder or archival file is its name, independent of its location on a disk or inside an archival file. The file name extension (if any) is part of its name. DROID treats all filenames as

# Dos menús



# 1. Crear un perfil

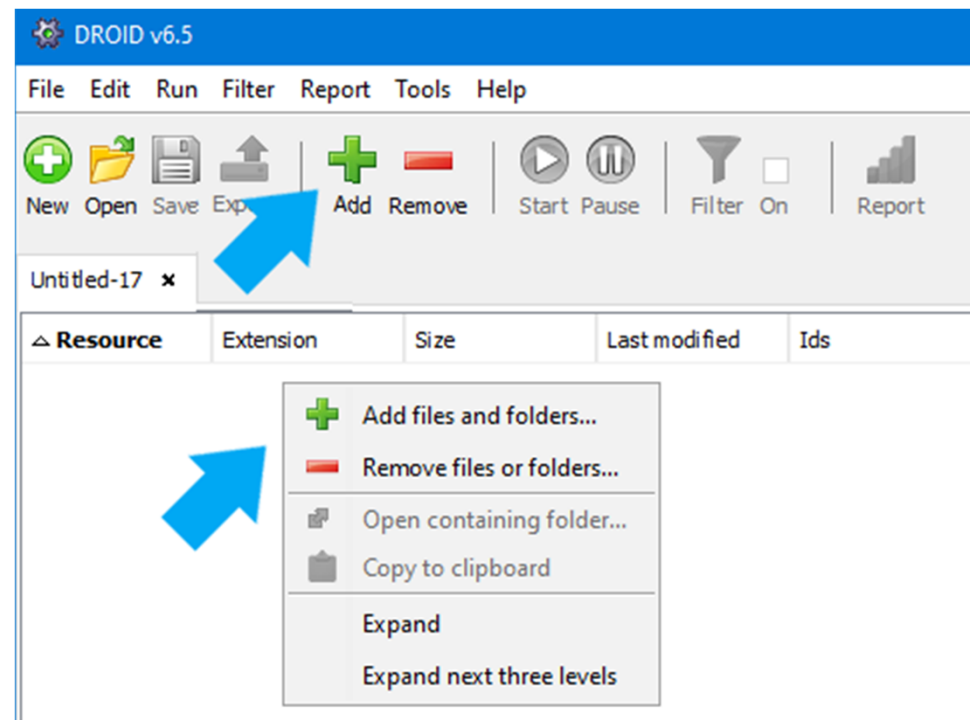
Desde el botón New o desde el menú File/New



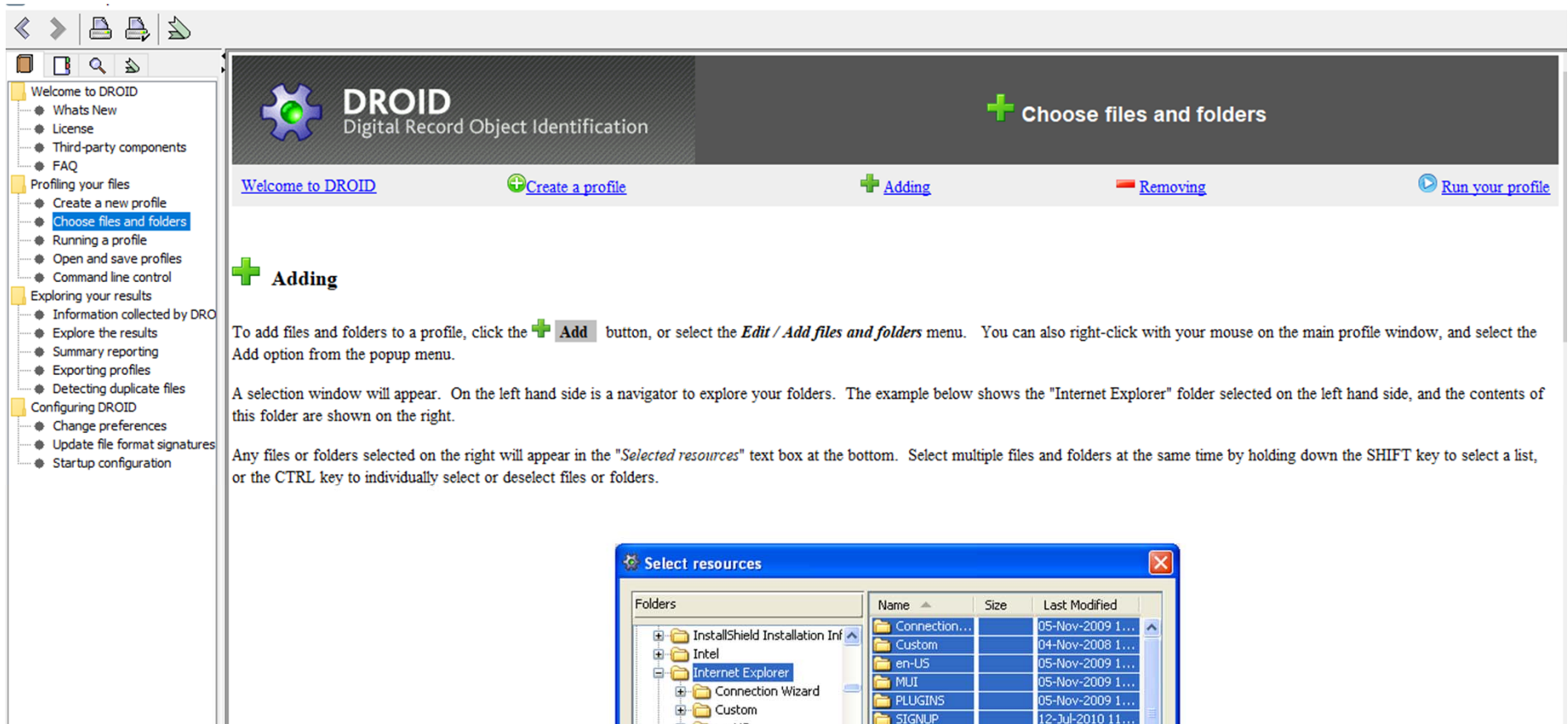
Si ya tengo un directorio con archivos para perfilar comienzo a hacer el perfil

2. En el caso de no tener ya un conjunto de archivos. Se los debe agregar: carpetas y subcarpetas o archivos

Desde el botón Add ó haciendo clic derecho



# Sumar archivos, elegirlos de un directorio



The screenshot shows the DROID (Digital Record Object Identification) application interface. On the left is a navigation tree with categories like 'Welcome to DROID', 'Profiling your files', 'Exploring your results', and 'Configuring DROID'. The 'Choose files and folders' option is highlighted. The main window has a header with the DROID logo and a 'Choose files and folders' button. Below the header are navigation links: 'Welcome to DROID', '+ Create a profile', '+ Adding', - Removing, and 'Run your profile'. The 'Adding' section is active, showing instructions on how to add files and folders to a profile. A 'Select resources' dialog box is open, showing a folder tree on the left and a list of files on the right. The 'Internet Explorer' folder is selected in the tree, and its contents are listed in the table below.

**+ Adding**

To add files and folders to a profile, click the **+ Add** button, or select the *Edit / Add files and folders* menu. You can also right-click with your mouse on the main profile window, and select the Add option from the popup menu.


A selection window will appear. On the left hand side is a navigator to explore your folders. The example below shows the "Internet Explorer" folder selected on the left hand side, and the contents of this folder are shown on the right.

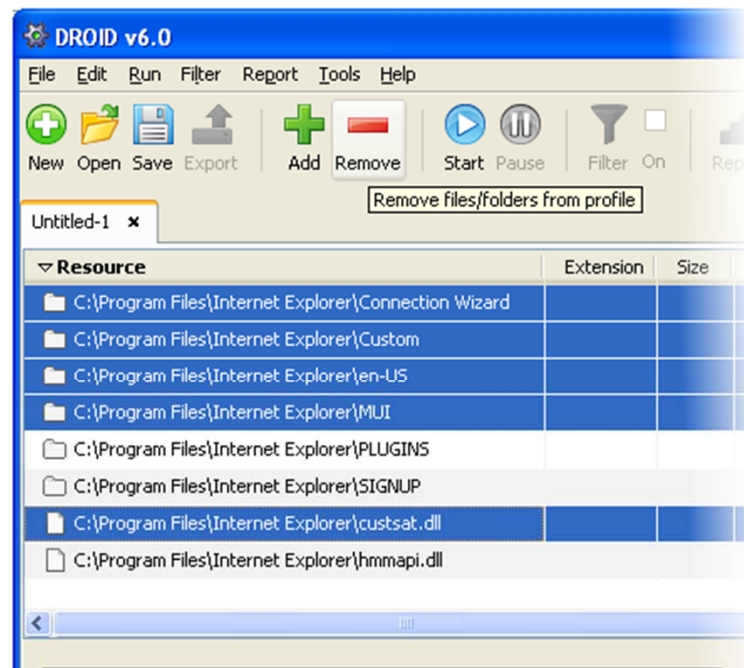
Any files or folders selected on the right will appear in the "Selected resources" text box at the bottom. Select multiple files and folders at the same time by holding down the SHIFT key to select a list, or the CTRL key to individually select or deselect files or folders.

| Folders | Name          | Size | Last Modified     |
|---------|---------------|------|-------------------|
| +       | Connection... |      | 05-Nov-2009 1...  |
| +       | Custom        |      | 04-Nov-2008 1...  |
| +       | en-US         |      | 05-Nov-2009 1...  |
| +       | MUI           |      | 05-Nov-2009 1...  |
| +       | PLUGINS       |      | 05-Nov-2009 1...  |
| +       | SIGNUP        |      | 12-JUL-2010 11... |

### 3. Remover archivos del conjunto

#### Removing

To remove files or folders from the profile, select them in the main window, and either press the  **Remove** button, or use the *Edit / Remove files or folders* menu item. You can also right-click with your mouse on the main profile window, and select the Remove option from the popup menu.

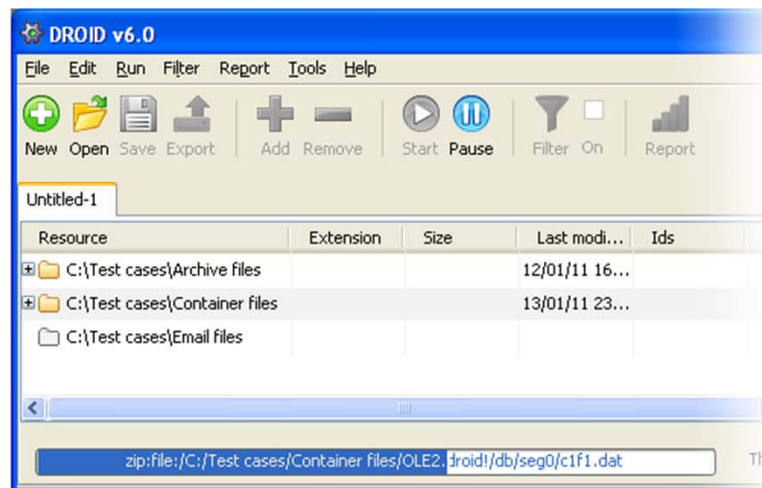




## 4. Ejecutar un perfil

### Running

You can start profiling files or folders using the  **Start** button, or by selecting the *Run / Start Identification* menu item.



## 5. Visualizar reporte



Report

### Comprehensive breakdown

#### Profile Summary

| Name        | Signature version | Container version | Started      | Finished     | Filters |
|-------------|-------------------|-------------------|--------------|--------------|---------|
| Untitled-20 | 96                | 20200121          | 26 ago. 2020 | 26 ago. 2020 |         |

#### File & folder count

| Report field | Grouping fields |
|--------------|-----------------|
| FILE_NAME    |                 |

| Profile               | Count    | Sum | Min | Max | Average |
|-----------------------|----------|-----|-----|-----|---------|
| Untitled-20           | 4        |     |     |     |         |
| <b>Profile totals</b> | <b>4</b> |     |     |     |         |

## Reporte de DROID

El perfil que realiza DROID sobre los formatos de los archivos de interés contrasta los mismos con los formatos del registro internacional PRONOM.

El reporte que ofrece DROID, es global, es decir que analiza agrupamientos de archivos por formatos y no hace reportes sobre objetos individuales.

**Tabla 5.1: Reportes provistos por DROID**

| REPORTES DE DROID                               | DESCRIPCIÓN DE LOS REPORTES   |
|---|---|
| <i>File count and sizes</i>                     | Suma, tamaño total, mínimo, máximo y promedio de todos los archivos del perfil.   |
| <i>Total count of files and folders</i>         | Suma de todos los archivos y carpetas del perfil.   |
| <i>Total unreadable files</i>                   | Suma de todos los archivos ilegibles.   |
| <i>Total unreadable folders</i>                 | Suma de todas las carpetas ilegibles.   |
| <i>File count and sizes by file extension</i>   | Suma, tamaño total, mínimo, máximo y promedio de todos los archivos del perfil, separados según su extensión.   |
| <i>File count and sizes by file format PUID</i> | Suma, tamaño total, mínimo, máximo y promedio de todos los archivos del perfil, separados de acuerdo a sus formatos PUID (PRONOM Unique IDentifiers). |
| <i>File count and sizes by MIME type</i>        | Suma, tamaño total, mínimo, máximo y promedio de todos los archivos del perfil, separados de acuerdo a sus MIME types.                                |

|  |   |
|--|---|
| <b><i>File count and sizes by month last modified</i></b>          | Suma, tamaño total, mínimo, máximo y promedio de todos los archivos del perfil, separados de acuerdo al último mes en que fueron modificados. Los meses son representados por números 1 (enero) hasta 12 (diciembre). |
| <b><i>File count and sizes by year last modified</i></b>           | Suma, tamaño total, mínimo, máximo y promedio de todos los archivos del perfil, separados por el último año en que fueron modificados.  |
| <b><i>File count and sizes by year and month last modified</i></b> | Suma, tamaño total, mínimo, máximo y promedio de todos los archivos del perfil, separados por año y mes de última modificación.   |
| <b><i>Comprehensive breakdown</i></b>                              | Reporte que combina todos los reportes precedentes.   |

Fuente: Tesis Completa Marisa De Giusti página 118:

<http://sedici.unlp.edu.ar/handle/10915/43157>

|       | Extension        | Size     | Last modified  | Ids <sup>1</sup> | Format                  | Version      | Mime type             | PUID <sup>2</sup>       |
|-------|------------------|----------|----------------|------------------|-------------------------|--------------|-----------------------|-------------------------|
| DI... | pdf              | 52,4 KB  | 25/8/20 8:31   |                  | Acrobat PDF 1.2 - Po... | 1.2          | application/pdf       | <a href="#">fmt/16</a>  |
| DI... | pdf <sup>5</sup> | -1 bytes | 31/12/69 21:00 |                  |                         |              |                       |                         |
| DI... | jpg <sup>6</sup> | 88,8 KB  | 26/8/20 8:02   |                  | Microsoft Word for ...  | 2007 onwards | application/vnd.op... | <a href="#">fmt/412</a> |
| DI... | pdf              | 52,4 KB  | 25/8/20 8:31   |                  | Acrobat PDF 1.2 - Po... | 1.2          | application/pdf       | <a href="#">fmt/16</a>  |

Algunos de los puntos importantes a tener en cuenta del análisis que realiza DROID

1. Ids: Cero si no puede identificar el formato Uno, si se identifica inequívocamente como un formato único. <sup>(2)</sup> más de uno, si DROID no puede decidir sin ambigüedades en qué formato está
2. PUID: es un identificador persistente y único a nivel mundial para un formato de archivo y una versión, asignado por los National Archives. Pueden consultarse los formatos en <http://www.nationalarchives.gov.uk/PRONOM/BasicSearch/proBasicSearch.aspx?status=new>

3. Método de identificación: **Extensión** se identificó únicamente sobre la base de su extensión de archivo. **Firma** se identificó al encontrar patrones de firma dentro del archivo que se sabe que ocurren en formatos y versiones de archivo particulares. **Contenedor**: se identificó al encontrar archivos incrustados (posiblemente con sus propias firmas) dentro del archivo principal. Por ejemplo, los archivos de procesamiento de texto de Microsoft Office 2007 son en realidad archivos zip que contienen archivos xml, imágenes u otros recursos utilizados en el documento.

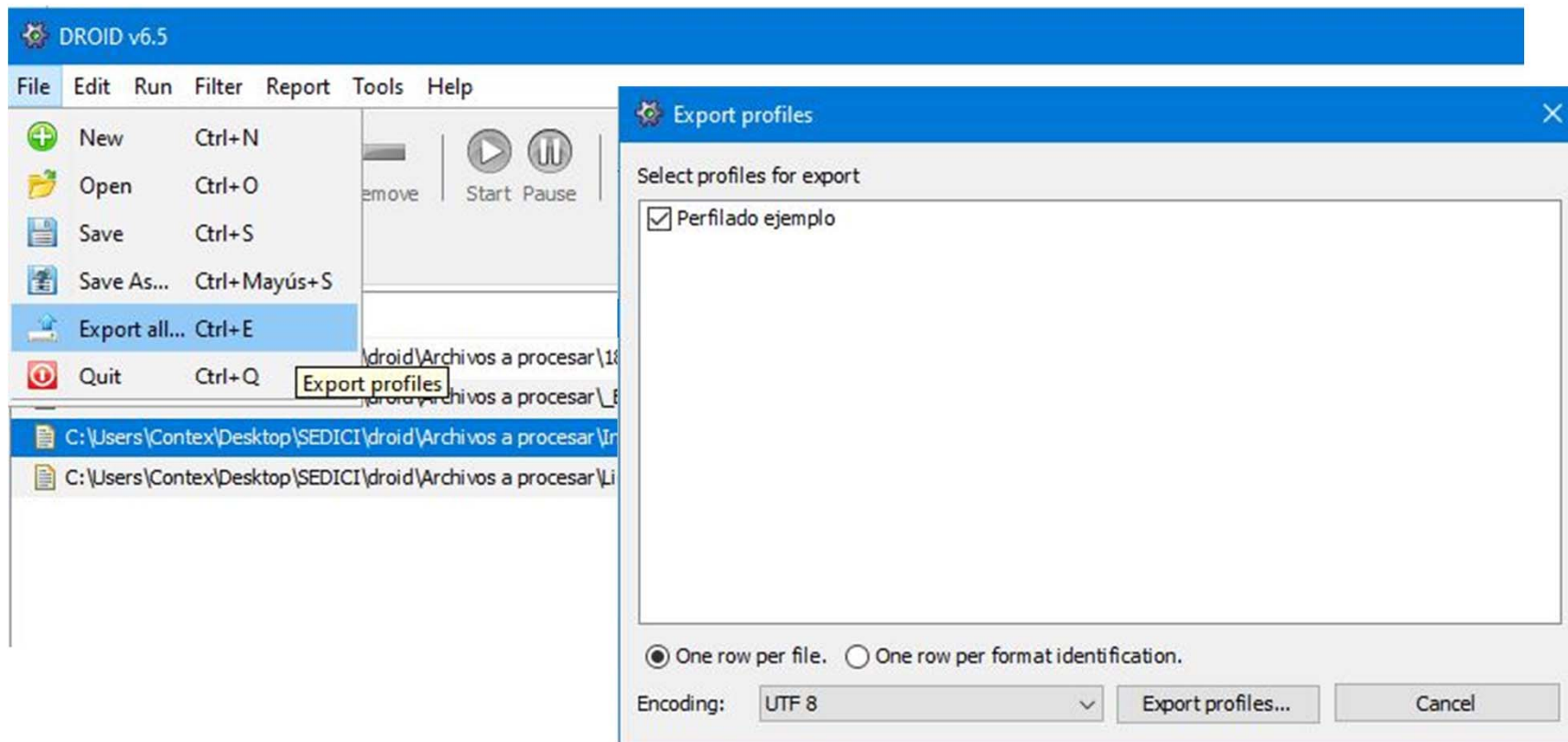
4. Generador de HASH 'MD5' 'SHA1' o 'SHA256' es importante para detectar los archivos duplicados (aunque DROID no tiene una herramienta de detección de duplicados se pueden exportar los resultados en un archivo csv para luego procesarlo e identificarlos)

5. Este es un ejemplo de un archivo que no tiene extensión, DROID detectó que es pdf, pero para que muestre más información es necesario colocar la extensión y volver a perfilarlo

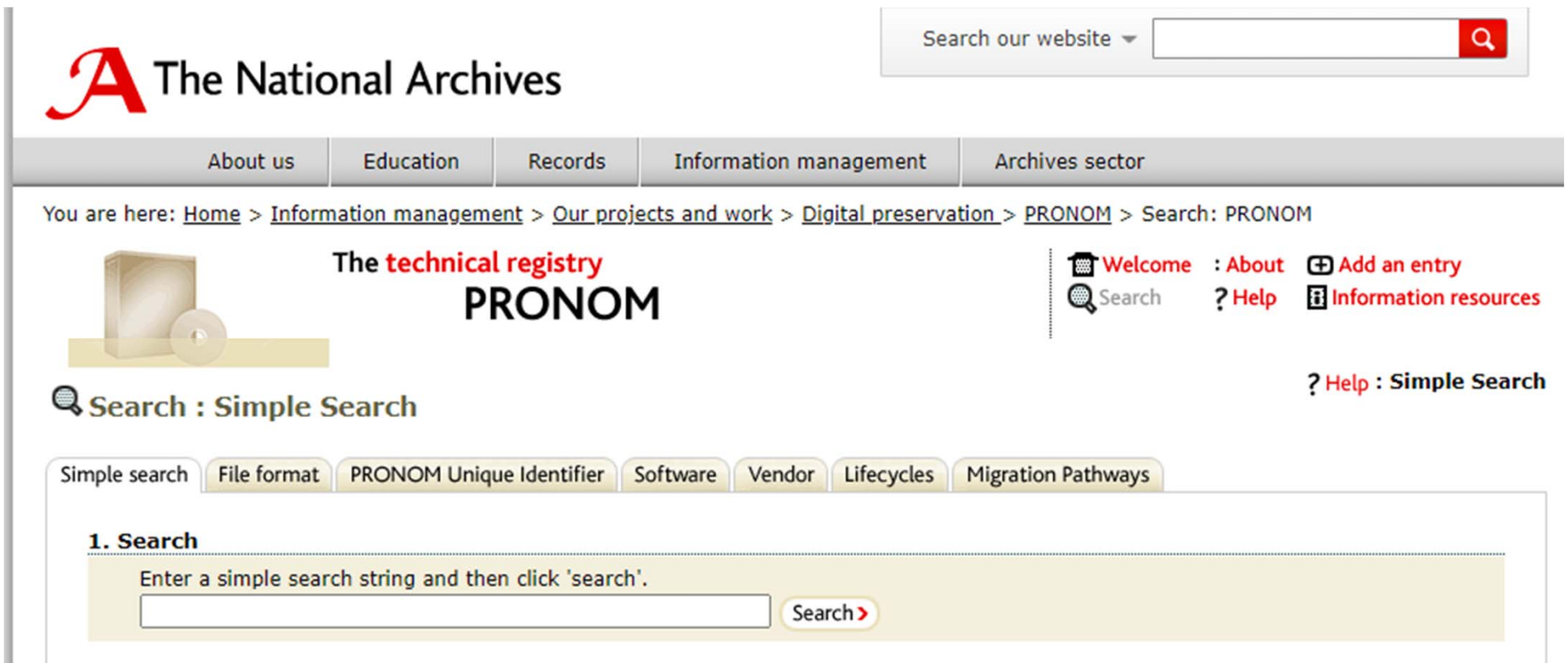
6. Este es un ejemplo de detección de archivos con extensión errónea. La extensión del documento es .jpg pero en realidad es un docx. DROID lo detectó y muestra la información que corresponde



# Exportar perfil como .csv



# Cómo buscar un formato en Pronom



The screenshot shows the website for The National Archives' PRONOM project. At the top left is the logo for The National Archives, featuring a large red letter 'A' followed by the text 'The National Archives'. To the right of the logo is a search bar with the text 'Search our website' and a magnifying glass icon. Below the logo is a navigation menu with five items: 'About us', 'Education', 'Records', 'Information management', and 'Archives sector'. Below the navigation menu is a breadcrumb trail: 'You are here: [Home](#) > [Information management](#) > [Our projects and work](#) > [Digital preservation](#) > [PRONOM](#) > Search: PRONOM'. Below the breadcrumb trail is a section for 'The technical registry PRONOM'. On the left of this section is an image of a CD-ROM and its jewel case. To the right of the image is the text 'The technical registry PRONOM'. Further right is a list of links: 'Welcome', 'Search', 'About', 'Help', 'Add an entry', and 'Information resources'. Below this list is a link for '? Help : Simple Search'. Below the 'The technical registry PRONOM' section is a search bar with the text 'Search : Simple Search'. Below the search bar is a row of tabs: 'Simple search', 'File format', 'PRONOM Unique Identifier', 'Software', 'Vendor', 'Lifecycles', and 'Migration Pathways'. Below the tabs is a section titled '1. Search' with a light yellow background. Inside this section is a text input field with the placeholder text 'Enter a simple search string and then click 'search''. To the right of the input field is a 'Search >' button.

# Cómo buscar un formato en Pronom

The screenshot shows the PRONOM search page on The National Archives website. At the top left is the National Archives logo. A search bar is located at the top right. Below the logo is a navigation menu with categories: About us, Education, Records, Information management, and Archives sector. A breadcrumb trail indicates the current location: Home > Information management > Our projects and work > Digital preservation > PRONOM > Search: Search by PRONOM Unique Identifier. The main heading is 'The technical registry PRONOM'. To the right are links for Welcome, About, Add an entry, Search, Help, and Information resources. Below this is a search prompt: 'Search : By PRONOM Unique Identifier' with a help link for 'search by PUID'. A row of tabs includes Simple search, File format, PRONOM Unique Identifier (selected), Software, Vendor, Lifecycles, and Migration Pathways. The selected tab shows instructions: '1. PRONOM Unique Identifier' followed by 'To search for a particular file format, piece of software, compression type, encoding type, or piece of hardware, enter the PRONOM Unique Identifier (PUID) and then click 'search''. Below the instructions is a search input field containing 'fmt/276' and a 'Search >' button.

Search our website



- About us
- Education
- Records
- Information management
- Archives sector

You are here: [Home](#) > [Information management](#) > [Our projects and work](#) > [Digital preservation](#) > [PRONOM](#) > Search: Search by PRONOM Unique Identifier



The **technical registry**  
**PRONOM**

- Welcome
- About
- Add an entry
- Search
- Help
- Information resources

? [Help](#) : search by PUID

Search : By PRONOM Unique Identifier

- Simple search
- File format
- PRONOM Unique Identifier
- Software
- Vendor
- Lifecycles
- Migration Pathways

## 1. PRONOM Unique Identifier

To search for a particular file format, piece of software, compression type, encoding type, or piece of hardware, enter the PRONOM Unique Identifier (PUID) and then click 'search'.

# Usos de DROID para perfilar contenidos de RIs

- DROID tiene un uso concreto muy importante para quiénes gestionan repositorios porque permite saber qué formatos tenemos e incluso si hay archivos duplicados, si hay archivos sin la extracción de texto completo, sin licencias...
- El desafío es extraer los contenidos del repositorio para ponerlos en DROID y hacer el perfil.
- La experiencia nuestra es con DSPACE.
- Una posibilidad es la exportación del “assetstore” y el desencarpetamiento del mismo y además de cada ítem tomar solamente el bitstream que está en el bundle original y también el del bundle text.
- Esta tarea requiere intervención de un administrador informático.

# Cómo obtener los contenidos de un repositorio

## desde el servidor, ejecutando comandos

### 1) Exportar contenido en formato AIP

- <https://wiki.lyrasis.org/display/DSDOC6x/AIP+Backup+and+Restore#AIPBackupandRestore-ExportingAIPs>

### 2) Exportar contenido en formato SAF

- <https://wiki.lyrasis.org/display/DSDOC6x/Importing+and+Exporting+Items+via+Simple+Archive+Format#ImportingandExportingItemsviaSimpleArchiveFormat-ExportingItems>

### 3) Copiar parte del assetstore

## desde interfaz web:

### 1) Exportar contenido en formato SAF

- A) exportar item, colección, comunidad
- B) revisar mail

# Desde Interfaz web SEDICI

← → ↻ No es seguro | sedici.unlp.edu.ar/handle/10915/101196 ☆ 📄 📁 📄 📄 📄 📄 📄 📄 📄 📄

Aplicaciones Recibidos (833) - m... UNLP BNA BP BBVA Francés Sueldo UNLP Datos Archivatic... Archivematica Archivematica DOCs SEDICI CIC Digital

## SEDICI

REPOSITORIO INSTITUCIONAL DE LA UNLP



UNIVERSIDAD NACIONAL DE LA PLATA

Inicio Buscar material Subir material Institucional Preguntas frecuentes Contacto Administración

Unidades académicas → Presidencia → Editorial de la Universidad Nacional de La Plata (EDULP) → Libros de Cátedra

### Buscar material

Busque entre los mas de **96500** recursos disponibles en el repositorio

Subir material

Suba sus trabajos a SEDICI, para mejorar notoriamente su visibilidad e impacto

[¿Cómo cargo mis trabajos?](#)

[Ayuda con...](#)

## Socio/logía

*Algunos recorridos por la teoría social*

Compiladores: Lacchini, Ana Julia | Casajús, Rocio | Manuele, Matías Eugenio 2020

Tipo de documento: Libro Like Share Twitter R

### Resumen

El texto se presenta como un material orientado a estudiantes de los colegios preuniversitarios de la Universidad y otras instituciones. El objetivo es acompañar el proceso de enseñanza aprendizaje de la teoría social en general y sus conceptos fundamentales, adecuando conceptos teóricos para el nivel, incorporando ejemplos, trabajos de investigación e intervención, exponiendo el trabajo de la propia UNLP e identificando los problemas del ámbito territorial de las y los estudiantes. Para ello se convocó a producción conjunta, en primer lugar entre docentes de los tres colegios preuniversitarios de la UNLP y en segundo lugar, en la articulación de los pregrados con docentes de diferentes facultades de la Universidad. Se trata fundamentalmente de dar cuenta del complejo mundo de las relaciones sociales y recuperar orígenes, autores y perspectivas del pensamiento sociológico.

**Marisa Raquel De Giusti** Salir

- Editar perfil
- Mis documentos
- Mis exportaciones
- Contexto
- Editar este ítem
- Exportar ítem
- Exportar metadatos
- Generar certificado de depósito
- Editar Item



# Desde Interfaz web CIC Digital

Investigadores en Universidades Nacionales de la provincia de Buenos Aires / Universidad Nacional del Sur (UNS) / Artículos

Artículo . European Physical Journal E ; vol. 41, no.9

## A first principles study of pristine and Al doped activated carbon interacting with 5-Fluorouracil anticancer drugs

Román, G. | Nosedá Grau, Emilia | Díaz Compañy, Andrés Carlos Daniel | Brizuela, G. | Juan, Alfredo | Simonetti, Sandra

Año  
2018

- Editar este ítem
- Exportar ítem
- Exportar metadatos

### Resumen:

The adsorption of the 5-Fluorouracil (5-FU) molecule on the pristine and Al-doped activated carbon (AC) was investigated by using the Vienna Ab-initio Simulation Package. It is found that the 5-FU molecule is only weakly adsorbed on the pristine AC with high adsorption energy and large surface distance. The adsorption of the 5-FU molecule on pristine AC is highly disfavored. In contrast, the molecule shows strong interactions with the Al-doped AC confirmed by the lesser adsorption energy, the charge transfers on the Al-modified zone and the significant changes in the DOS at the Fermi level. The results of our study suggest that the Al dopant increases the adsorption capacity of AC enhancing its interactions with polar atoms of the adsorbate, hence improving its adsorption properties.



Descargas

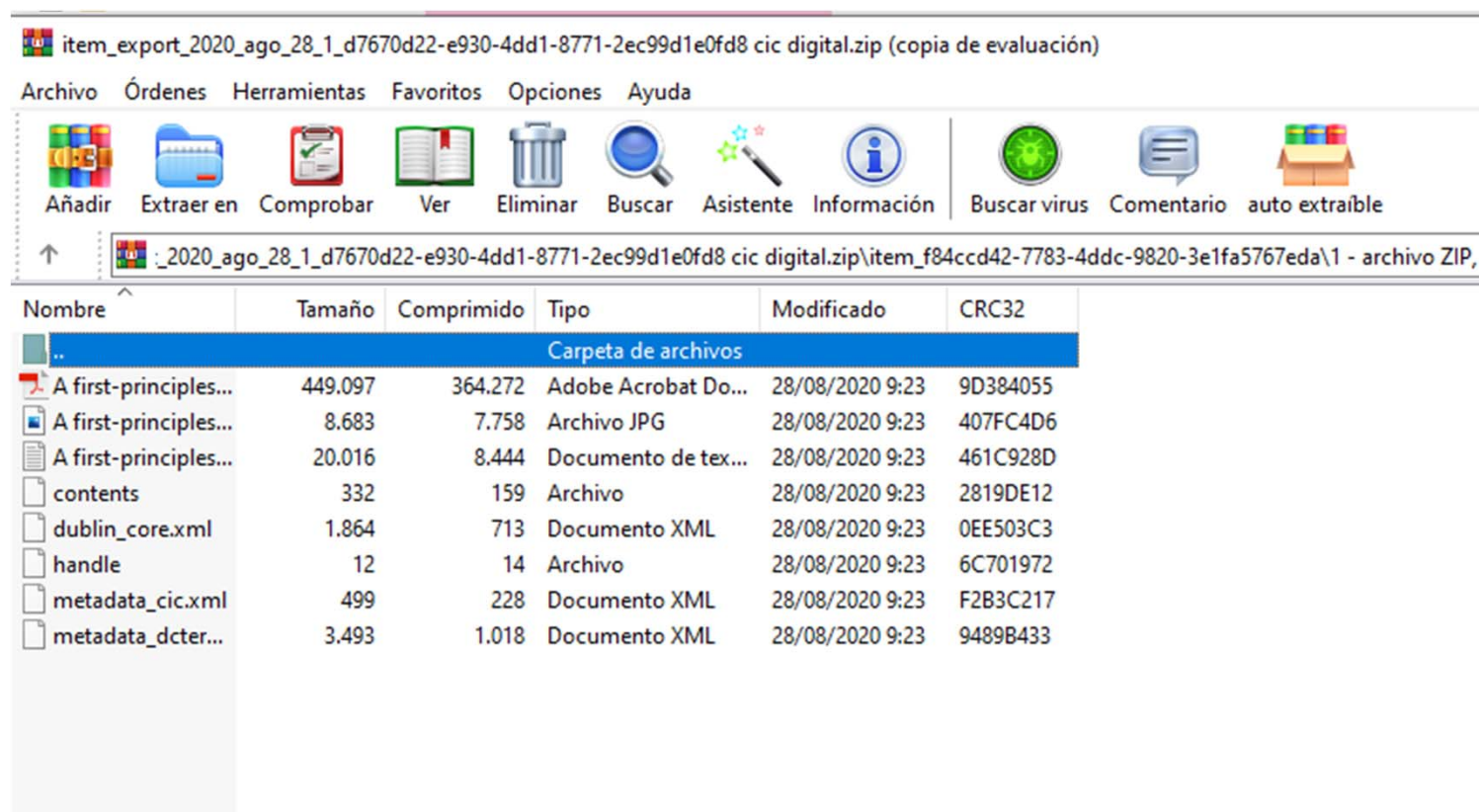
# Recepción de email con enlace al item exportado en SAF



The screenshot shows the Gmail interface. On the left, the navigation menu includes 'Redactar', 'Recibidos' (647), 'Destacados', 'Pospuestos', 'Importantes', and 'Meet' (with options to 'Iniciar una reunión' and 'Unirse a una reunión'). The main area displays an email from 'Repositorio CIC-DIGITAL' with the subject 'CIC-Digital - Exportación del Item solicitado está lista para descargar'. The email body contains the text: 'La exportación del item que ha solicitado desde el repositorio está lista para descargar. Puede descargar el archivo comprimido usando el siguiente link: [https://digital.cic.gba.gob.ar/exportdownload/item\\_export\\_2020\\_ago\\_28\\_1\\_d7670d22-e930-4dd1-8771-2ec99d1e0fd8.zip](https://digital.cic.gba.gob.ar/exportdownload/item_export_2020_ago_28_1_d7670d22-e930-4dd1-8771-2ec99d1e0fd8.zip) Este archivo permanecerá disponible para su descarga por al menos 48 horas.'



# SAF exportado desde la interfaz de CIC Digital



# Muchas gracias!



**Todos nuestros trabajos están abiertos en la colección:**

**<http://sedici.unlp.edu.ar/handle/10915/25293>**

**marisa.degiusti@sedici.unlp.edu.ar**