Cómo gestionar los datos de investigación: buenas prácticas, estándares, herramientas

Isabel Bernal, Juan Román Molina. Oficina Técnica de DIGITAL.CSIC

isabel.bernal@bib.csic.es, j.roman@bib.csic.es

19 - 21 de octubre / 2021, CSIC

DIGITAL.CSIC

CC BY NC SA

A free, open source, powerful tool forworking with messy data

High accuracy Data Anonymization

Pandoc

A universal document converter

Mr. Data Converter

I will convert your Excel data into one of several web-friendly formats

ZANRAN

Unlock the intelligence within your data

This is an experimental prototype that aims to render scientific papers in HTML so they can be more easily read by screen readers or on mobile devices

Conceptos

Datos abiertos: Data that are used as primary sources to support technical or scientific enquiry, research, scholarship, or artistic activity, and that are used as evidence in the research process and/or are commonly accepted in the research community as necessary to validate research findings and results. All other digital and non-digital content have the potential of becoming research data. Research data may be experimental data, observational data, operational data, third party data, public sector data, monitoring data, processed data, or repurposed data.

http://dictionary.casrai.org/Research_data

Etapas de la Ciencia del dato

- Gestión de datos: captura, depuración y almacenamiento de la información para su análisis

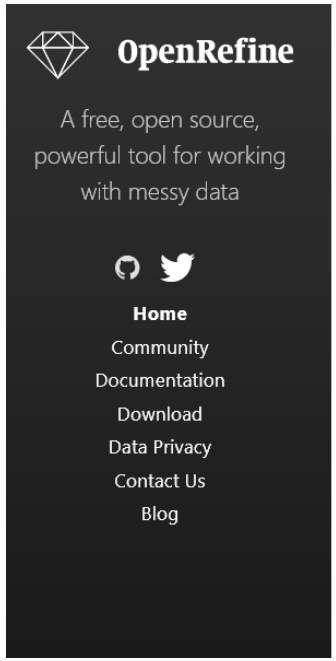
- Análisis de los datos

- Visualización: presentación de los resultados de forma clara, atractiva, concisa y comprensible. Simplificar los datos, darle forma visual (no textual).

DIGITAL.CSIC

Open Refine

1. Panel de trabajo / guías
2. Subir fichero
3. Crear proyecto
4. Editar dataset
5. Exportar proyecto
6. Exportar en otros formatos
7. Abrir / importar proyecto



Welcome!

OpenRefine (previously Google Refine) is a powerful tool for working with messy data: cleaning it; transforming it from one format into another; and extending it with web services and external data.

OpenRefine always keeps your data private on your own computer until YOU want to share or collaborate. Your private data never leaves your computer unless you want it to. (It works by running a small server on your computer and you use your web browser to interact with it)

OpenRefine is available in more than 15 languages.

OpenRefine is part of Code for Science & Society.

Introduction to OpenRefine

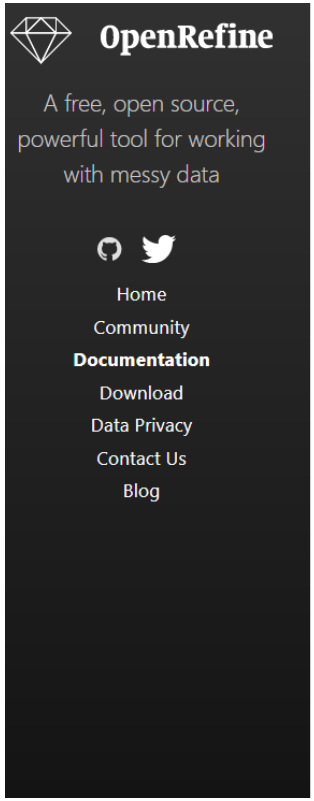
1. Explore Data

OpenRefine can help you explore large data sets with ease. You can find out more about this functionality by watching the video below.

https://openrefine.org

Open Refine

1. Panel de trabajo / guías



Documentation for users

- Official documentation
- FAQ

Documentation for developers

- Information for developers
- Contributor guidelines
- Developer discussion list

Support

- OpenRefine support mailing list
- StackOverflow tag
- File a bug report or feature request

Online courses

- OpenRefine Foundation Course (with video)
- DataCarpentry: OpenRefine for Social Science Data
- DataCarpentry: OpenRefine for Ecology
- The Programming Historian: Cleaning Data with OpenRef
- The Programming Historian: Fetching and Parsing Data fr OpenRefine
- Nettoyer et préparer des données avec OpenRefine : ateli consortium MASA - 14/11/2018 (fr)
- Curated Tutorial List

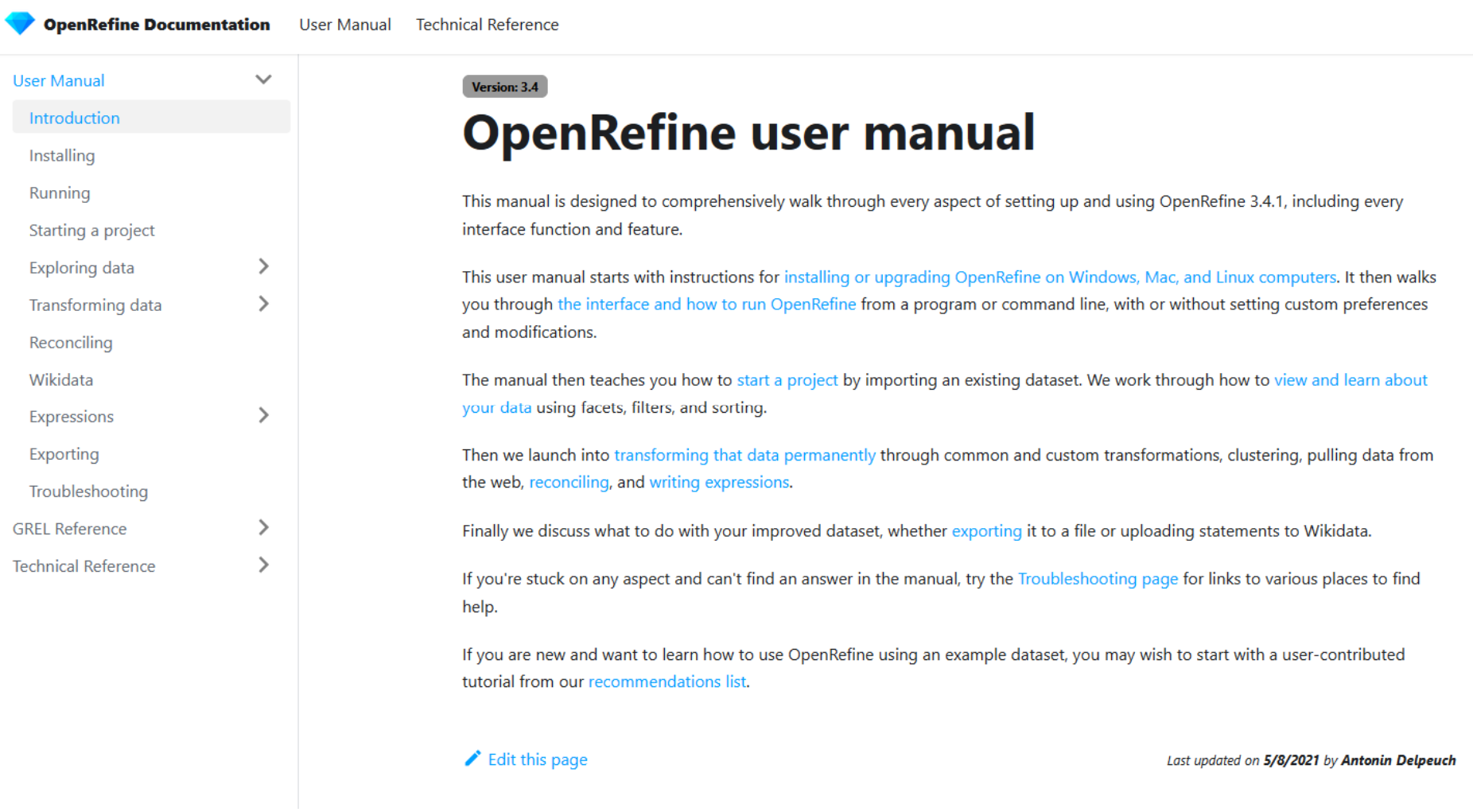
Using OpenRefine - The Book

Using OpenRefine by Ruben Verborgh, Max De Wilde offers anyone a great introduction to OpenRefine. Organized by recipes with hands on examples, the book covers the following topics:

1. Import data in various formats
2. Explore datasets in a matter of seconds
3. Apply basic and advanced cell transformations
4. Deal with cells that contain multiple values
5. Create instantaneous links between datasets
6. Filter and partition your data easily with regular expressions
7. Use named-entity extraction on full-text fields to automatically identify topics
8. Perform advanced data operations with the General Refine Expression Language

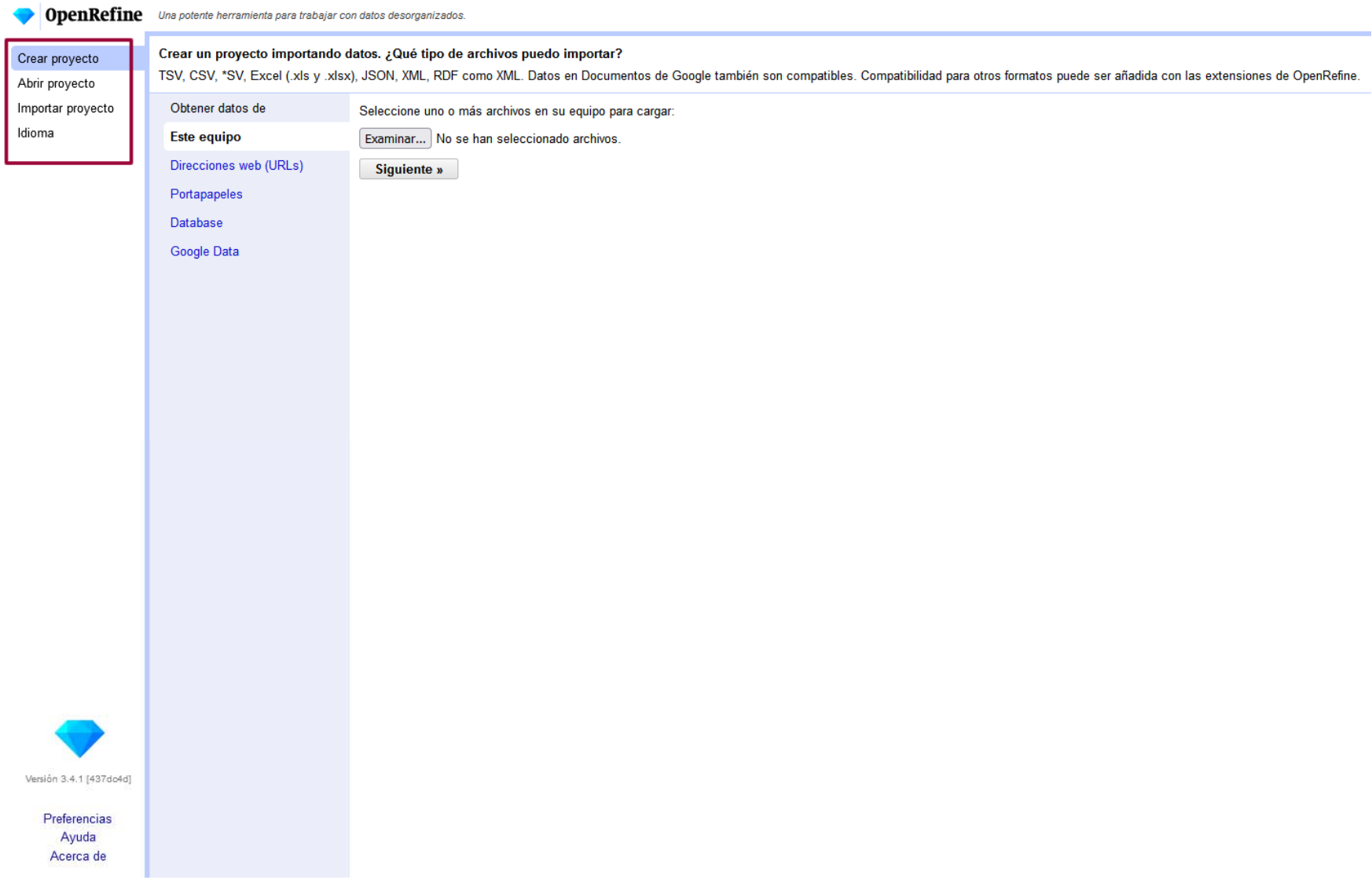
Open Refine

1. Panel de trabajo / guías



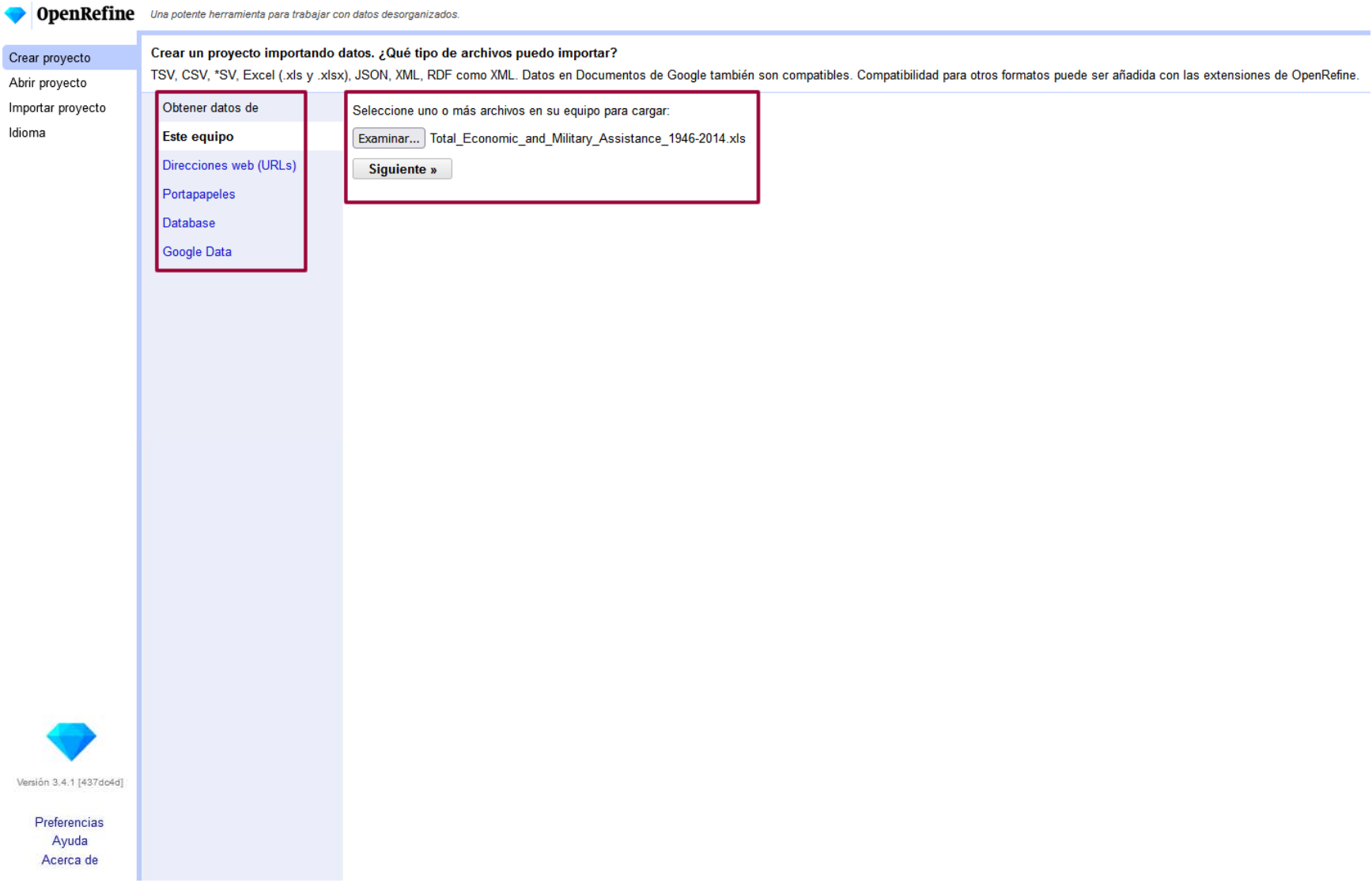
Open Refine

1. Panel de trabajo / guías



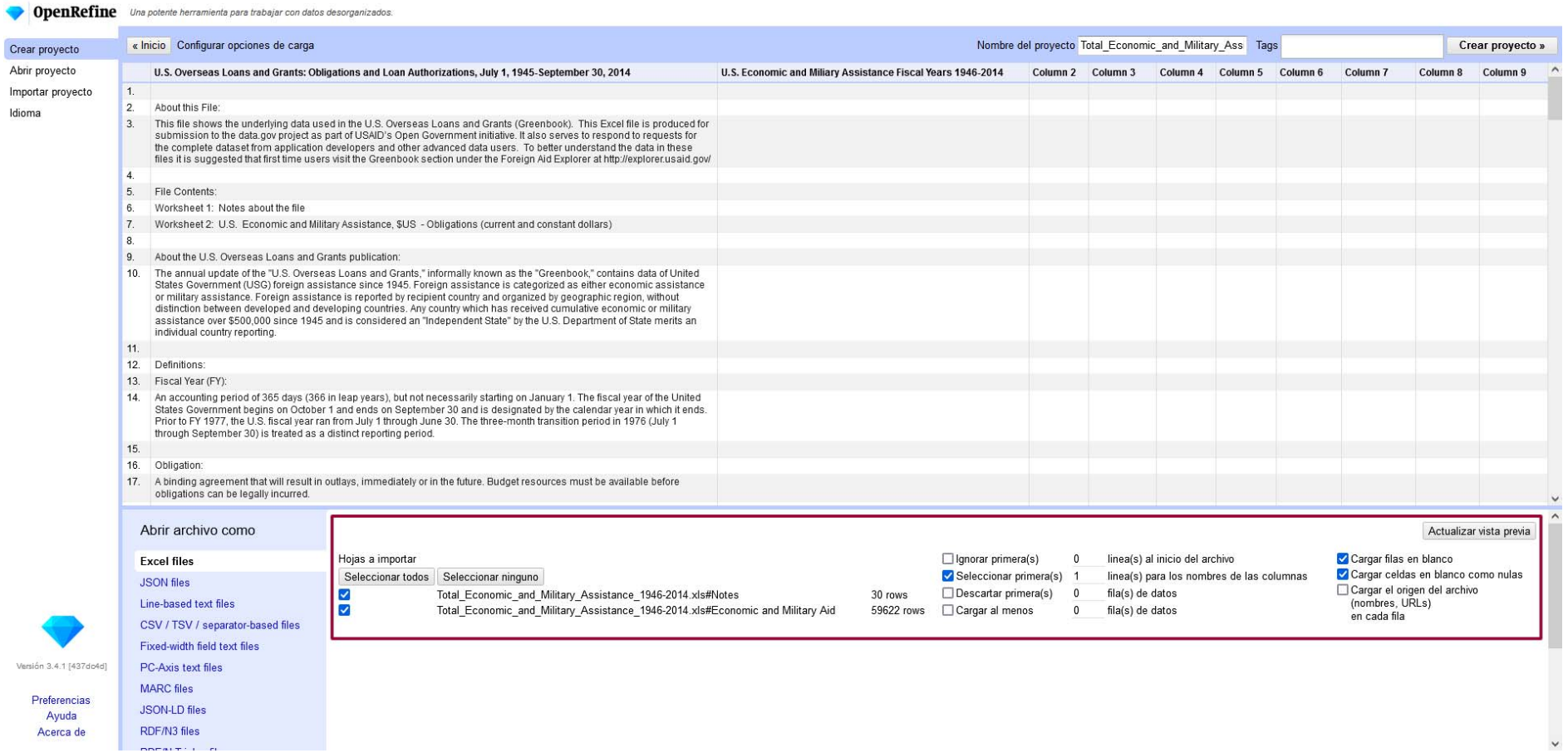
Open Refine

2. Subir fichero



Open Refine

2. Subir fichero



Open Refine

2. Subir fichero

Hojas a importar

Seleccionar todos | Seleccionar ninguno

☐	Total_Economic_and_Military_Assistance_1946-2014.xls#Notes	30 rows
☑	Total_Economic_and_Military_Assistance_1946-2014.xls#Economic and Military Aid	59622 rows

☐ Ignorar primera(s) 0 linea(s) al inicio del archivo

☑ Seleccionar primera(s) 1 linea(s) para los nombres de las columnas

☑ Descartar primera(s) 5 fila(s) de datos

☐ Cargar al menos 0 fila(s) de datos

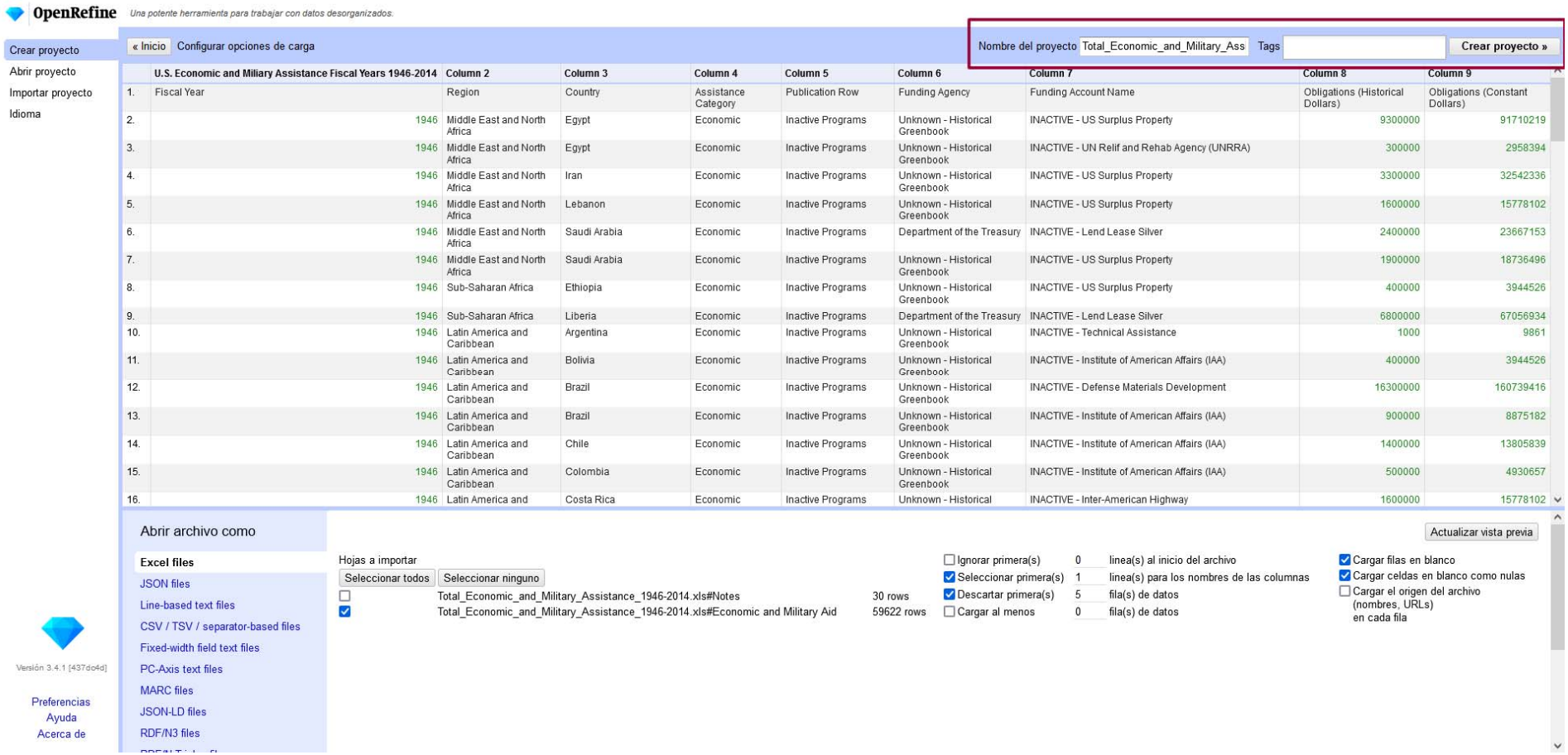
☑ Cargar filas en blanco

☑ Cargar celdas en blanco como nulas

☐ Cargar el origen del archivo (nombres, URLs) en cada fila

Open Refine

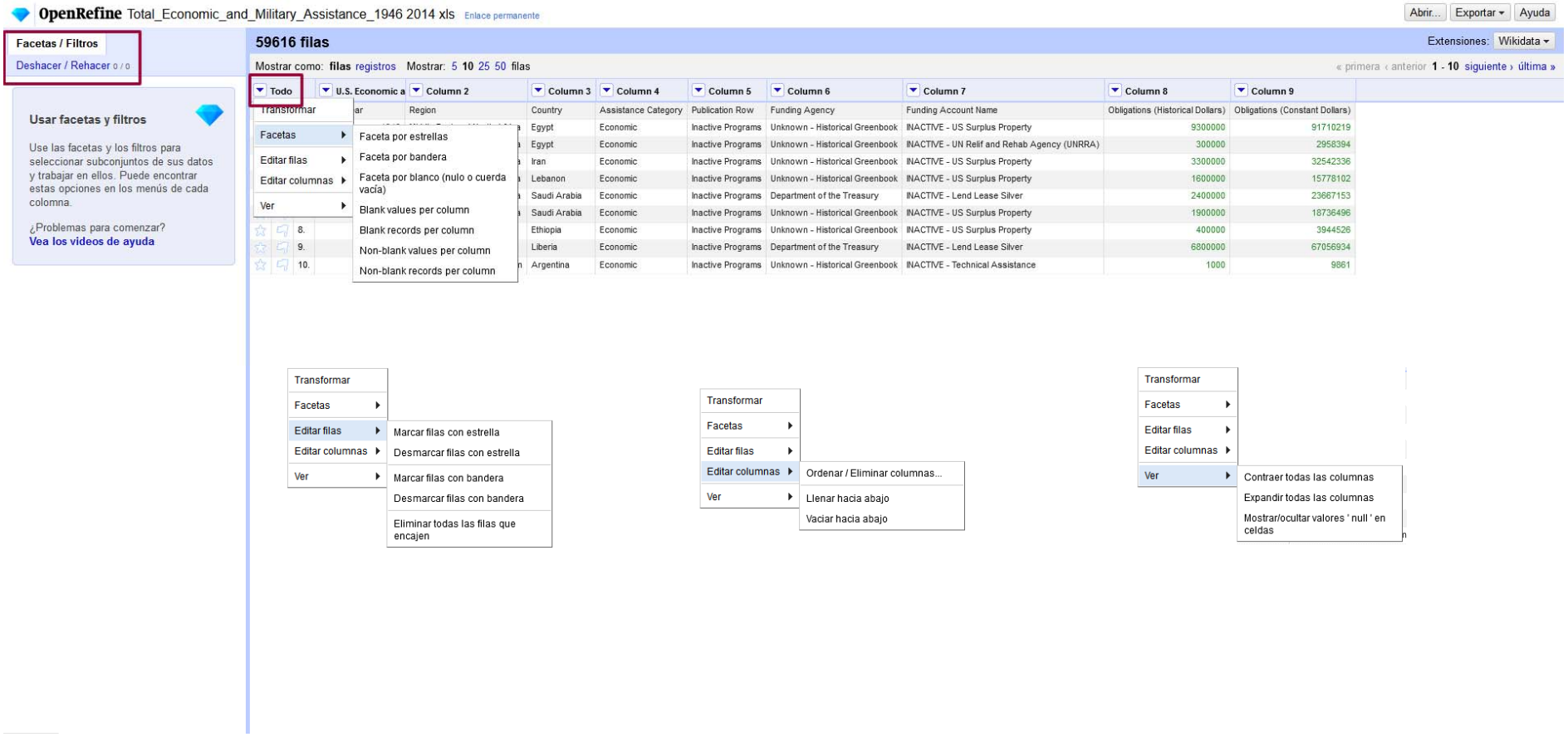
3. Crear proyecto



	U.S. Economic and Miliary Assistance Fiscal Years 1946-2014	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9
1.	Fiscal Year	Region	Country	Assistance Category	Publication Row	Funding Agency	Funding Account Name	Obligations (Historical Dollars)	Obligations (Constant Dollars)
2.	1946	Middle East and North Africa	Egypt	Economic	Inactive Programs	Unknown - Historical Greenbook	INACTIVE - US Surplus Property	9300000	91710219
3.	1946	Middle East and North Africa	Egypt	Economic	Inactive Programs	Unknown - Historical Greenbook	INACTIVE - UN Relif and Rehab Agency (UNRRA)	300000	2958394
4.	1946	Middle East and North Africa	Iran	Economic	Inactive Programs	Unknown - Historical Greenbook	INACTIVE - US Surplus Property	3300000	32542336
5.	1946	Middle East and North Africa	Lebanon	Economic	Inactive Programs	Unknown - Historical Greenbook	INACTIVE - US Surplus Property	1600000	15778102
6.	1946	Middle East and North Africa	Saudi Arabia	Economic	Inactive Programs	Department of the Treasury	INACTIVE - Lend Lease Silver	2400000	23667153
7.	1946	Middle East and North Africa	Saudi Arabia	Economic	Inactive Programs	Unknown - Historical Greenbook	INACTIVE - US Surplus Property	1900000	18736496
8.	1946	Sub-Saharan Africa	Ethiopia	Economic	Inactive Programs	Unknown - Historical Greenbook	INACTIVE - US Surplus Property	400000	3944526
9.	1946	Sub-Saharan Africa	Liberia	Economic	Inactive Programs	Department of the Treasury	INACTIVE - Lend Lease Silver	6800000	67056934
10.	1946	Latin America and Caribbean	Argentina	Economic	Inactive Programs	Unknown - Historical Greenbook	INACTIVE - Technical Assistance	1000	9861
11.	1946	Latin America and Caribbean	Bolivia	Economic	Inactive Programs	Unknown - Historical Greenbook	INACTIVE - Institute of American Affairs (IAA)	400000	3944526
12.	1946	Latin America and Caribbean	Brazil	Economic	Inactive Programs	Unknown - Historical Greenbook	INACTIVE - Defense Materials Development	16300000	160739416
13.	1946	Latin America and Caribbean	Brazil	Economic	Inactive Programs	Unknown - Historical Greenbook	INACTIVE - Institute of American Affairs (IAA)	900000	8875182
14.	1946	Latin America and Caribbean	Chile	Economic	Inactive Programs	Unknown - Historical Greenbook	INACTIVE - Institute of American Affairs (IAA)	1400000	13805839
15.	1946	Latin America and Caribbean	Colombia	Economic	Inactive Programs	Unknown - Historical Greenbook	INACTIVE - Institute of American Affairs (IAA)	500000	4930657
16.	1946	Latin America and	Costa Rica	Economic	Inactive Programs	Unknown - Historical	INACTIVE - Inter-American Highway	1600000	15778102

Open Refine

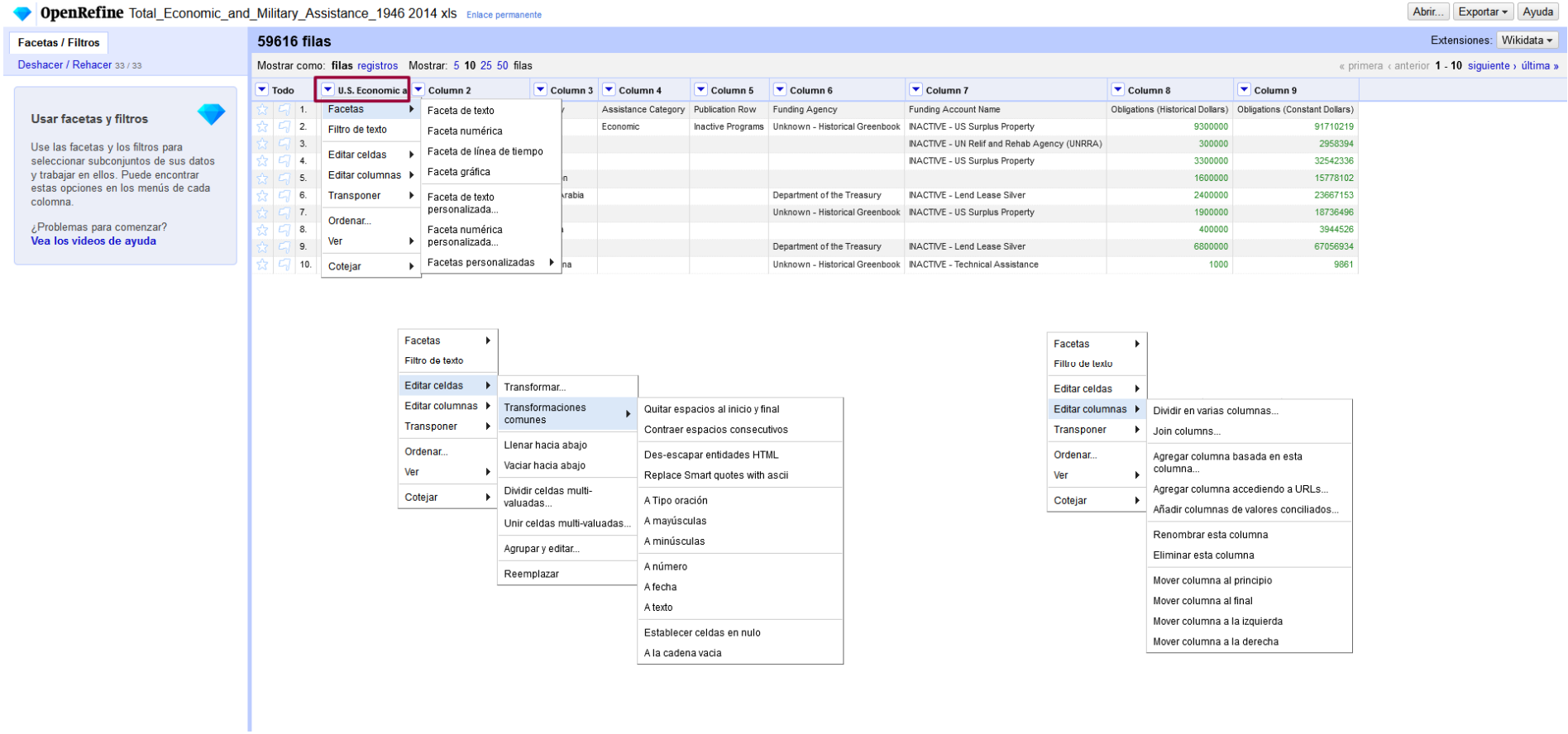
4. Editar dataset



DIGITAL.CSIC

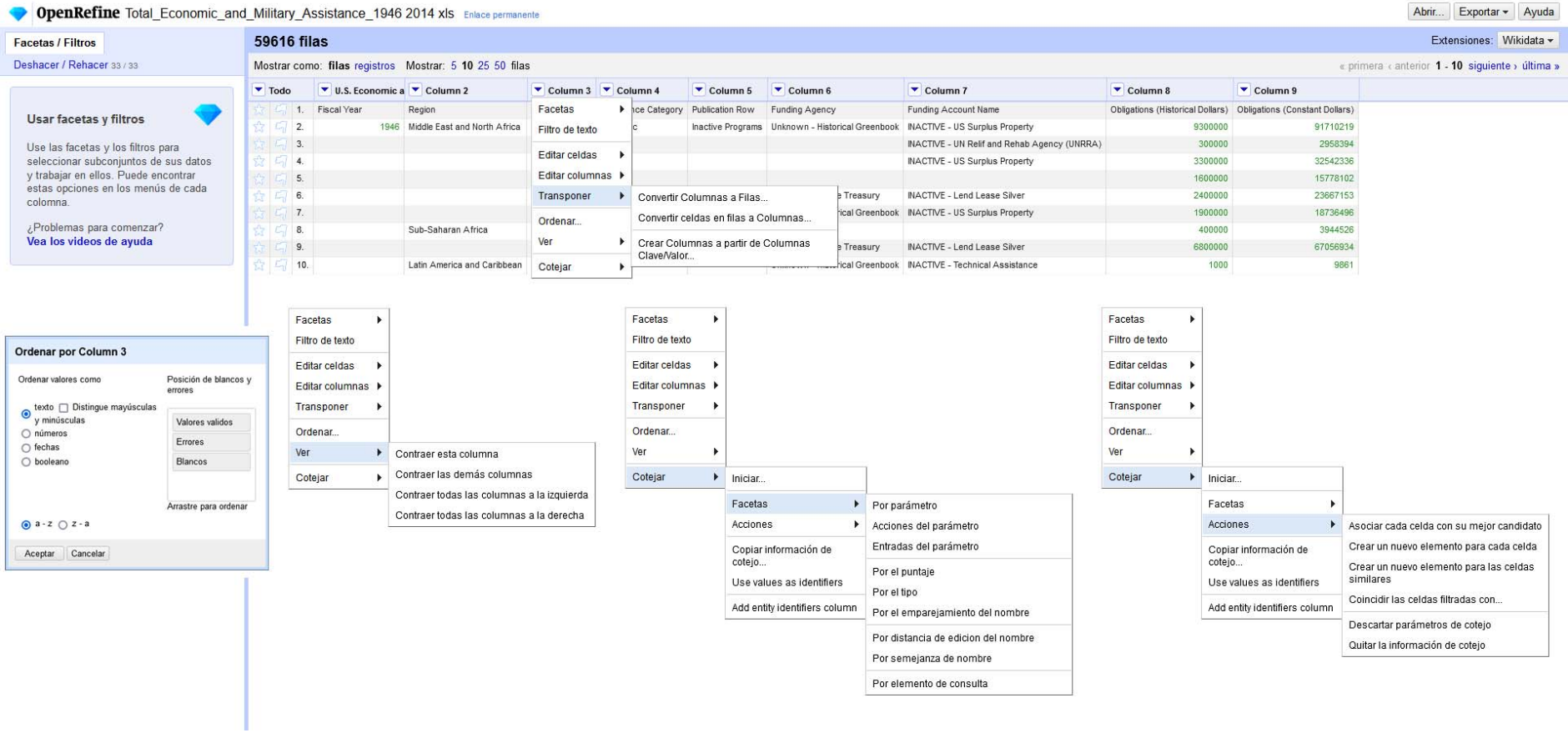
Open Refine

4. Editar dataset



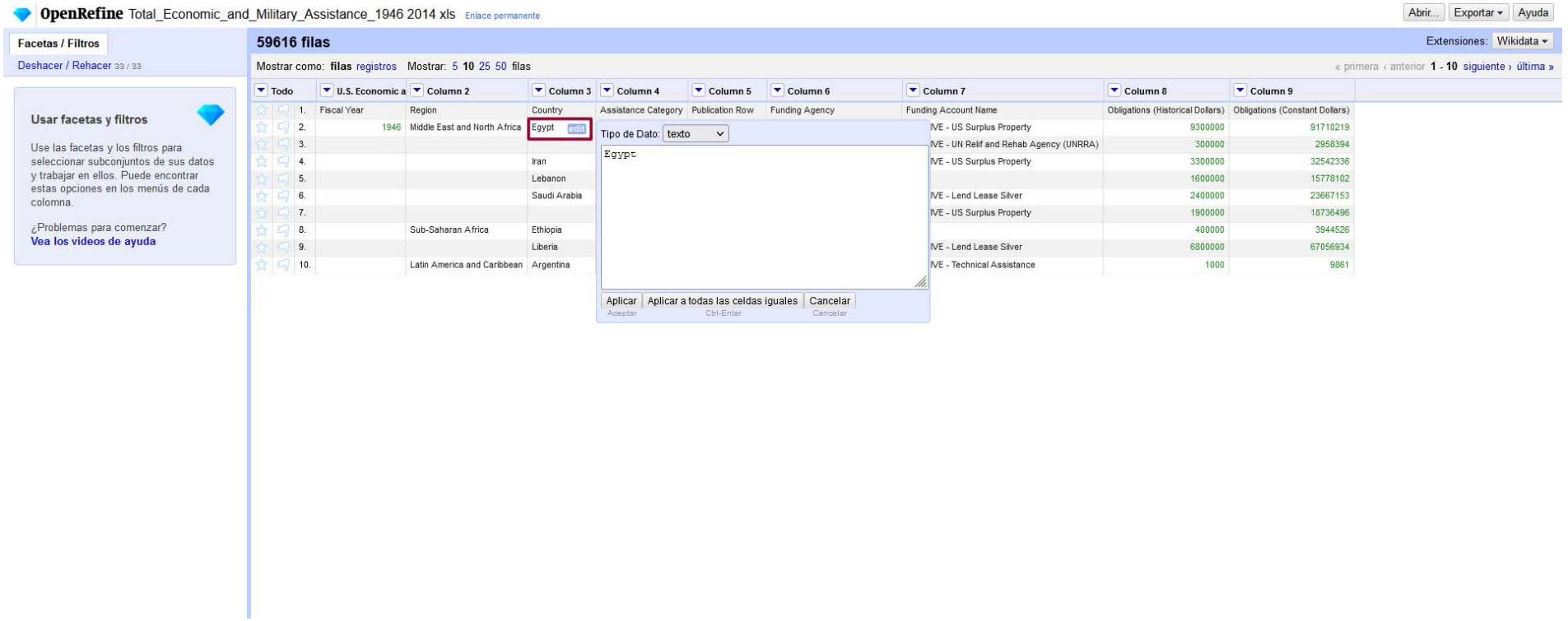
Open Refine

4. Editar dataset



Open Refine

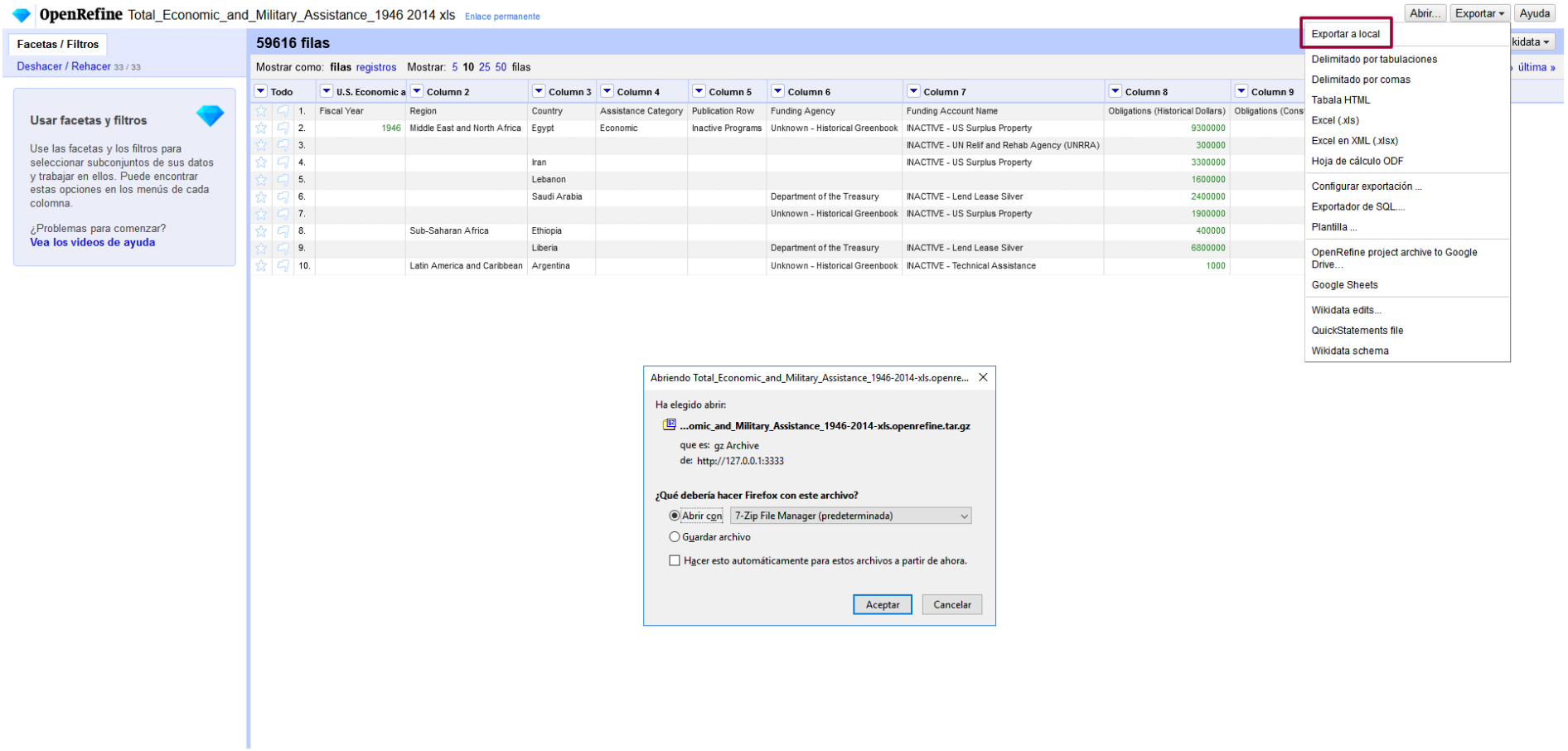
4. Editar dataset



DIGITAL.CSIC

Open Refine

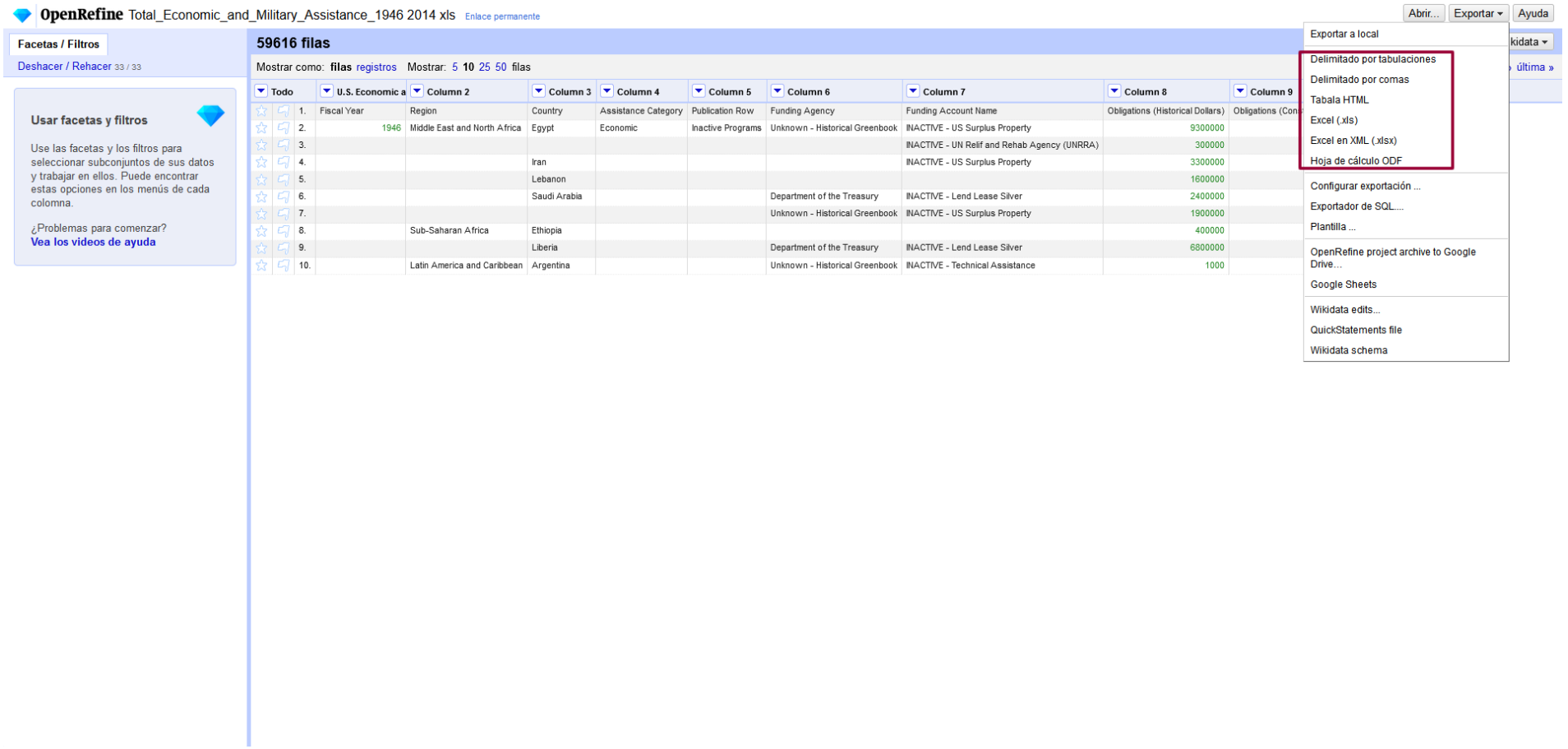
5. Exportar proyecto



Todo		U.S. Economic a	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9
	1.	Fiscal Year	Region	Country	Assistance Category	Publication Row	Funding Agency	Funding Account Name	Obligations (Historical Dollars)	Obligations (Cons
	2.	1946	Middle East and North Africa	Egypt	Economic	Inactive Programs	Unknown - Historical Greenbook	INACTIVE - US Surplus Property	9300000	
	3.							INACTIVE - UN Relif and Rehab Agency (UNRRA)	300000	
	4.			Iran				INACTIVE - US Surplus Property	3300000	
	5.			Lebanon					1600000	
	6.			Saudi Arabia			Department of the Treasury	INACTIVE - Lend Lease Silver	2400000	
	7.						Unknown - Historical Greenbook	INACTIVE - US Surplus Property	1900000	
	8.		Sub-Saharan Africa	Ethiopia					400000	
	9.			Liberia			Department of the Treasury	INACTIVE - Lend Lease Silver	6800000	
	10.		Latin America and Caribbean	Argentina			Unknown - Historical Greenbook	INACTIVE - Technical Assistance	1000	

Open Refine

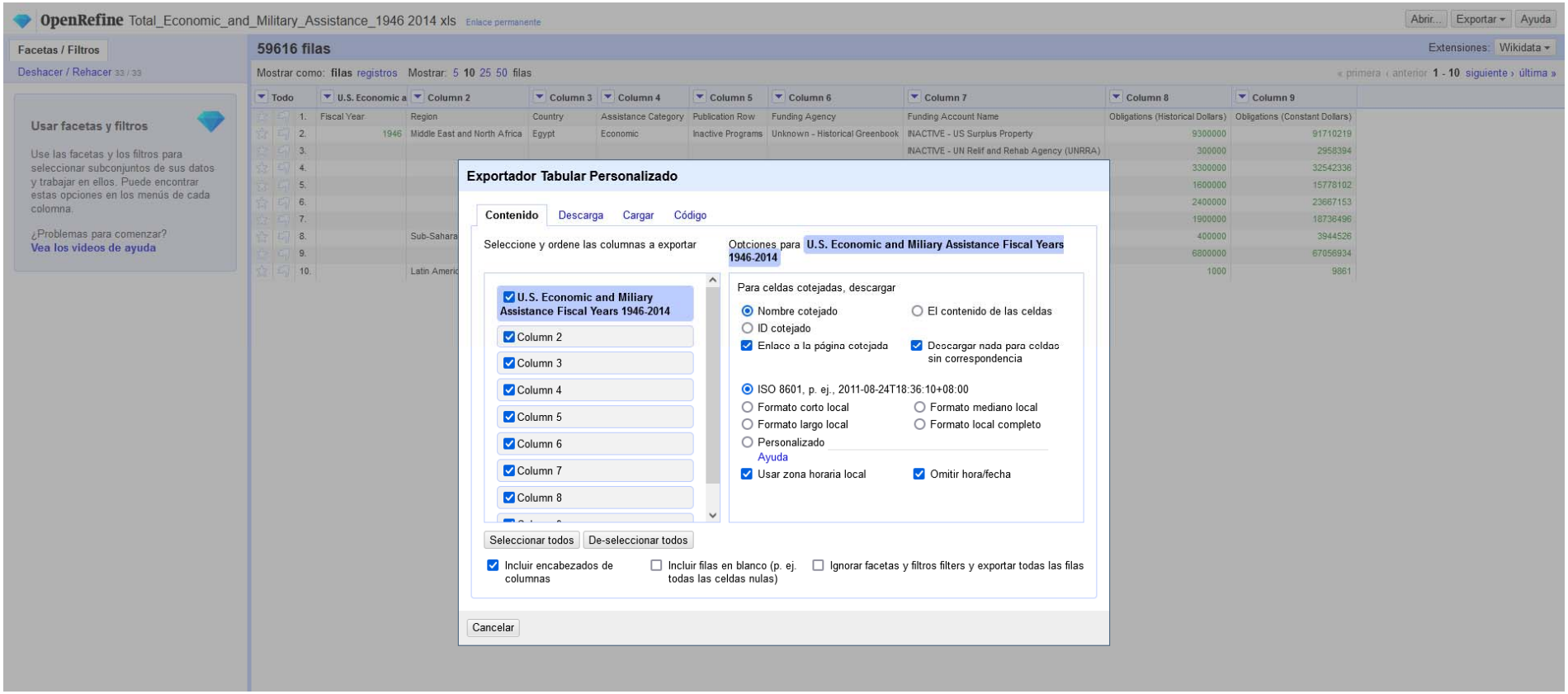
6. Exportar en otros formatos



DIGITAL.CSIC

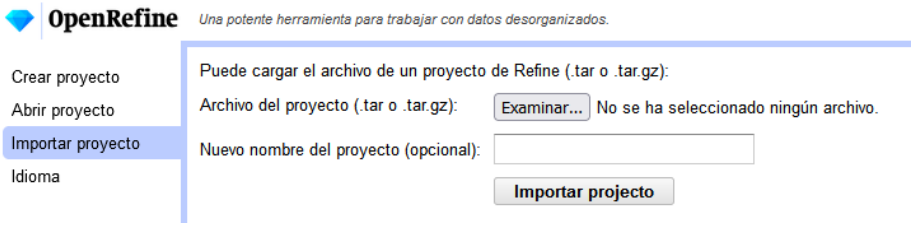
Open Refine

6. Exportar en otros formatos: personalizado



Open Refine

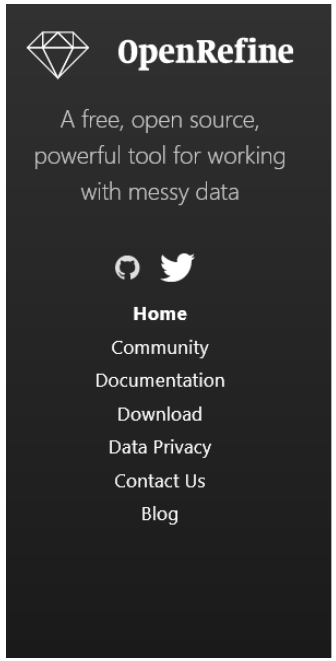
7. Abrir / importar proyecto



DIGITAL.CSIC

Cómo gestionar los datos de investigación: buenas prácticas, estándares, herramientas

PROGRAMA



Welcome!

OpenRefine (previously Google Refine) is a powerful tool for working with messy data: cleaning it; transforming it from one format into another; and extending it with web services and external data.

OpenRefine always keeps your data private on your own computer until YOU want to share or collaborate. Your private data never leaves your computer unless you want it to. (It works by running a small server on your computer and you use your web browser to interact with it)

OpenRefine is available in more than 15 languages.

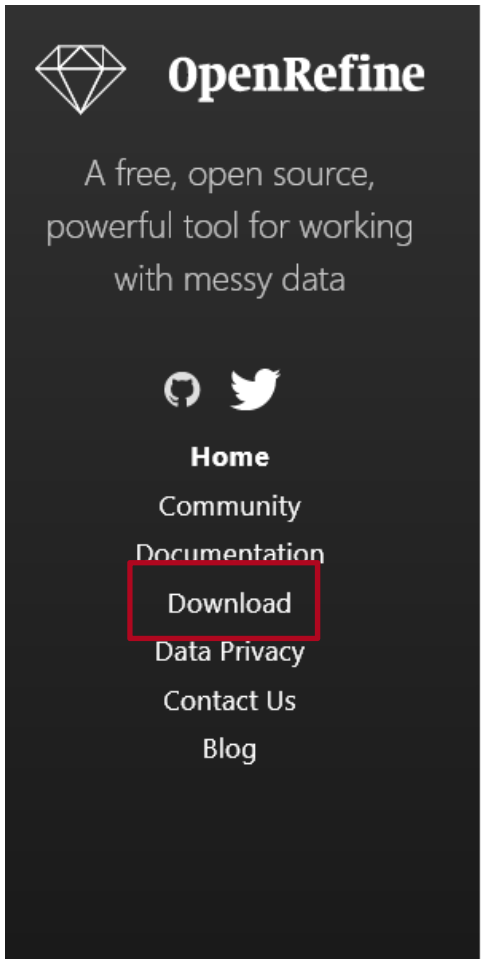
OpenRefine is part of Code for Science & Society.

Introduction to OpenRefine

1. Explore Data

OpenRefine can help you explore large data sets with ease. You can find out more about this functionality by watching the video below.

https://openrefine.org



Welcome!

OpenRefine (previously Google Refine) is a powerful tool for working with messy data: cleaning it; transforming it from one format into another; and extending it with web services and external data.

OpenRefine always keeps your data private on your own computer until YOU want to share or collaborate. Your private data never leaves your computer unless you want it to. (It works by running a small server on your computer and you use your web browser to interact with it)

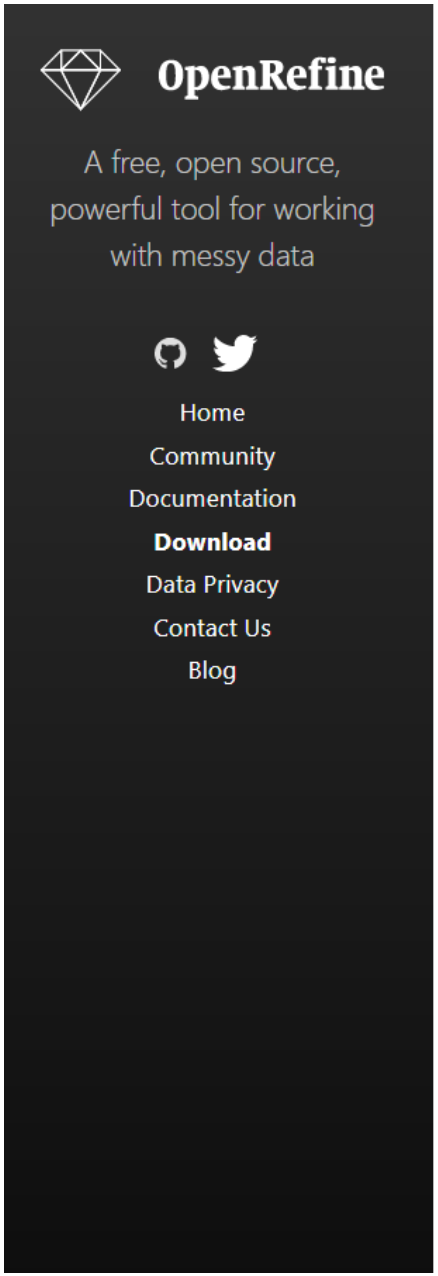
OpenRefine is available in more than 15 languages.

OpenRefine is part of Code for Science & Society.

Introduction to OpenRefine

1. Explore Data

OpenRefine can help you explore large data sets with ease. You can find out more about this functionality by watching the video below.



Download

You will find on this page a list of OpenRefine distributions and extensions available for download. Are we missing something? Want to fix a typo? You can submit changes (pull request) from here.

Official Distribution

Read the installation instructions.

You can also download all official releases and source from our GitHub releases page

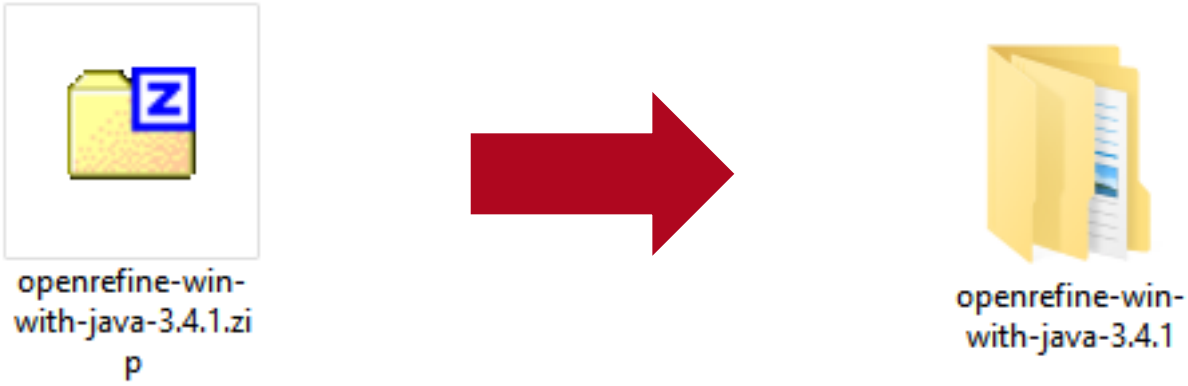
OpenRefine 3.4.1

The latest stable release of OpenRefine 3.4.1, released on September 24, 2020. Please backup your workspace directory before installing and report any problems that you encounter. A change log is provided on the release page.

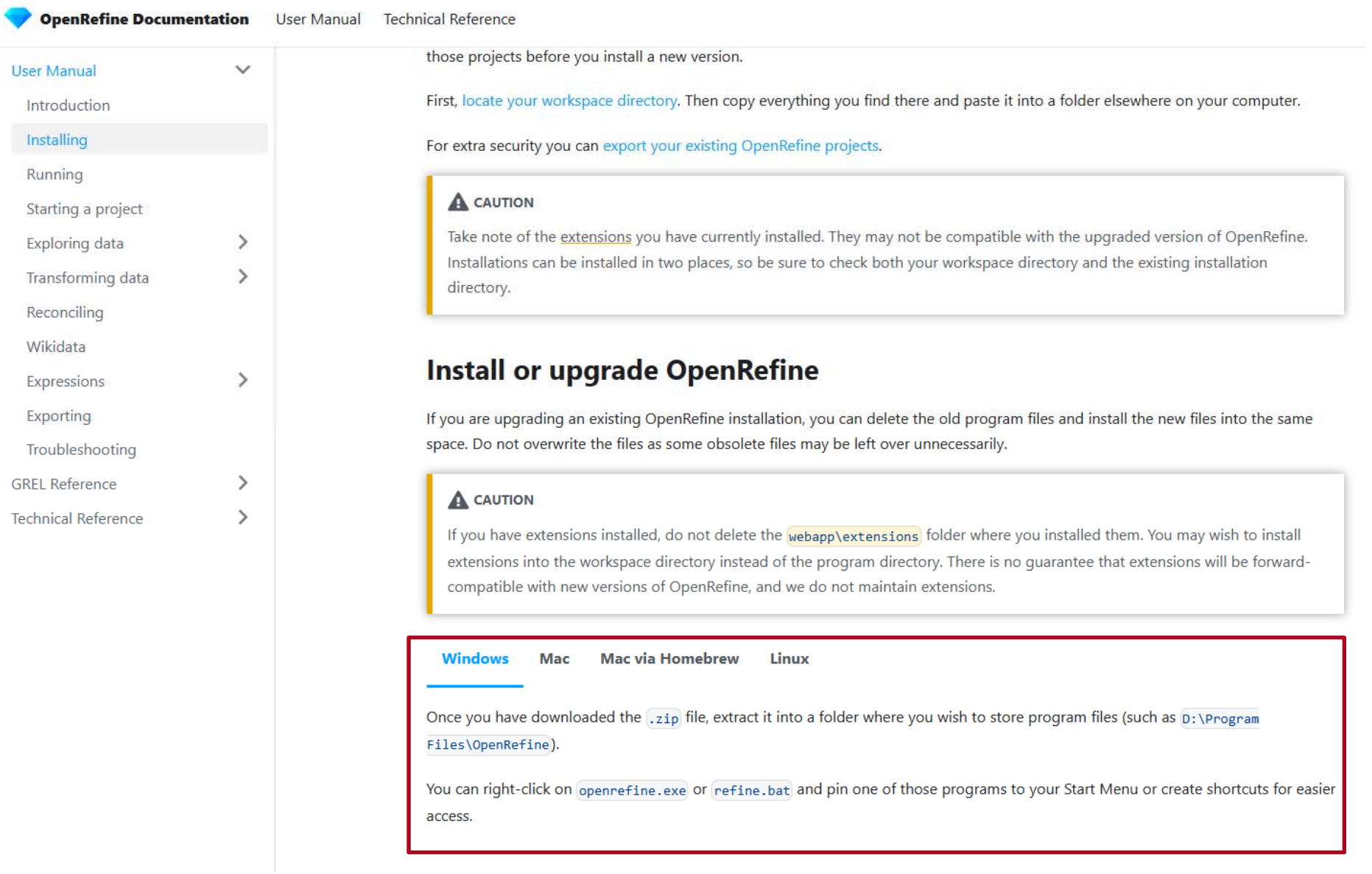
- **Windows kit**, This requires Java to be installed on your computer. Download, unzip, and double-click on *openrefine.exe* or *refine.bat* if the former does not work.
- **Windows kit with embedded Java**, includes OpenJDK Java, available under the GPLv2+CE license. Download, unzip, and double-click on *openrefine.exe* or *refine.bat* if the former does not work.
- **Mac kit**, Download, open, drag icon into the Applications folder and double click on it. You do not need to install Java separately.
- **Linux kit**, Download, extract, then type *./refine* to start. This requires Java to be installed on your computer.

OpenRefine 3.5 beta1

The first beta release of OpenRefine 3.5, released on May 29, 2021. Please backup your workspace directory before installing and report any problems that you encounter. A change log is provided on the release page.


openrefine-win-with-java-3.4.1.zip
openrefine-win-with-java-3.4.1

Cómo gestionar los datos de investigación: buenas prácticas, estándares, herramientas



OpenRefine Documentation | User Manual | Technical Reference

- User Manual
  - Introduction
  - Installing
  - Running
  - Starting a project
  - Exploring data
  - Transforming data
  - Reconciling
  - Wikidata
  - Expressions
  - Exporting
  - Troubleshooting
- GREL Reference
- Technical Reference

those projects before you install a new version.

First, locate your workspace directory. Then copy everything you find there and paste it into a folder elsewhere on your computer.

For extra security you can export your existing OpenRefine projects.

CAUTION

Take note of the extensions you have currently installed. They may not be compatible with the upgraded version of OpenRefine. Installations can be installed in two places, so be sure to check both your workspace directory and the existing installation directory.

Install or upgrade OpenRefine

If you are upgrading an existing OpenRefine installation, you can delete the old program files and install the new files into the same space. Do not overwrite the files as some obsolete files may be left over unnecessarily.

CAUTION

If you have extensions installed, do not delete the `webapp\extensions` folder where you installed them. You may wish to install extensions into the workspace directory instead of the program directory. There is no guarantee that extensions will be forward-compatible with new versions of OpenRefine, and we do not maintain extensions.

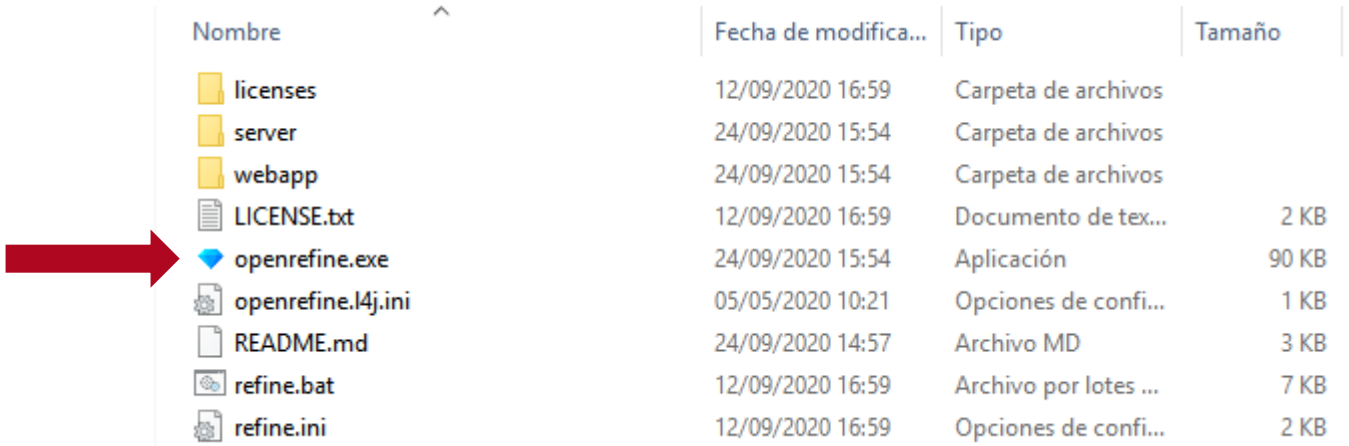
Windows | Mac | Mac via Homebrew | Linux

Once you have downloaded the `.zip` file, extract it into a folder where you wish to store program files (such as `D:\Program Files\OpenRefine`).

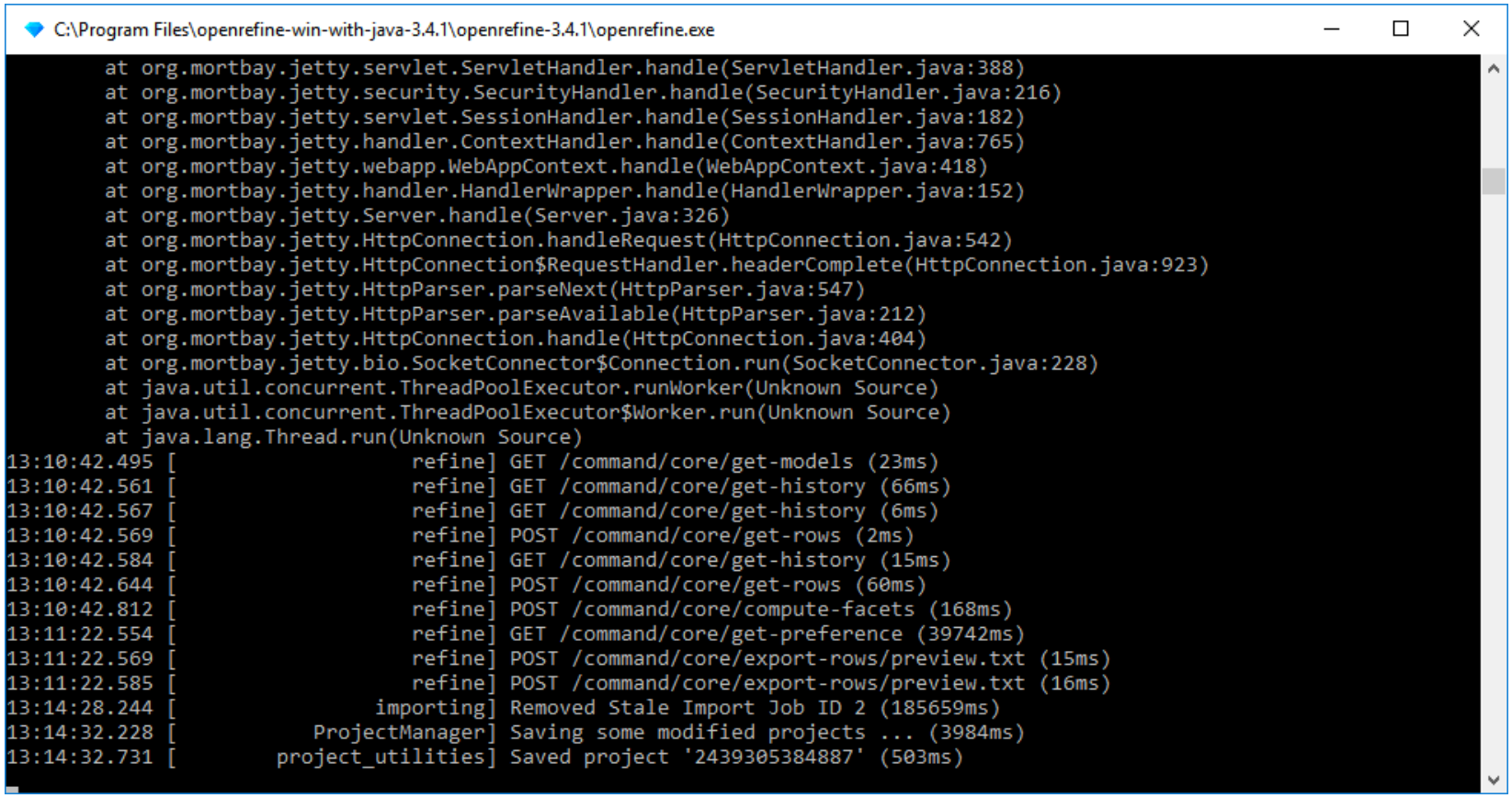
You can right-click on `openrefine.exe` or `refine.bat` and pin one of those programs to your Start Menu or create shortcuts for easier access.

DIGITAL.CSIC

Cómo gestionar los datos de investigación: buenas prácticas, estándares, herramientas



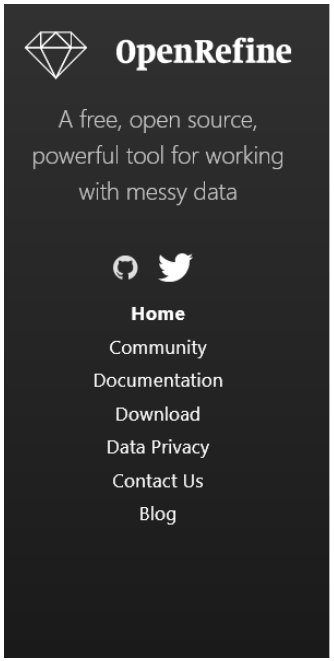
Nombre	Fecha de modifica...	Tipo	Tamaño
licenses	12/09/2020 16:59	Carpeta de archivos	
server	24/09/2020 15:54	Carpeta de archivos	
webapp	24/09/2020 15:54	Carpeta de archivos	
LICENSE.txt	12/09/2020 16:59	Documento de tex...	2 KB
openrefine.exe	24/09/2020 15:54	Aplicación	90 KB
openrefine.l4j.ini	05/05/2020 10:21	Opciones de confi...	1 KB
README.md	24/09/2020 14:57	Archivo MD	3 KB
refine.bat	12/09/2020 16:59	Archivo por lotes ...	7 KB
refine.ini	12/09/2020 16:59	Opciones de confi...	2 KB


C:\Program Files\openrefine-win-with-java-3.4.1\openrefine-3.4.1\openrefine.exe
at org.mortbay.jetty.servlet.ServletHandler.handle(ServletHandler.java:388)
at org.mortbay.jetty.security.SecurityHandler.handle(SecurityHandler.java:216)
at org.mortbay.jetty.servlet.SessionHandler.handle(SessionHandler.java:182)
at org.mortbay.jetty.handler.ContextHandler.handle(ContextHandler.java:765)
at org.mortbay.jetty.webapp.WebAppContext.handle(WebAppContext.java:418)
at org.mortbay.jetty.handler.HandlerWrapper.handle(HandlerWrapper.java:152)
at org.mortbay.jetty.Server.handle(Server.java:326)
at org.mortbay.jetty.HttpConnection.handleRequest(HttpConnection.java:542)
at org.mortbay.jetty.HttpConnection$RequestHandler.headerComplete(HttpConnection.java:923)
at org.mortbay.jetty.HttpParser.parseNext(HttpParser.java:547)
at org.mortbay.jetty.HttpParser.parseAvailable(HttpParser.java:212)
at org.mortbay.jetty.HttpConnection.handle(HttpConnection.java:404)
at org.mortbay.jetty.bio.SocketConnector$Connection.run(SocketConnector.java:228)
at java.util.concurrent.ThreadPoolExecutor.runWorker(Unknown Source)
at java.util.concurrent.ThreadPoolExecutor$Worker.run(Unknown Source)
at java.lang.Thread.run(Unknown Source)
13:10:42.495 [refine] GET /command/core/get-models (23ms)
13:10:42.561 [refine] GET /command/core/get-history (66ms)
13:10:42.567 [refine] GET /command/core/get-history (6ms)
13:10:42.569 [refine] POST /command/core/get-rows (2ms)
13:10:42.584 [refine] GET /command/core/get-history (15ms)
13:10:42.644 [refine] POST /command/core/get-rows (60ms)
13:10:42.812 [refine] POST /command/core/compute-facets (168ms)
13:11:22.554 [refine] GET /command/core/get-preference (39742ms)
13:11:22.569 [refine] POST /command/core/export-rows/preview.txt (15ms)
13:11:22.585 [refine] POST /command/core/export-rows/preview.txt (16ms)
13:14:28.244 [importing] Removed Stale Import Job ID 2 (185659ms)
13:14:32.228 [ProjectManager] Saving some modified projects ... (3984ms)
13:14:32.731 [project_utilities] Saved project '2439305384887' (503ms)

DIGITAL.CSIC

Cómo gestionar los datos de investigación: buenas prácticas, estándares, herramientas

PRÁCTICAS



Welcome!

OpenRefine (previously Google Refine) is a powerful tool for working with messy data: cleaning it; transforming it from one format into another; and extending it with web services and external data.

OpenRefine always keeps your data private on your own computer until YOU want to share or collaborate. Your private data never leaves your computer unless you want it to. (It works by running a small server on your computer and you use your web browser to interact with it)

OpenRefine is available in more than 15 languages.

OpenRefine is part of Code for Science & Society.

Introduction to OpenRefine

1. Explore Data

OpenRefine can help you explore large data sets with ease. You can find out more about this functionality by watching the video below.

https://openrefine.org

OpenRefine

valoresclimatologicos_bilbao-aeropuerto-csv
valoresclimatologicos_bilbao-aeropuerto-csv.xls

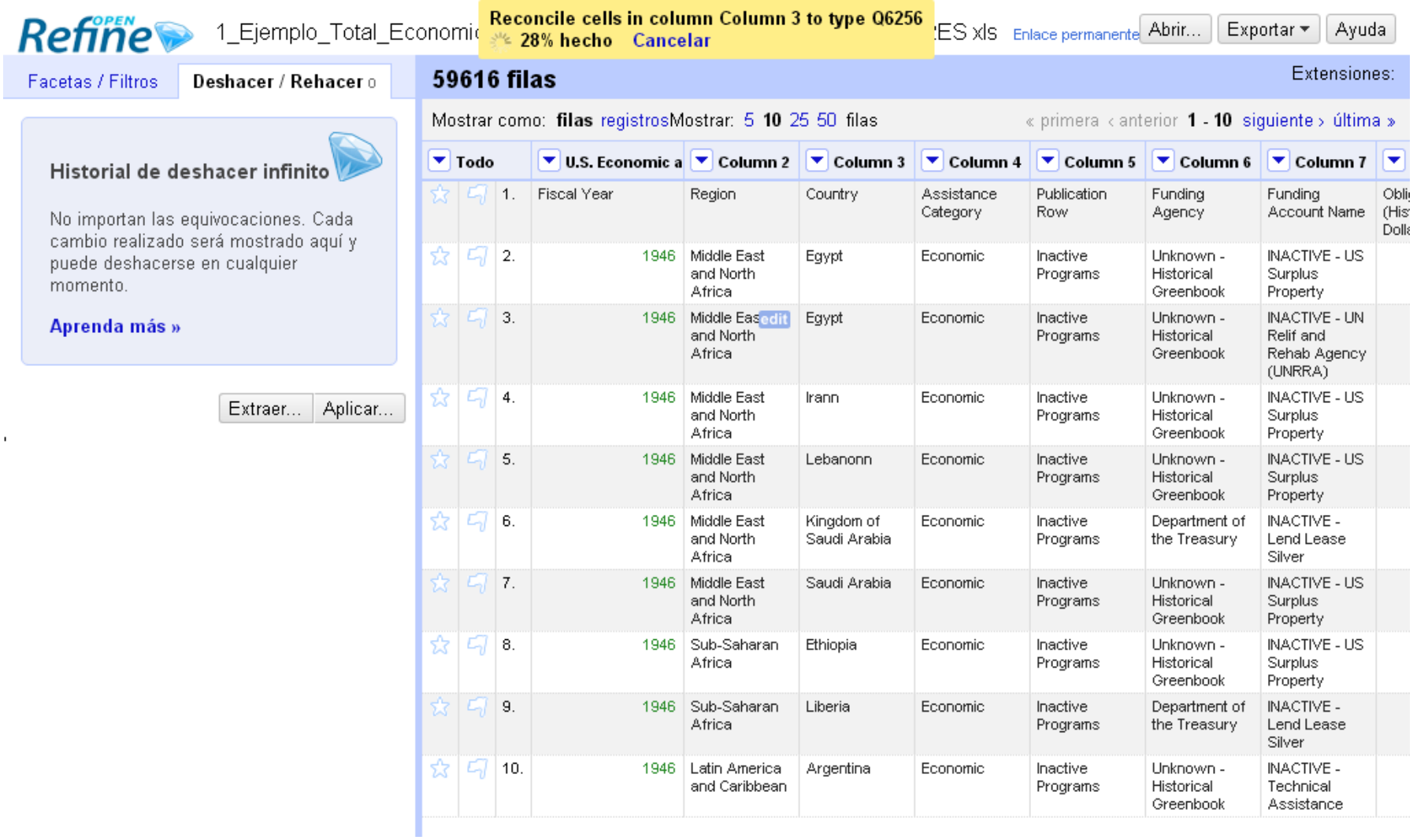
1. Primera toma de contacto con la herramienta
2. Comprobar las funcionalidades principales
3. Formatos xls y csv

Total_Economic_and_Military_Assistance_1946-2014.xls

1. Limpieza previa del dataset
2. Crear proyecto
3. Comprobar columna de países
4. Visualiza la ayuda por años
5. ¿Cuántos tipos de asistencia aparecen en el dataset?
6. ¿Existen países duplicados?
7. Detecta errores en “Chile” debido a las mayúsculas/minúsculas
8. Eliminar espacios vacíos en la columna “Países”, ¿alguna corección?
9. Renombra todas las columnas
10. Elimina la columna “Region”
11. Mueve las dos últimas columnas a continuación de “Assistance category”
12. Ordena por año más reciente
13. Ordena alfabéticamente los países
14. Contrae alguna columna y expándela posteriormente
15. Coteja la columna “Países”
16. Corrige celdas erróneas en países
17. Guarda el excel manipulado
18. Guarda el proyecto (gz)
19. Configura la exportación a tu gusto
20. Importa el proyecto

Cómo gestionar los datos de investigación: buenas prácticas, estándares, herramientas

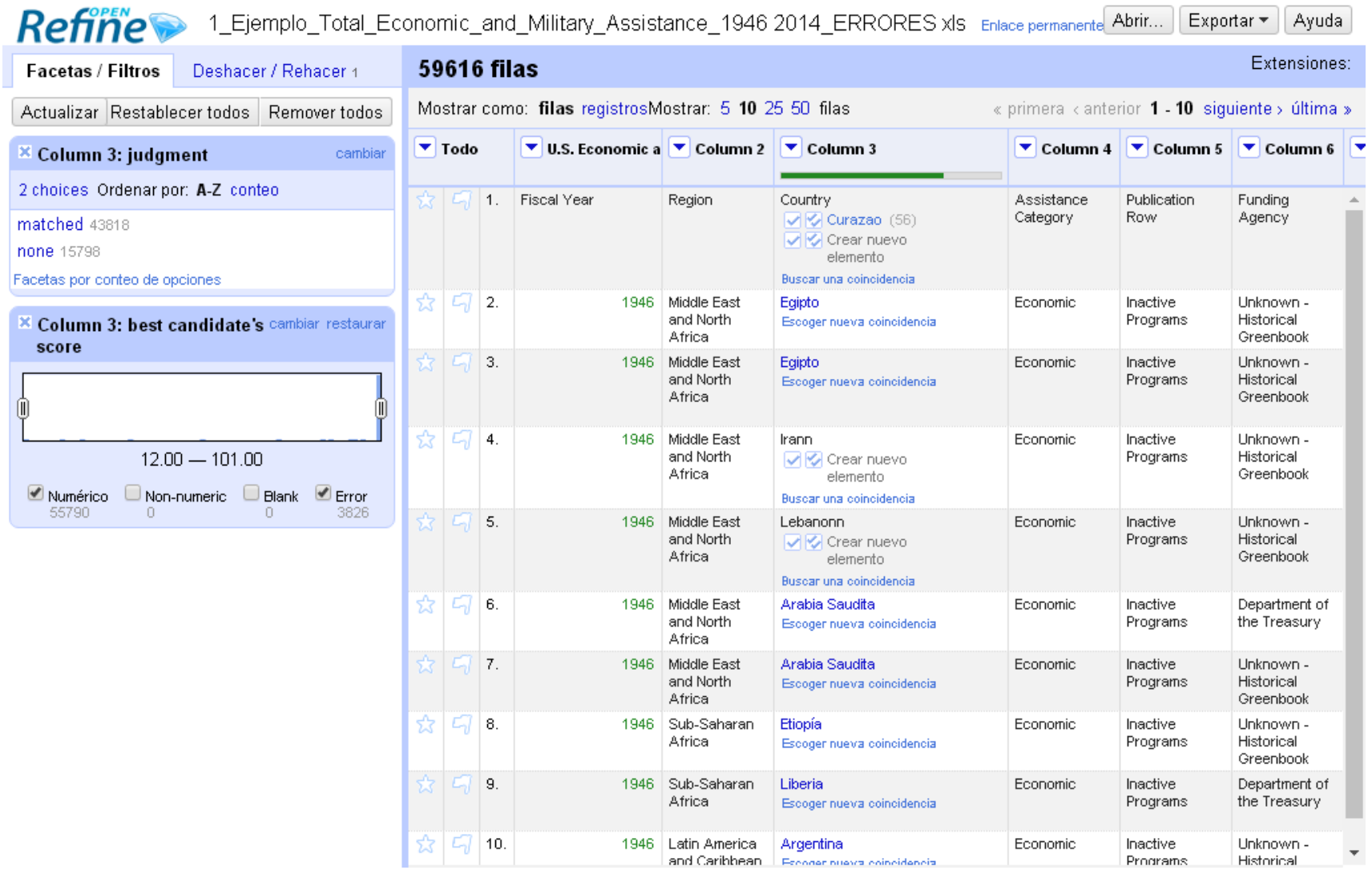
OpenRefine



Todo		U.S. Economic a	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	
1.	Fiscal Year		Region	Country	Assistance Category	Publication Row	Funding Agency	Funding Account Name	Obli (His Doll
2.	1946		Middle East and North Africa	Egypt	Economic	Inactive Programs	Unknown - Historical Greenbook	INACTIVE - US Surplus Property	
3.	1946		Middle Eas edit and North Africa	Egypt	Economic	Inactive Programs	Unknown - Historical Greenbook	INACTIVE - UN Relif and Rehab Agency (UNRRA)	
4.	1946		Middle East and North Africa	Irann	Economic	Inactive Programs	Unknown - Historical Greenbook	INACTIVE - US Surplus Property	
5.	1946		Middle East and North Africa	Lebanonn	Economic	Inactive Programs	Unknown - Historical Greenbook	INACTIVE - US Surplus Property	
6.	1946		Middle East and North Africa	Kingdom of Saudi Arabia	Economic	Inactive Programs	Department of the Treasury	INACTIVE - Lend Lease Silver	
7.	1946		Middle East and North Africa	Saudi Arabia	Economic	Inactive Programs	Unknown - Historical Greenbook	INACTIVE - US Surplus Property	
8.	1946		Sub-Saharan Africa	Ethiopia	Economic	Inactive Programs	Unknown - Historical Greenbook	INACTIVE - US Surplus Property	
9.	1946		Sub-Saharan Africa	Liberia	Economic	Inactive Programs	Department of the Treasury	INACTIVE - Lend Lease Silver	
10.	1946		Latin America and Caribbean	Argentina	Economic	Inactive Programs	Unknown - Historical Greenbook	INACTIVE - Technical Assistance	

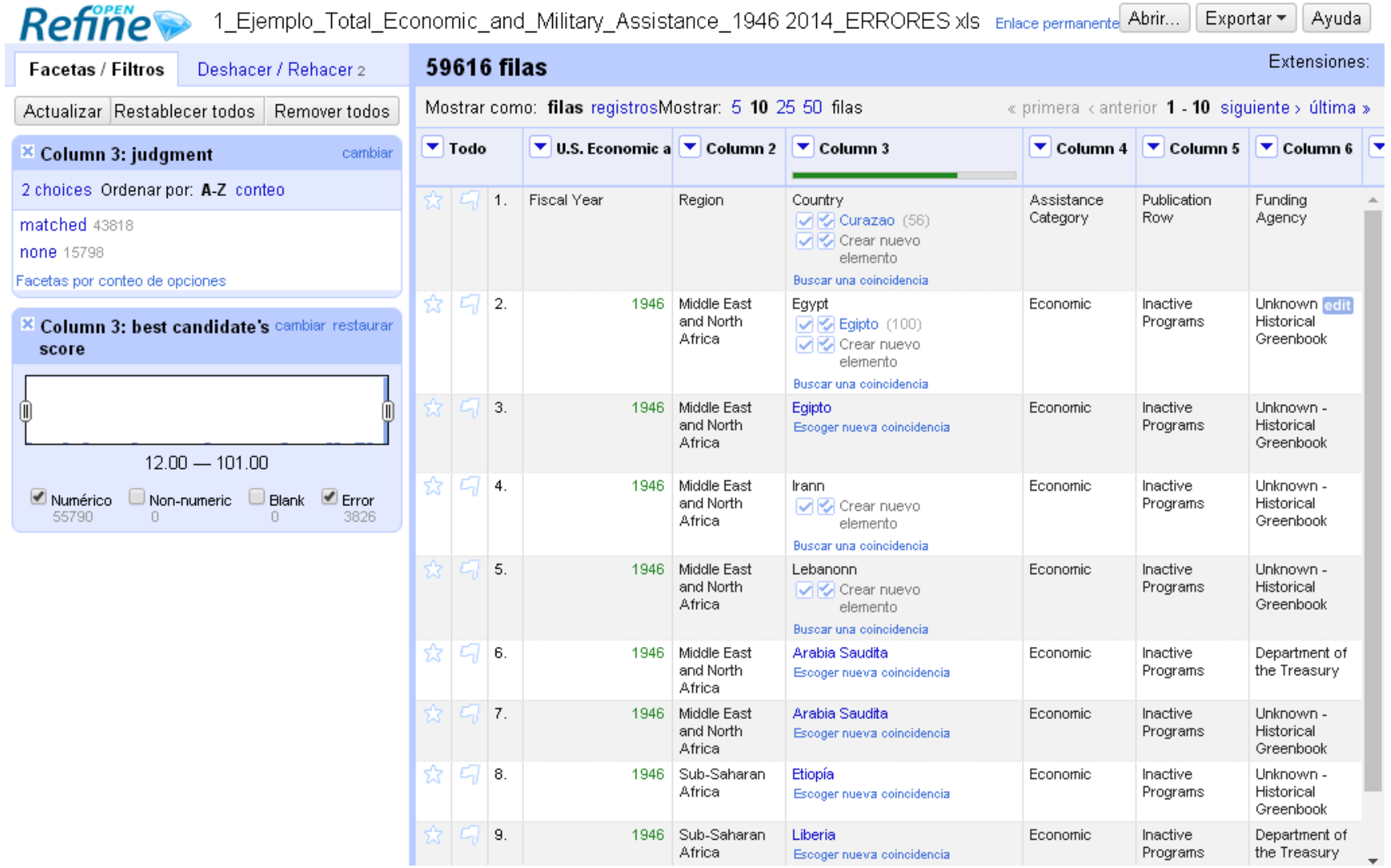
Cómo gestionar los datos de investigación: buenas prácticas, estándares, herramientas

OpenRefine



Cómo gestionar los datos de investigación: buenas prácticas, estándares, herramientas

OpenRefine



OpenRefine

tesis_univ_pub_2016.xls

1. Usar el fichero xls
2. Limpieza del dataset
3. Crear proyecto
4. Renombra las columnas
5. Agrupa las universidades de Andalucía
6. Trabaja con el dataset previa eliminación de “universidades públicas especiales”
7. Mueve las columnas de África al final del dataset
8. Comprueba que no hay universidades repetidas
9. Busca una universidad en concreto
10. Lleva a cabo algunas transformaciones comunes
11. ¿En qué universidad se han presentado más tesis en 2016?
12. Muestra únicamente las columnas de universidades, totales, España y UE 28
13. Edita celdas erróneas
14. Exporta el proyecto (gz), excel y por medio de la exportación personalizada

---

tesis_univ_pub_2016.csv

1. Limpia el dataset en formato csv
2. Trabaja con una selección de universidades a tu libre elección

OpenRefine

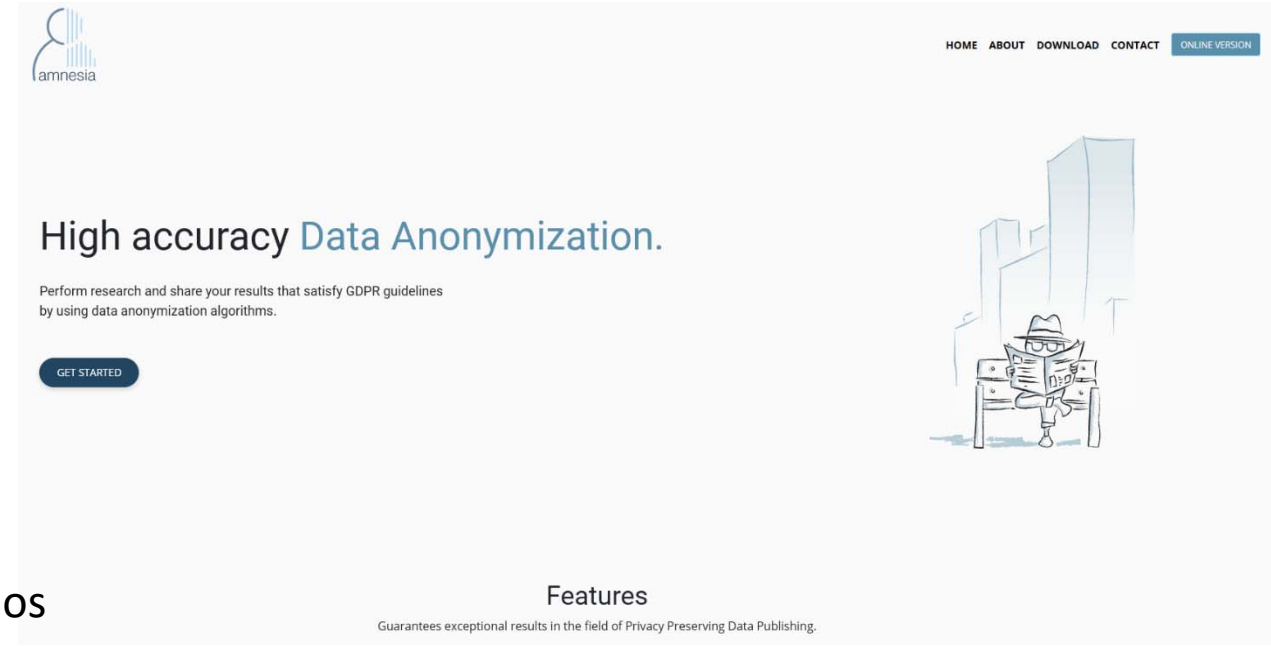
Ejercicio libre **ElectedOfficials.xls**

Trabaja el fichero usando las opciones vistas en ejercicios anteriores

DIGITAL.CSIC

Amnesia

1. Panel de trabajo / guías
2. Subir fichero
3. Generar jerarquías
4. Anonimizar los datos
5. Guardar datos anonimizados



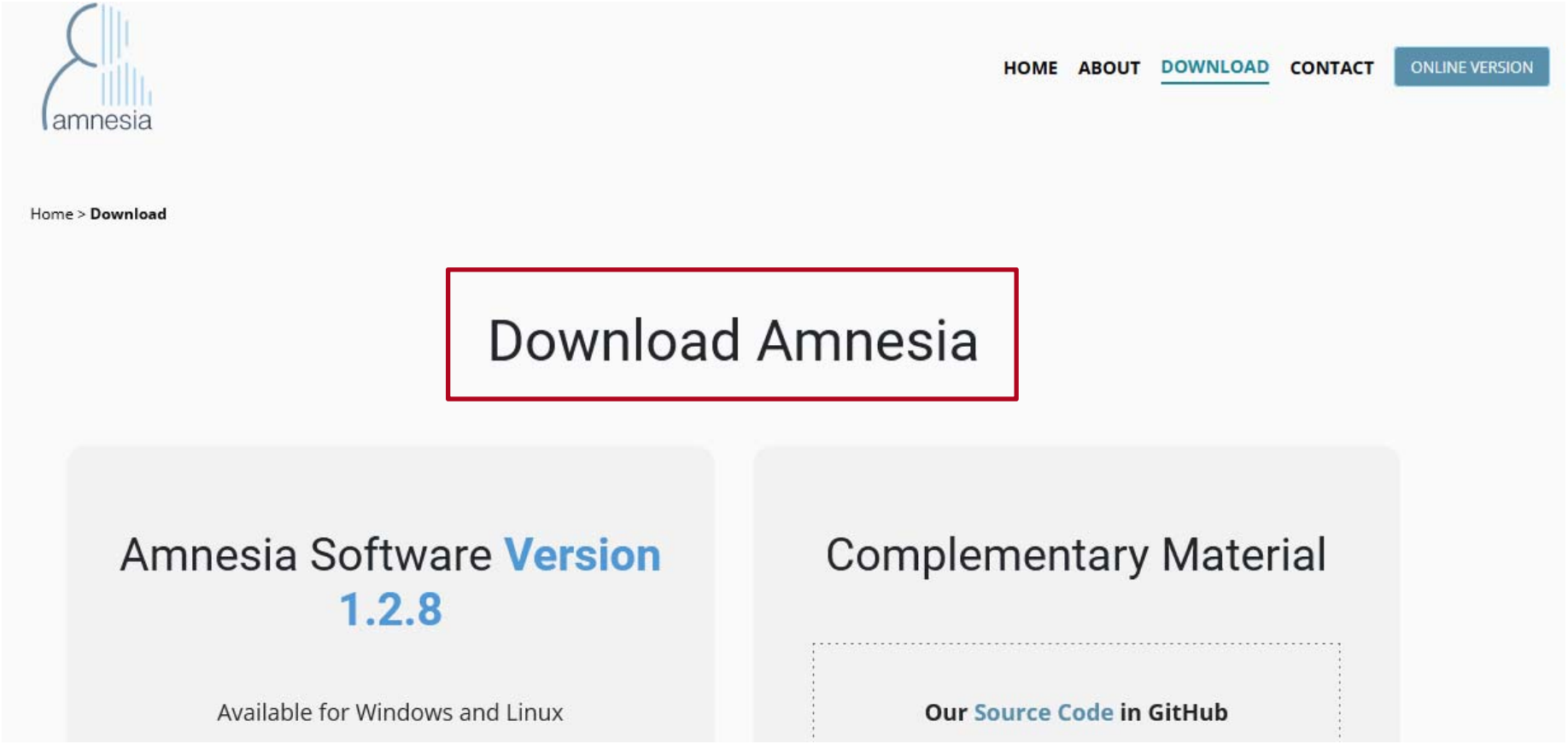
https://amnesia.openaire.eu

Amnesia

1. Panel de trabajo / guías

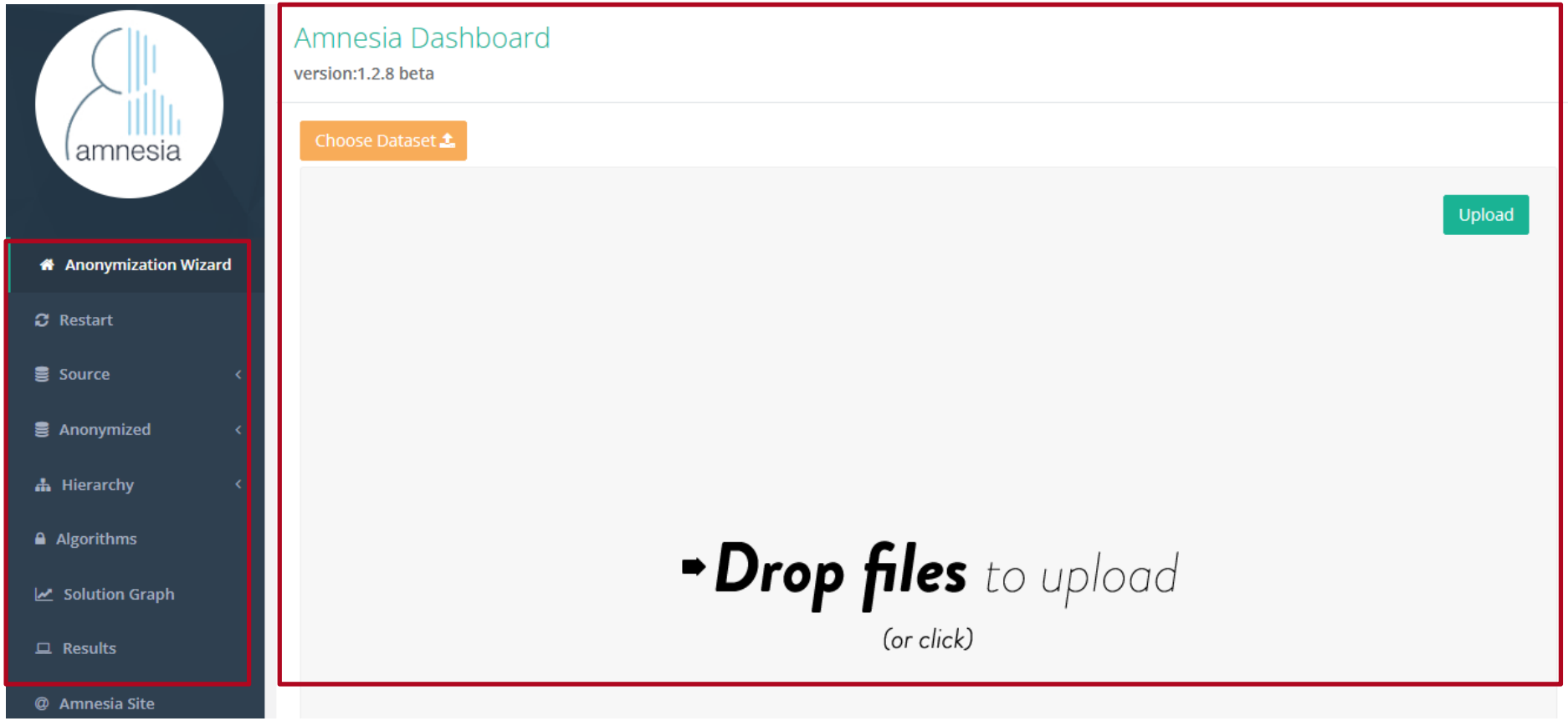
Amnesia

1. Panel de trabajo / guías



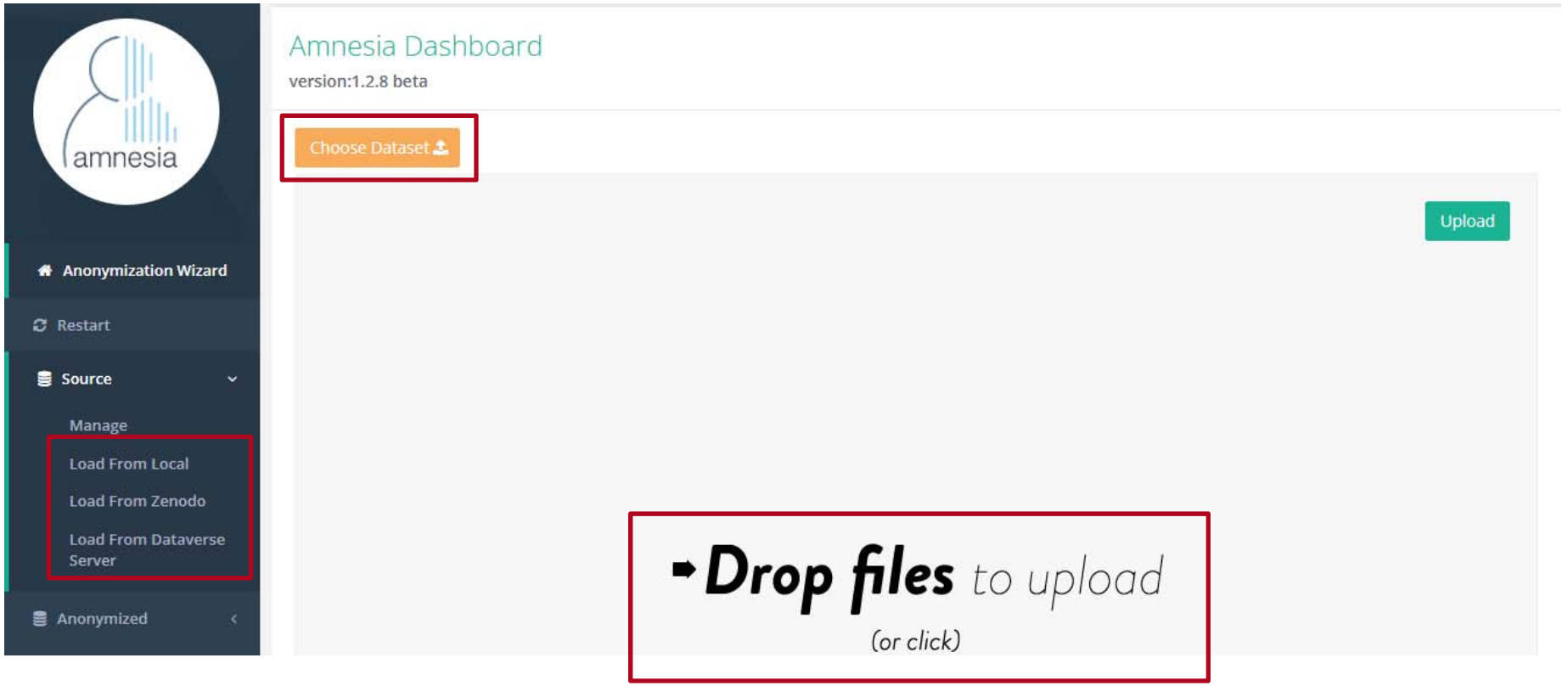
Amnesia

1. Panel de trabajo / guías



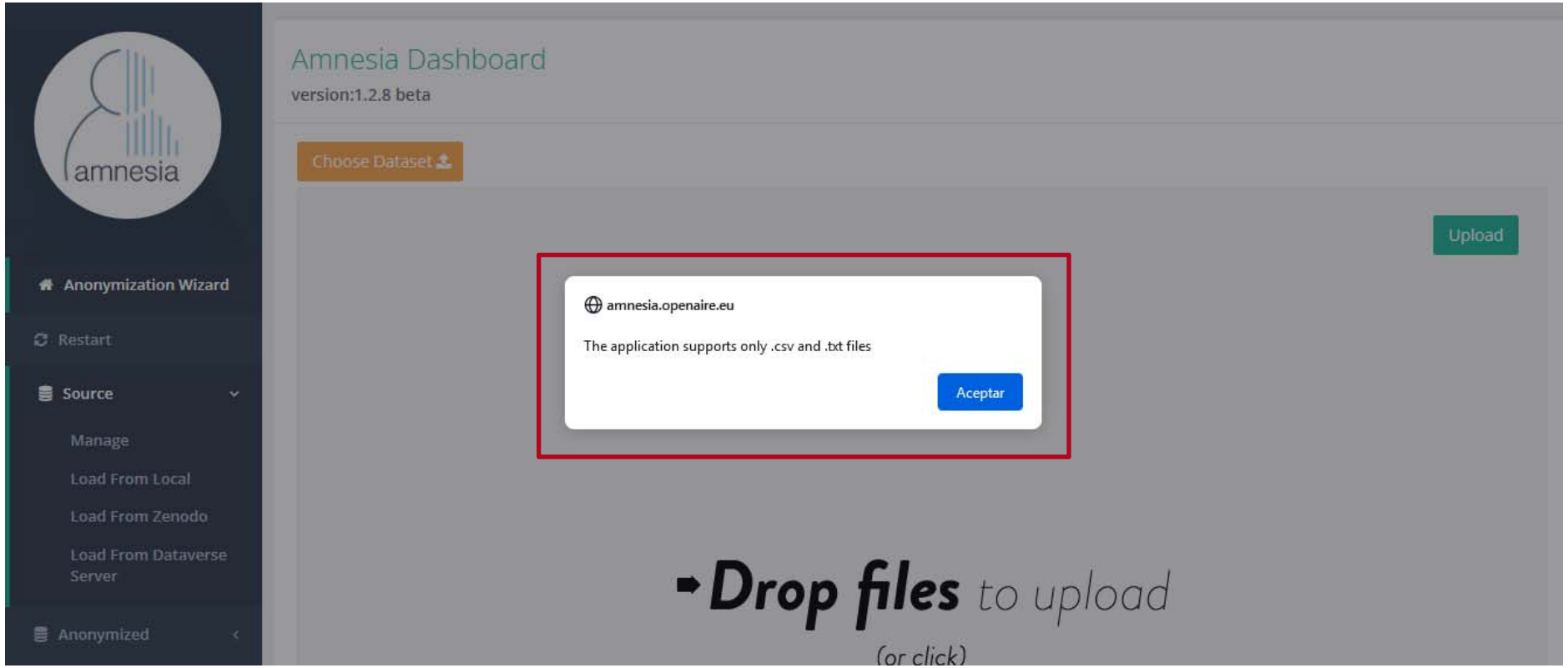
Amnesia

2. Subir fichero



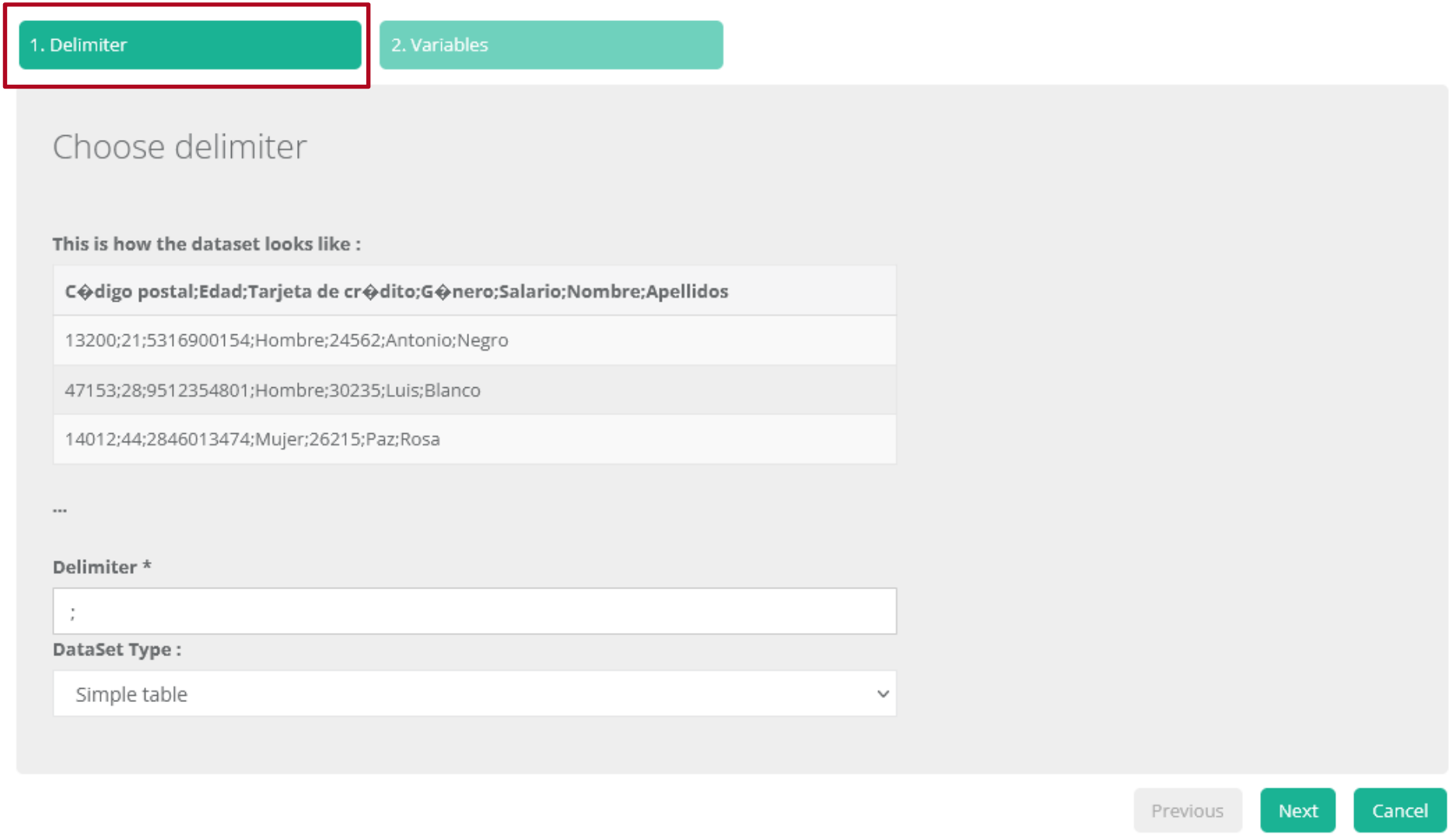
Amnesia

2. Subir fichero



Amnesia

2. Subir fichero



DIGITAL.CSIC

Amnesia

2. Subir fichero

1. Delimiter | 2. Variables

What type is your data?

Toggle All

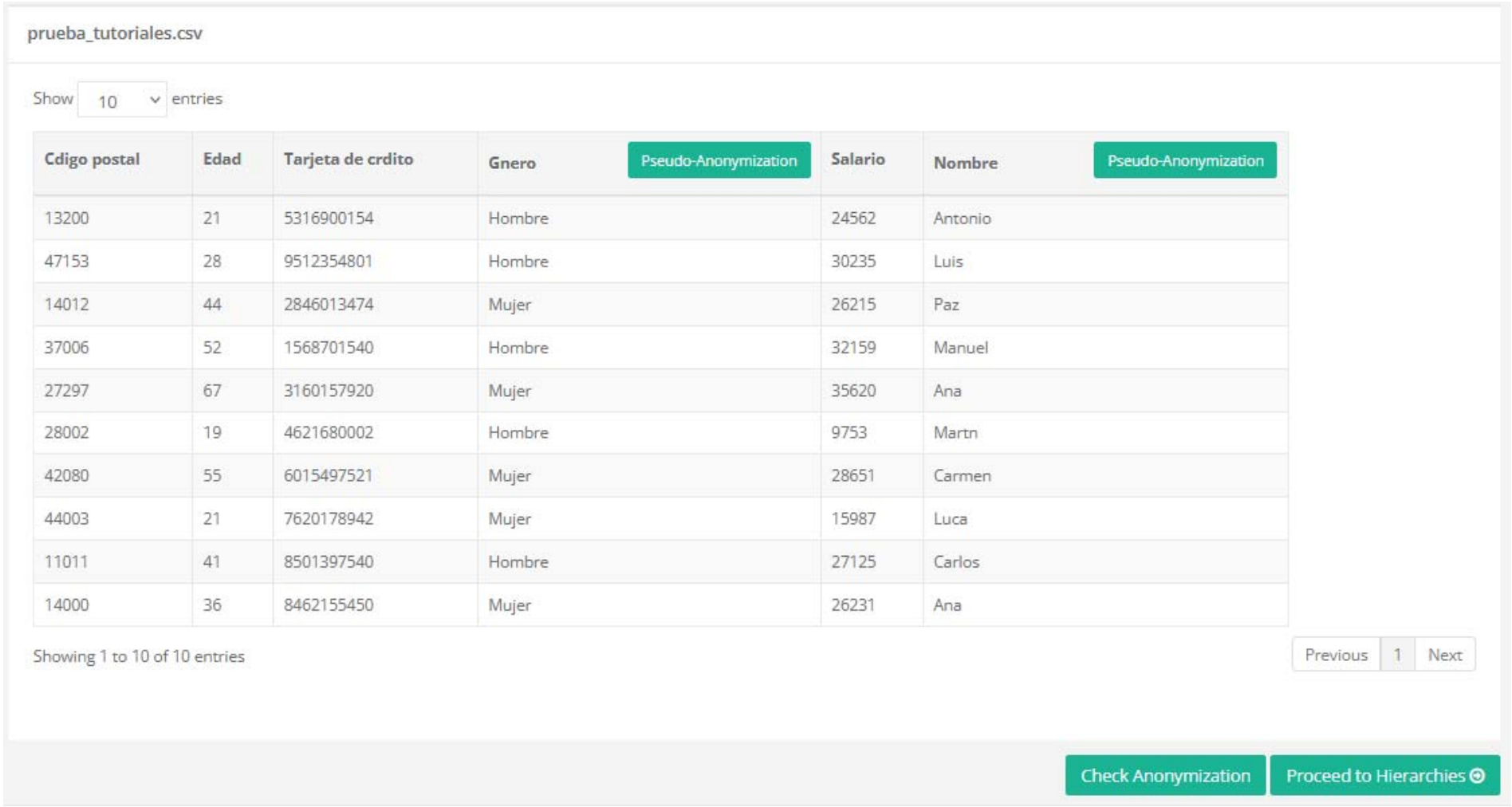
Choose the columns and their types.

Cdigo postal	Edad	Tarjeta de crdito	Gnero	Salario	Nombre	Apellidos
int	int	decimal	string	int	string	string
47153	28	9512354801	Hombre	30235	Luis	Blanco
14012	44	2846013474	Mujer	26215	Paz	Rosa
37006	52	1568701540	Hombre	32159	Manuel	Gris
27297	67	3160157920	Mujer	35620	Ana	Morado
28002	19	4621680002	Hombre	9753	Martn	Azul

Previous | Finish | Cancel

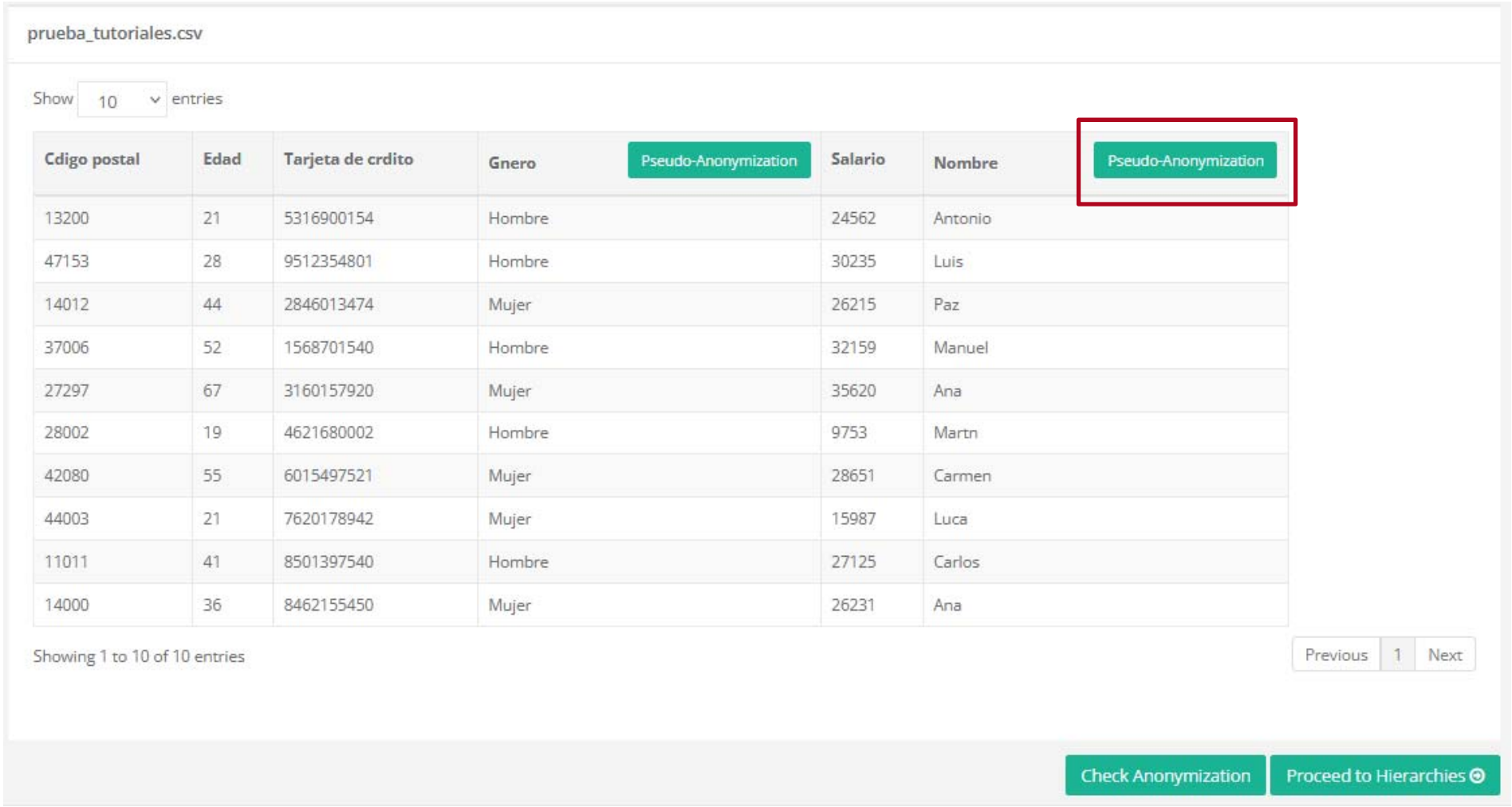
Amnesia

2. Subir fichero



Amnesia

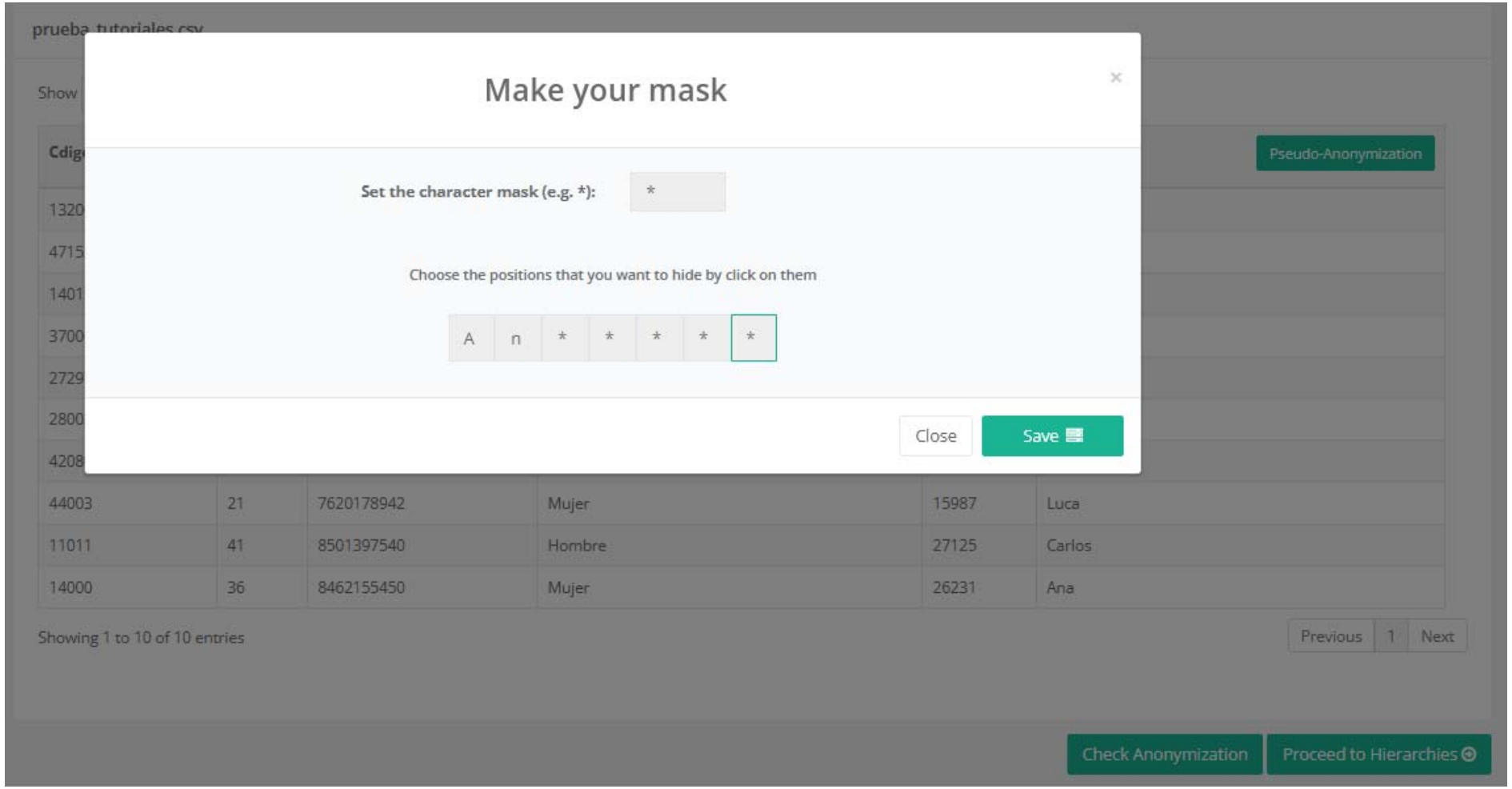
2. Subir fichero



DIGITAL.CSIC

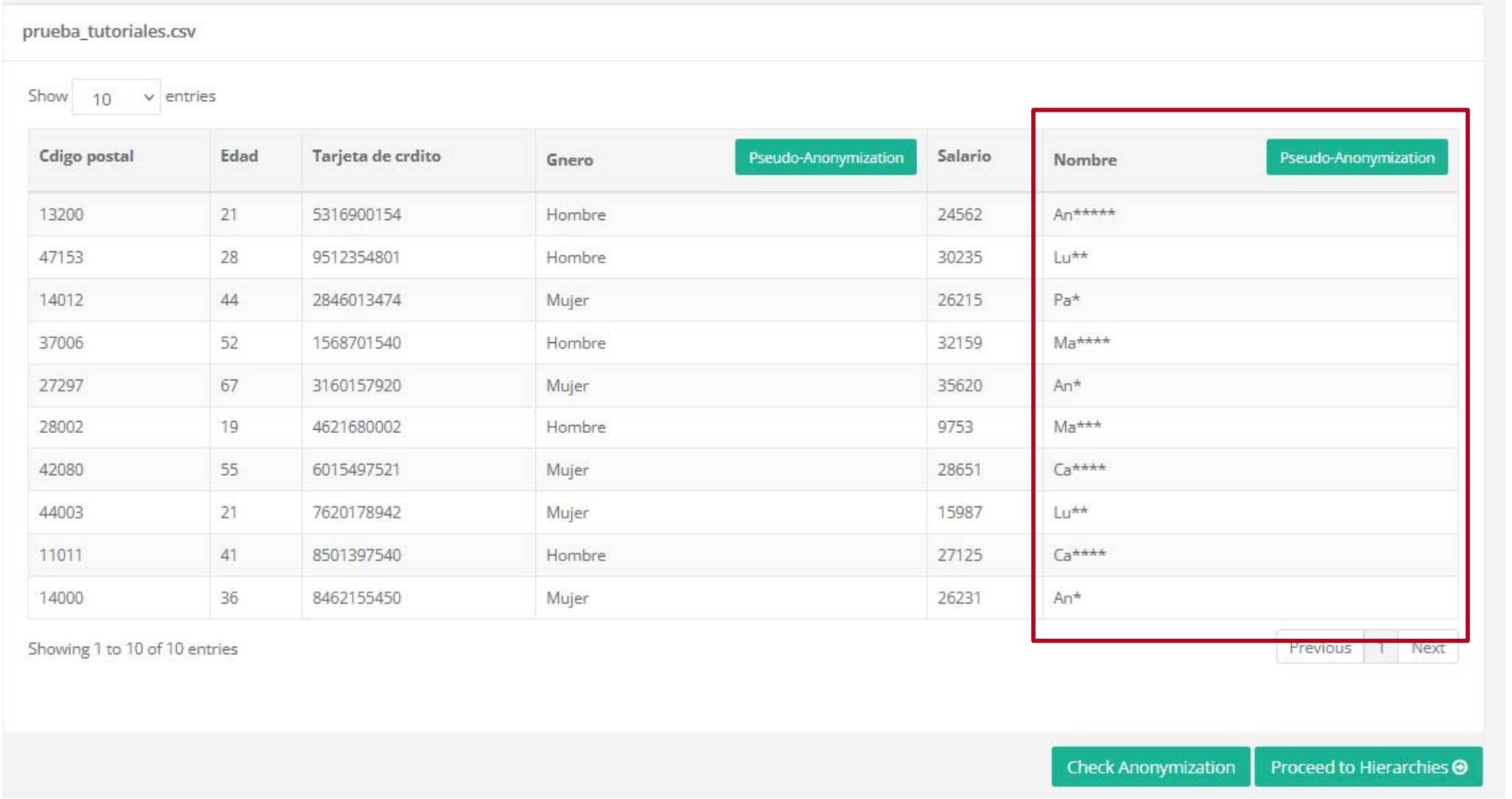
Amnesia

2. Subir fichero



Amnesia

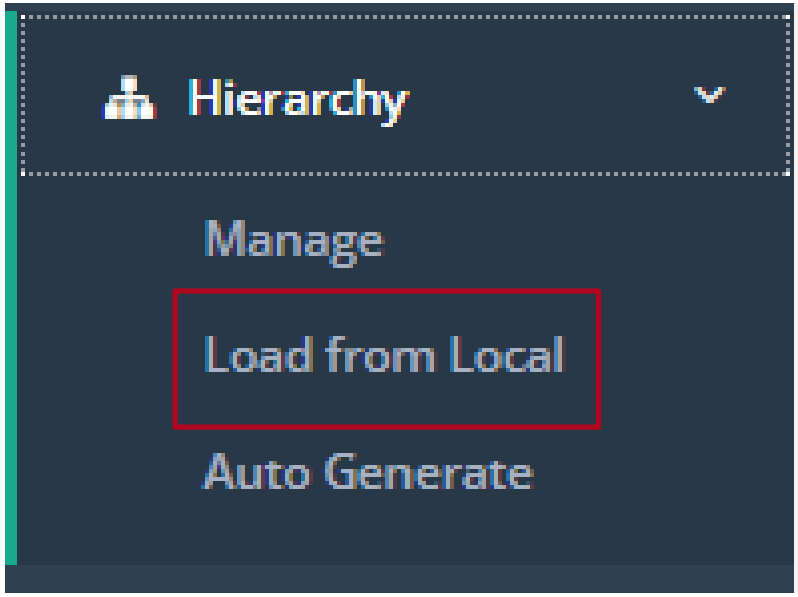
2. Subir fichero



DIGITAL.CSIC

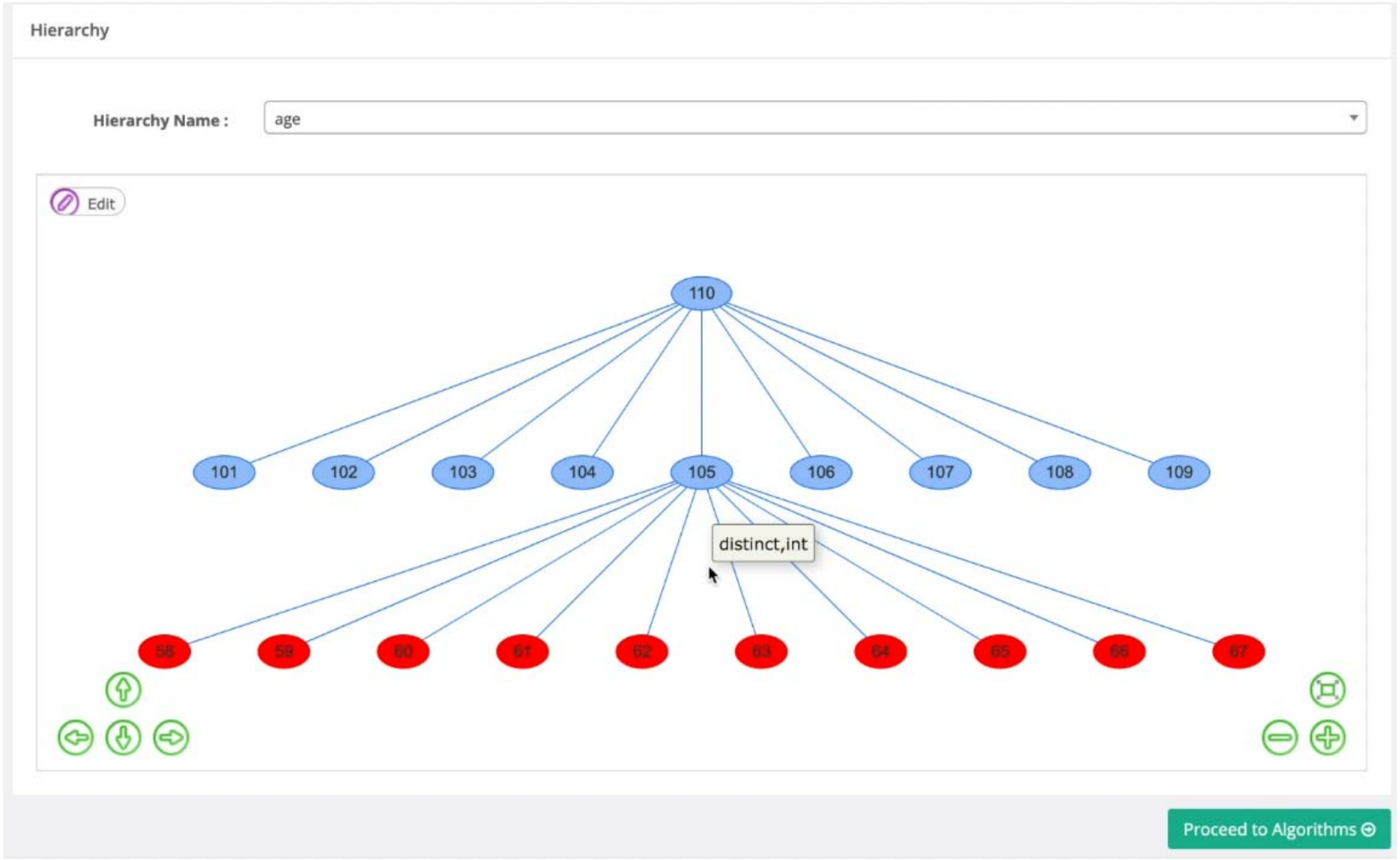
Amnesia

3. Generar jerarquías



Amnesia

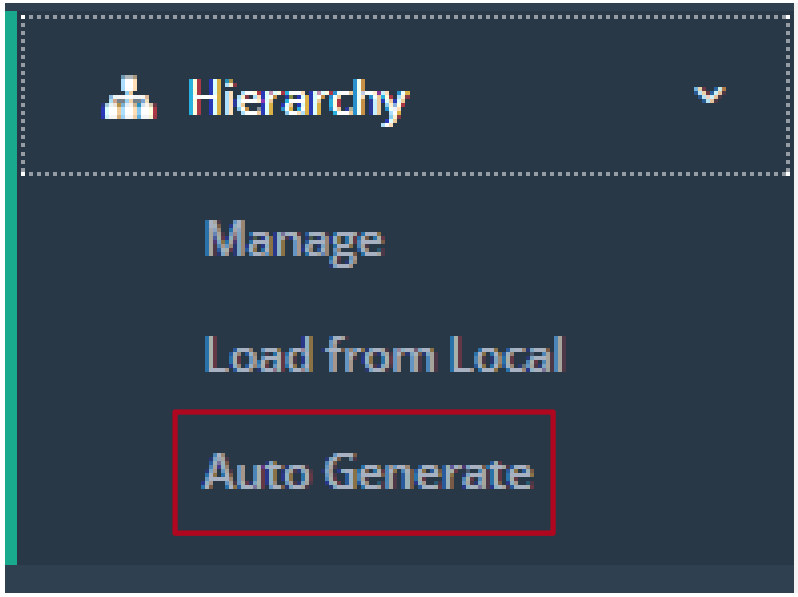
3. Generar jerarquías



DIGITAL.CSIC

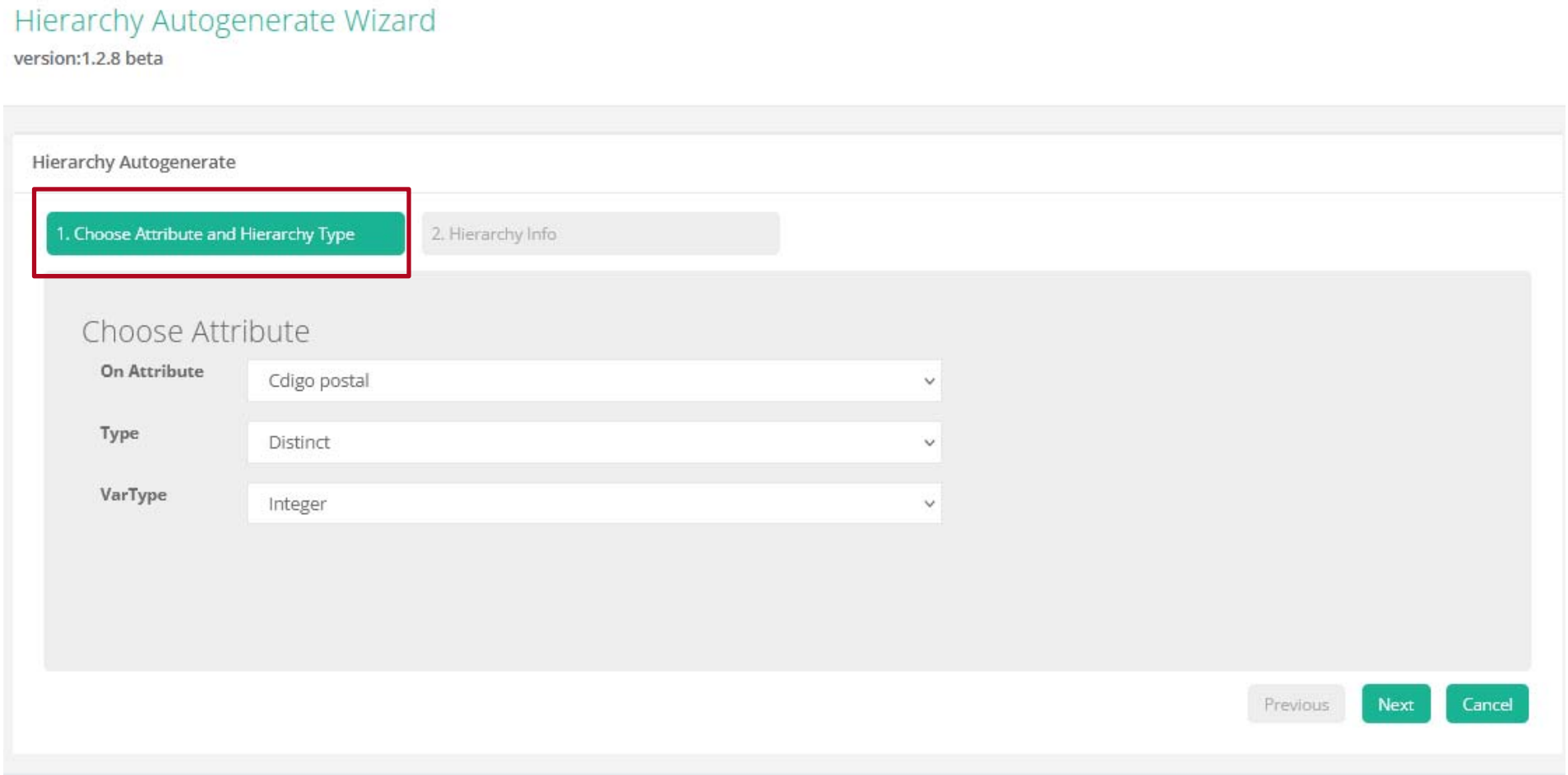
Amnesia

3. Generar jerarquías



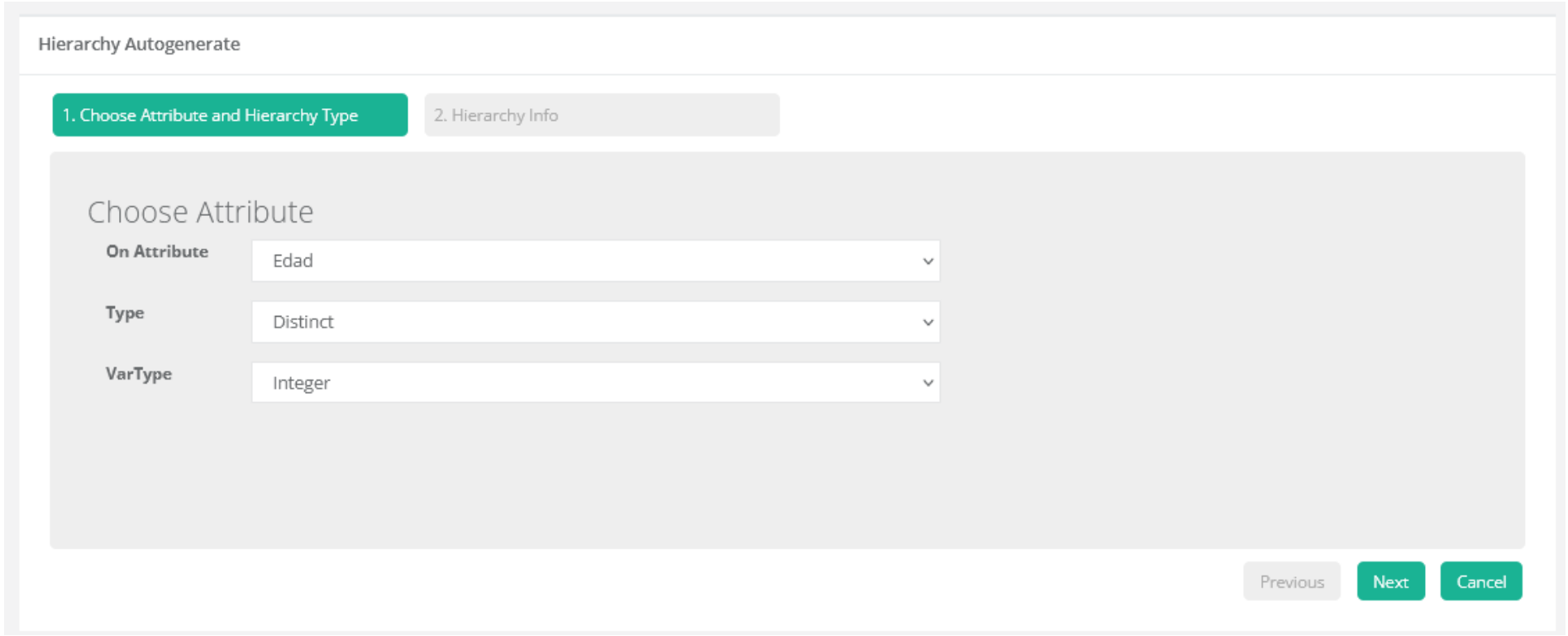
Amnesia

3. Generar jerarquías



Amnesia

3. Generar jerarquías

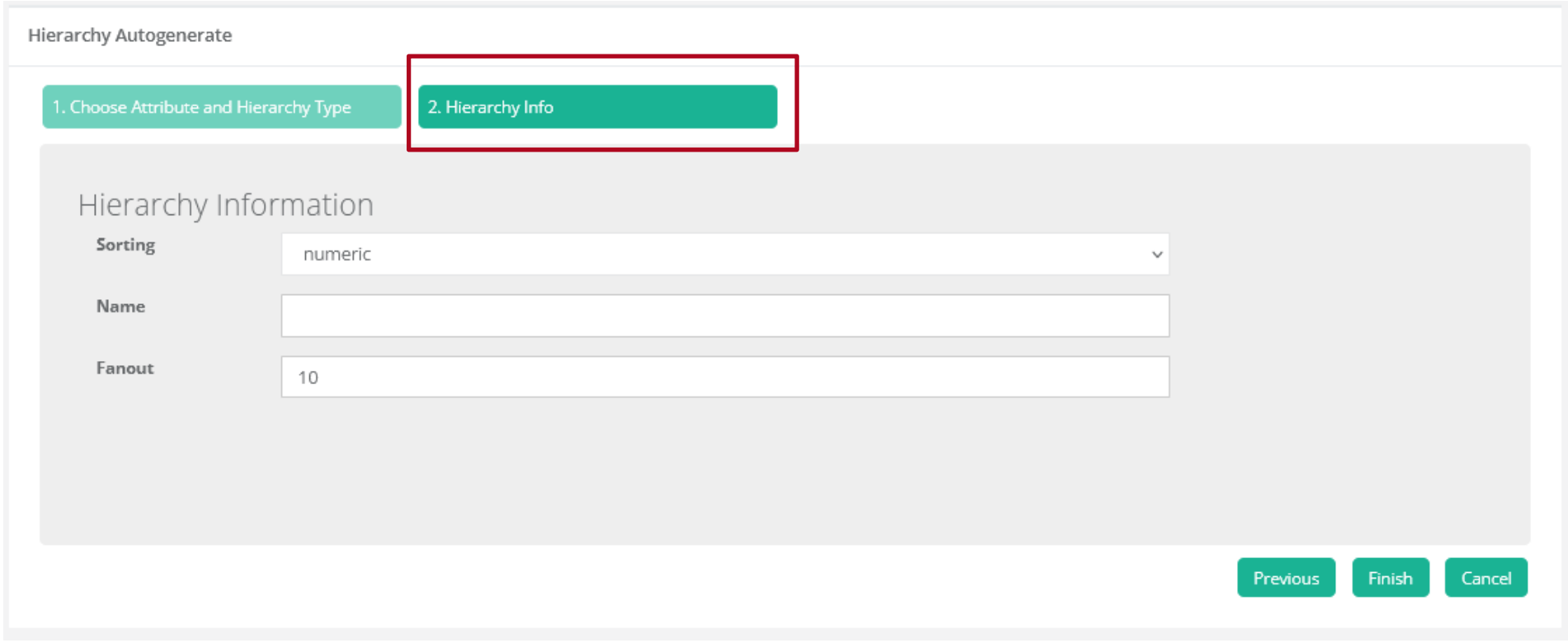


Distinct

Range

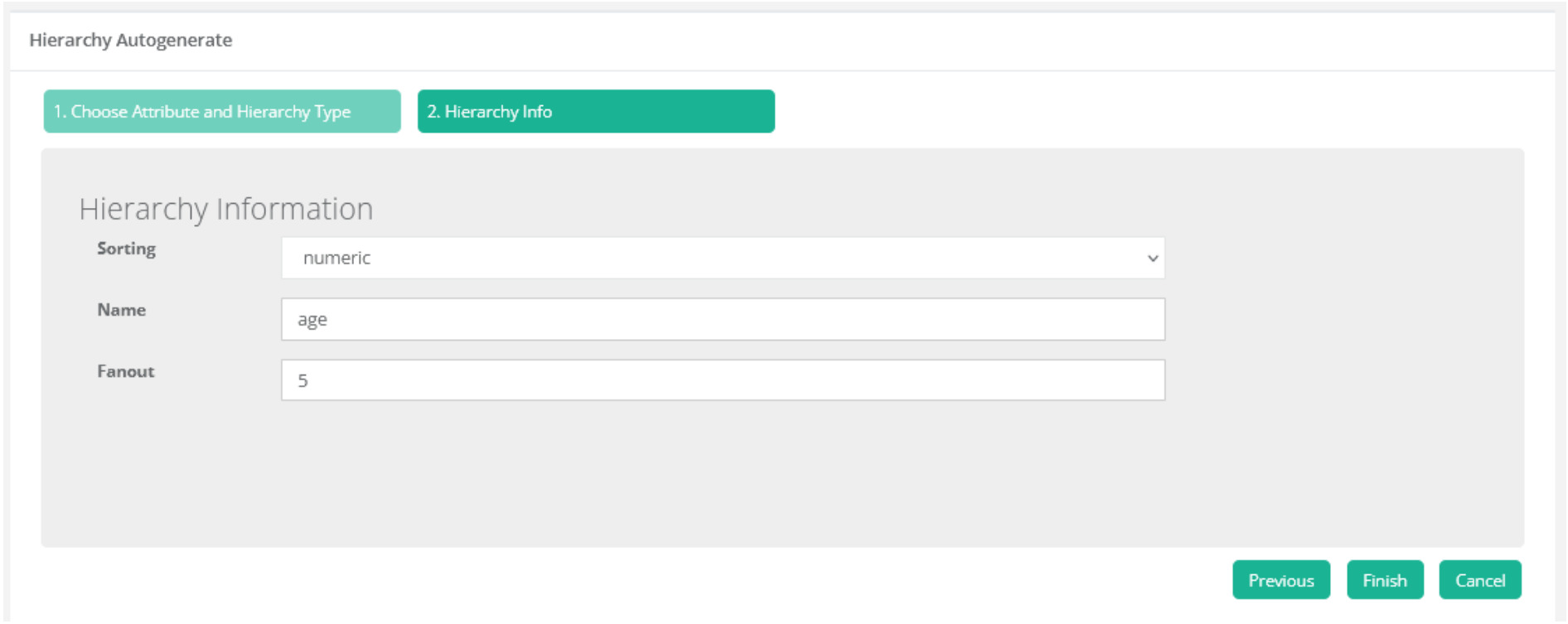
Amnesia

3. Generar jerarquías



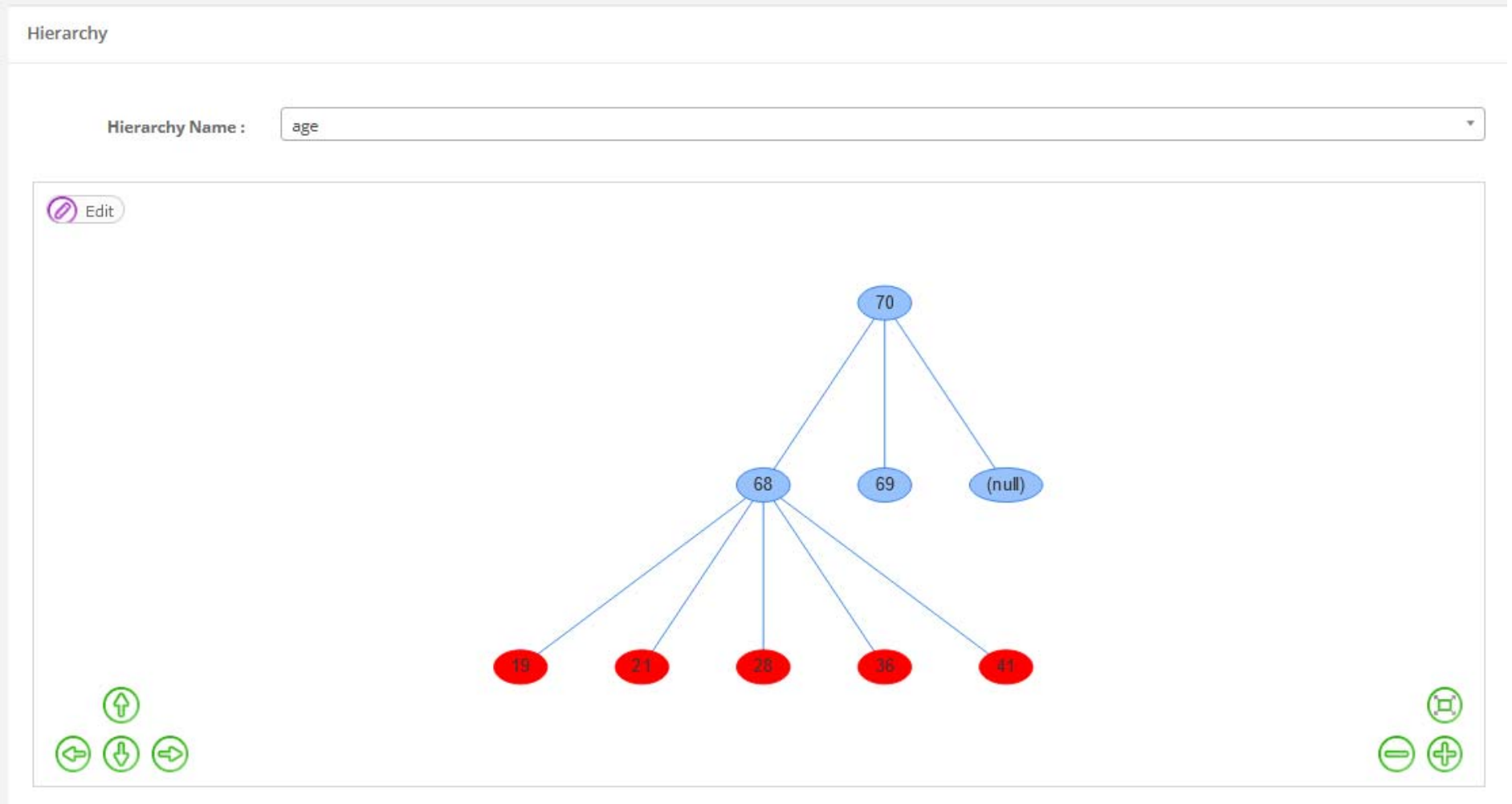
Amnesia

3. Generar jerarquías



Amnesia

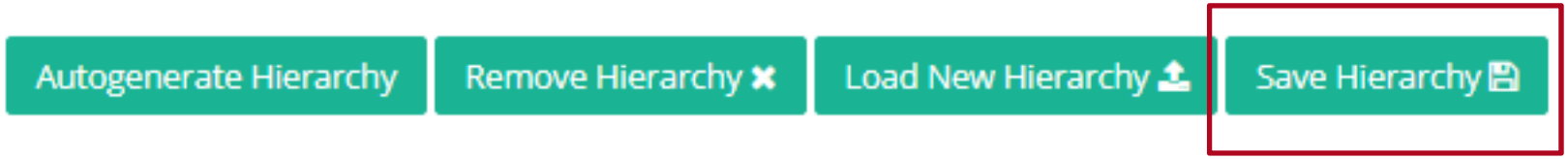
3. Generar jerarquías

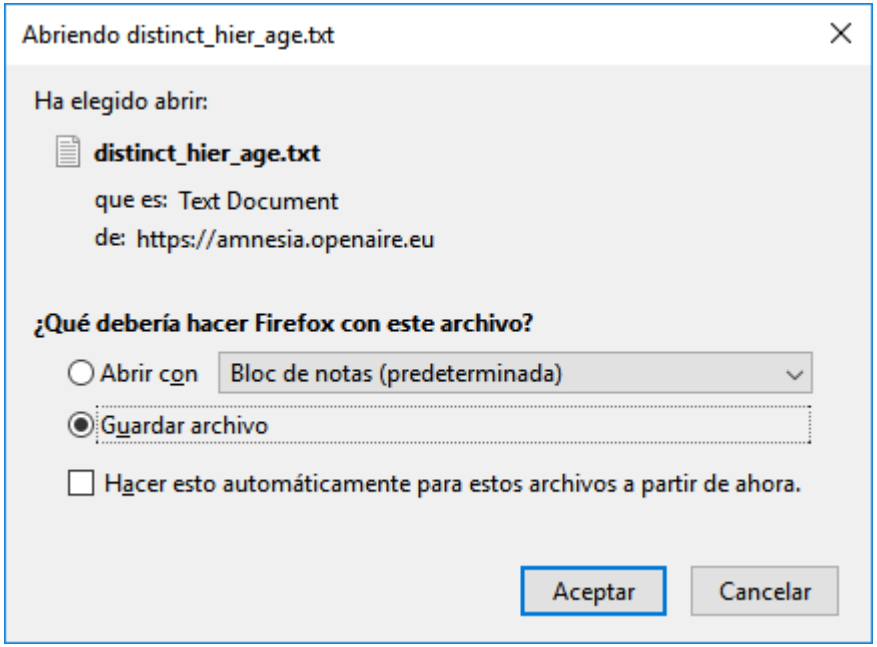


Los valores se han clasificado numéricamente y colocados en grupos de 5

Amnesia

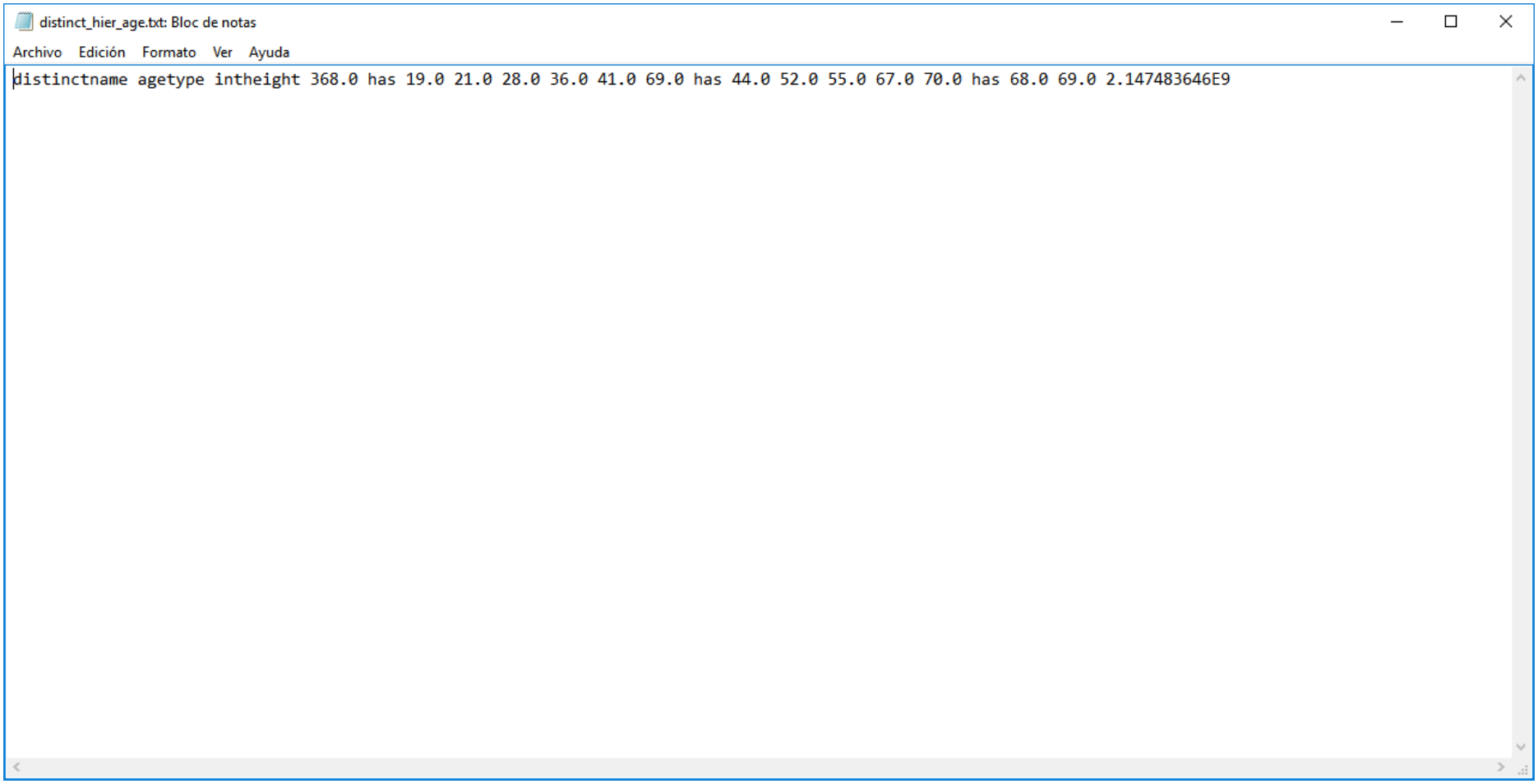
3. Generar jerarquías





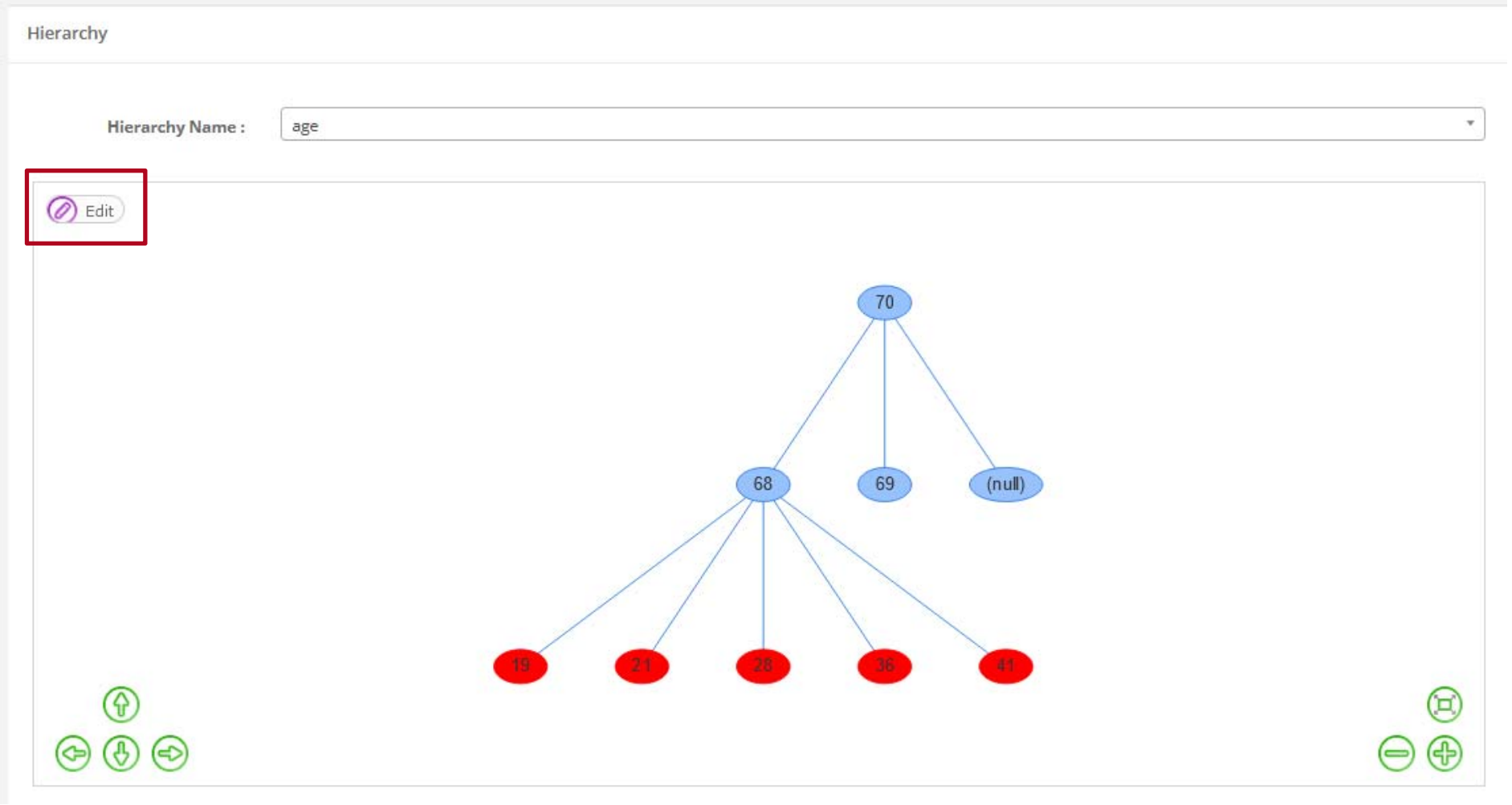
Amnesia

3. Generar jerarquías



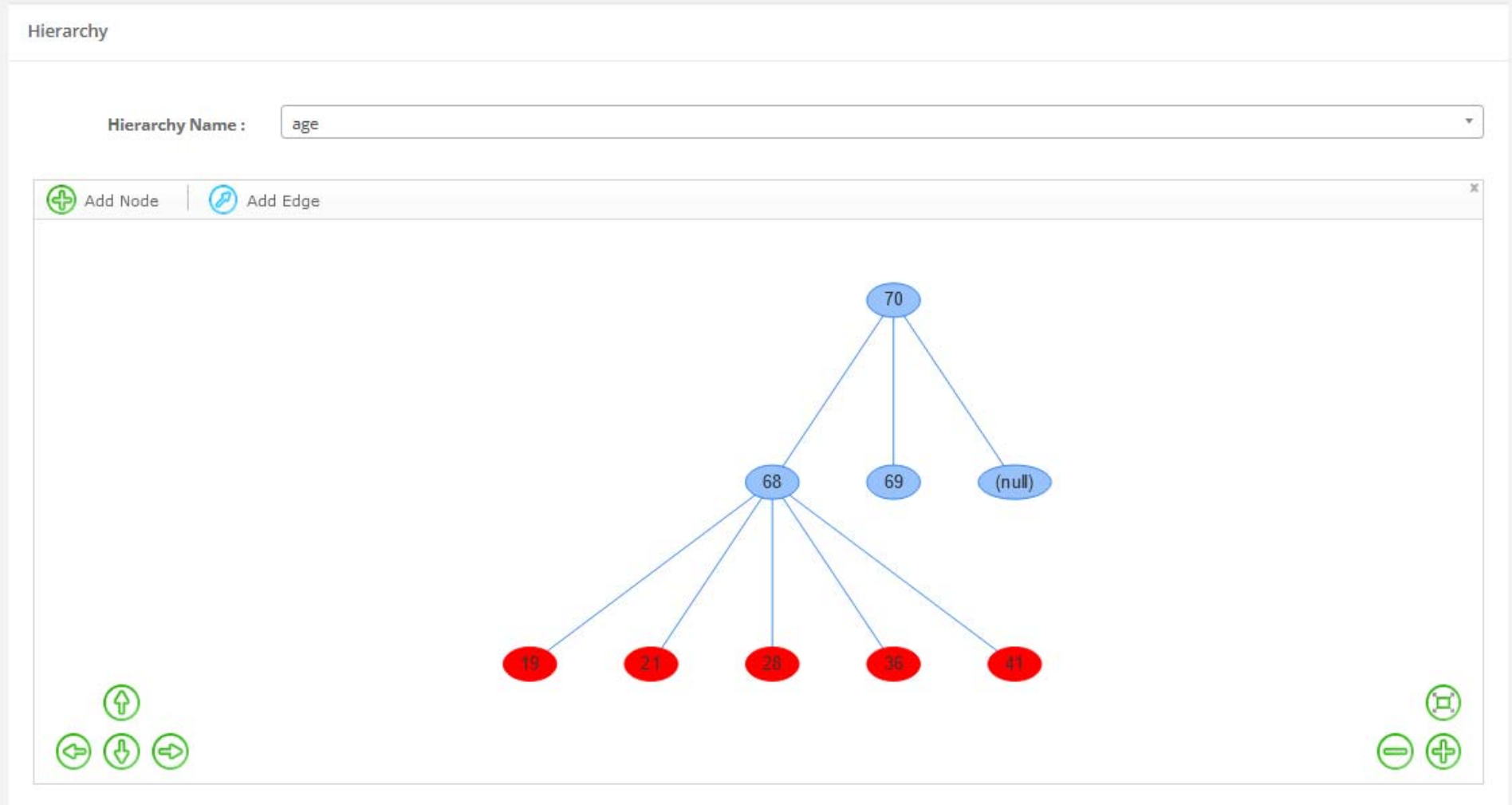
Amnesia

3. Generar jerarquías



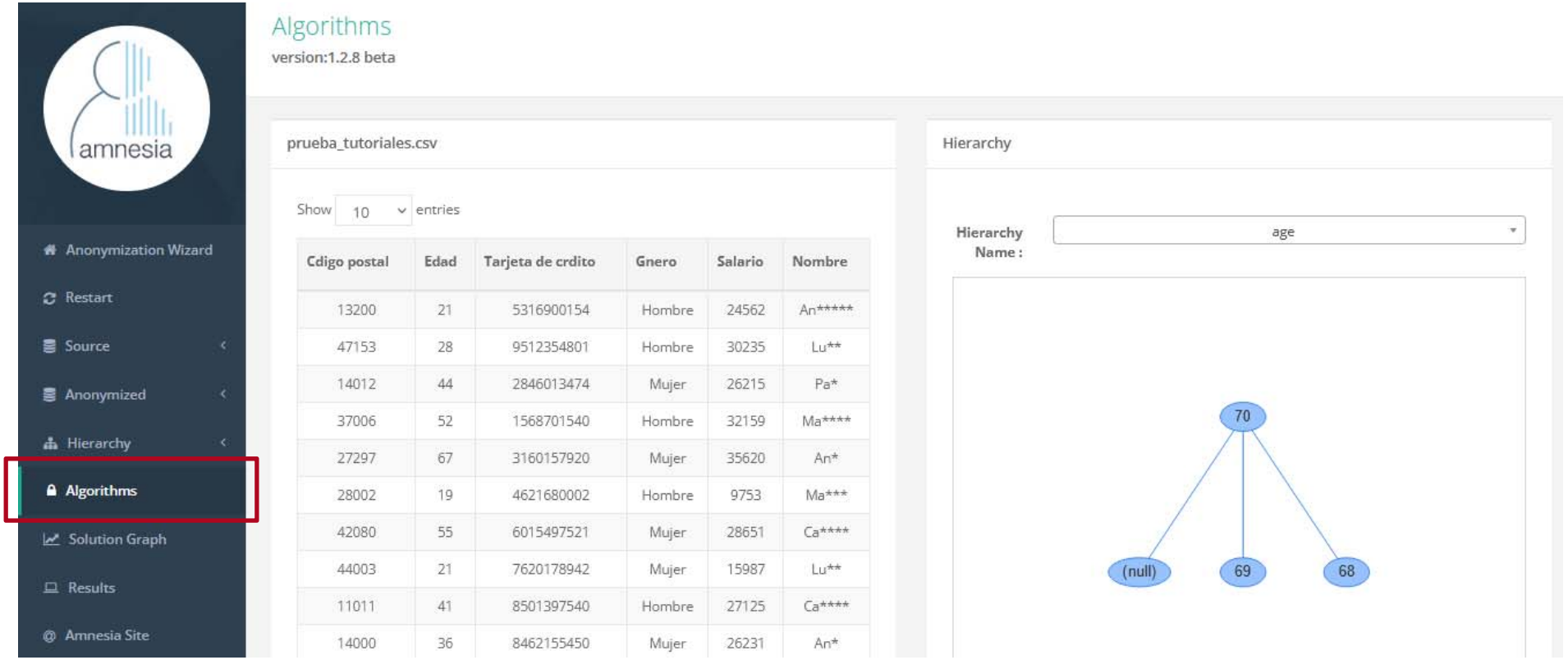
Amnesia

3. Generar jerarquías



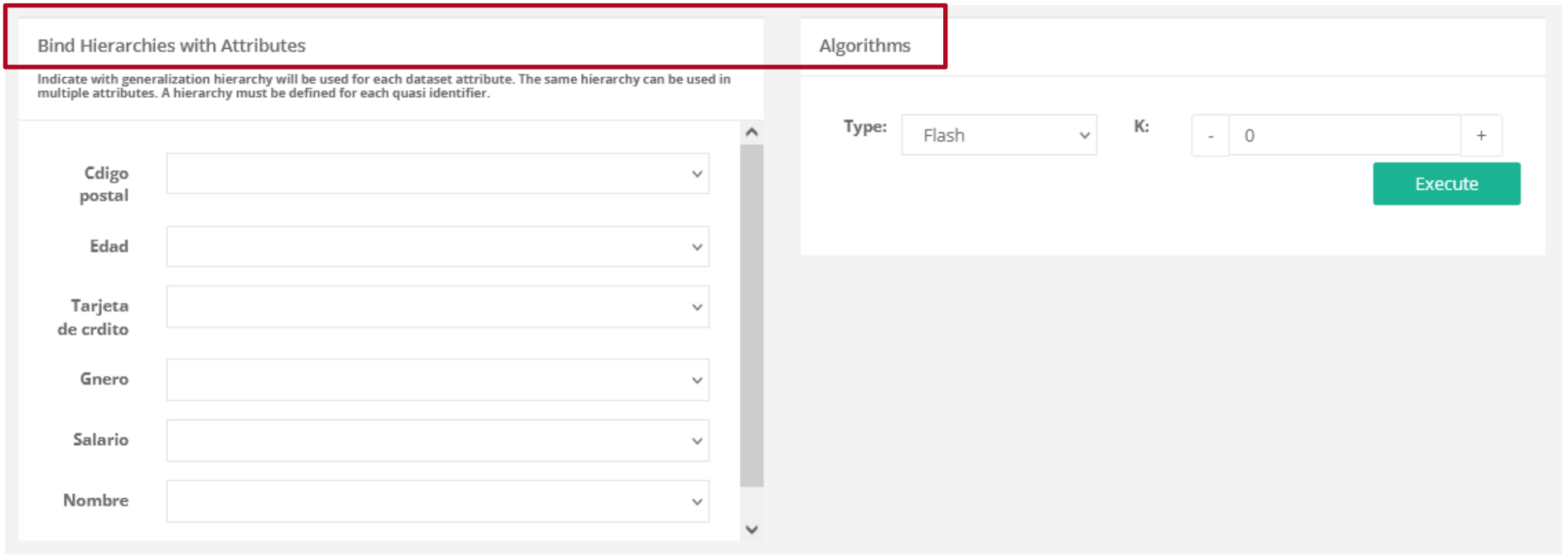
Amnesia

4. Anonimizar los datos



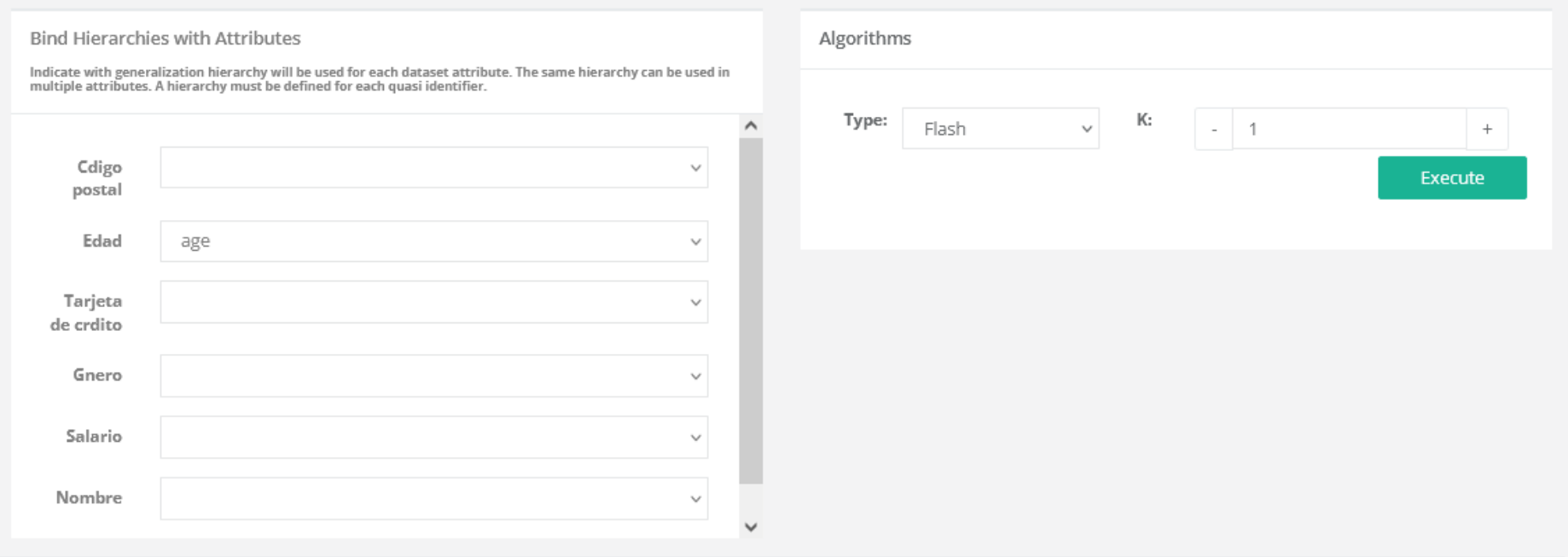
Amnesia

4. Anonimizar los datos



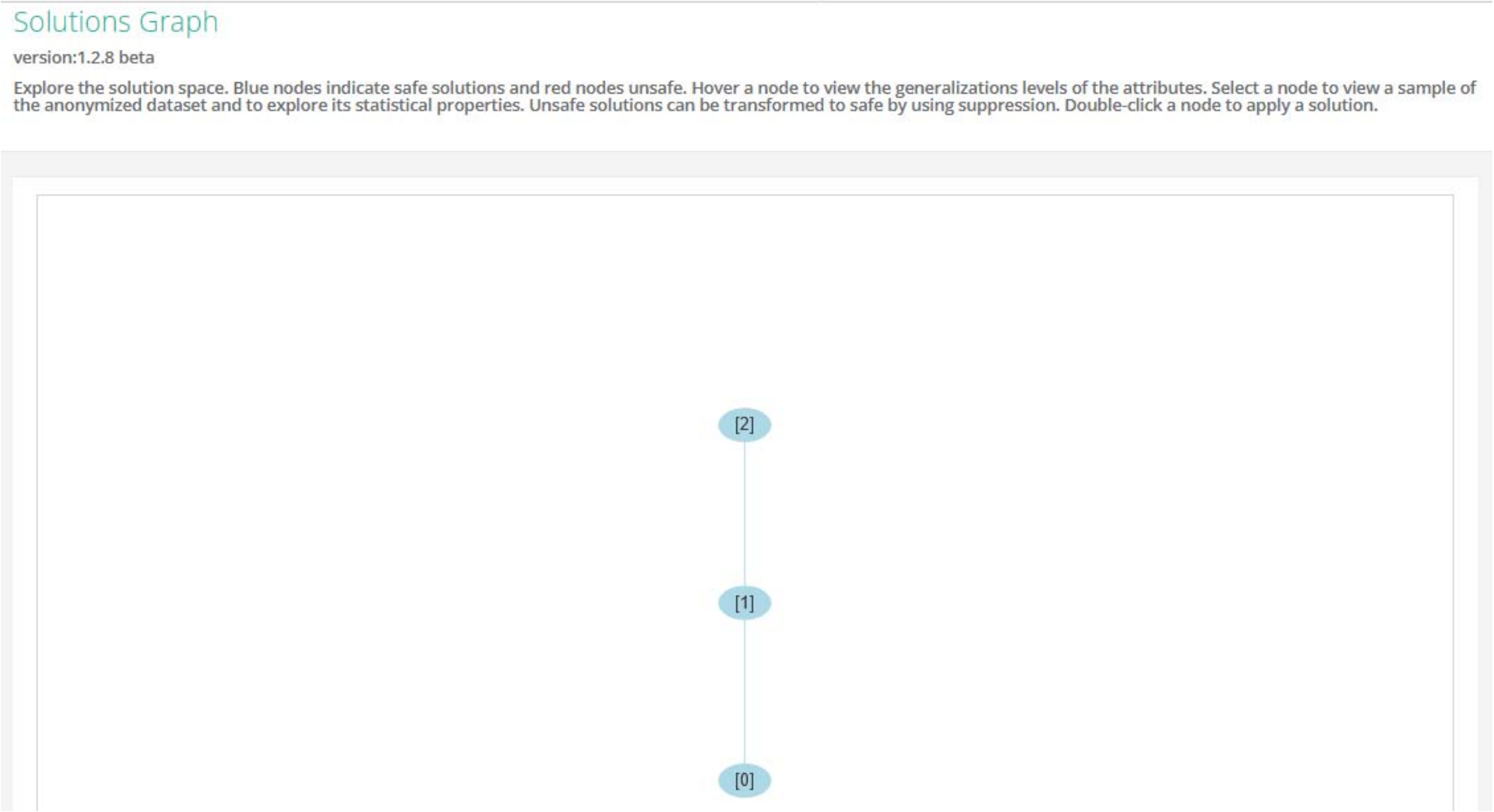
Amnesia

4. Anonimizar los datos



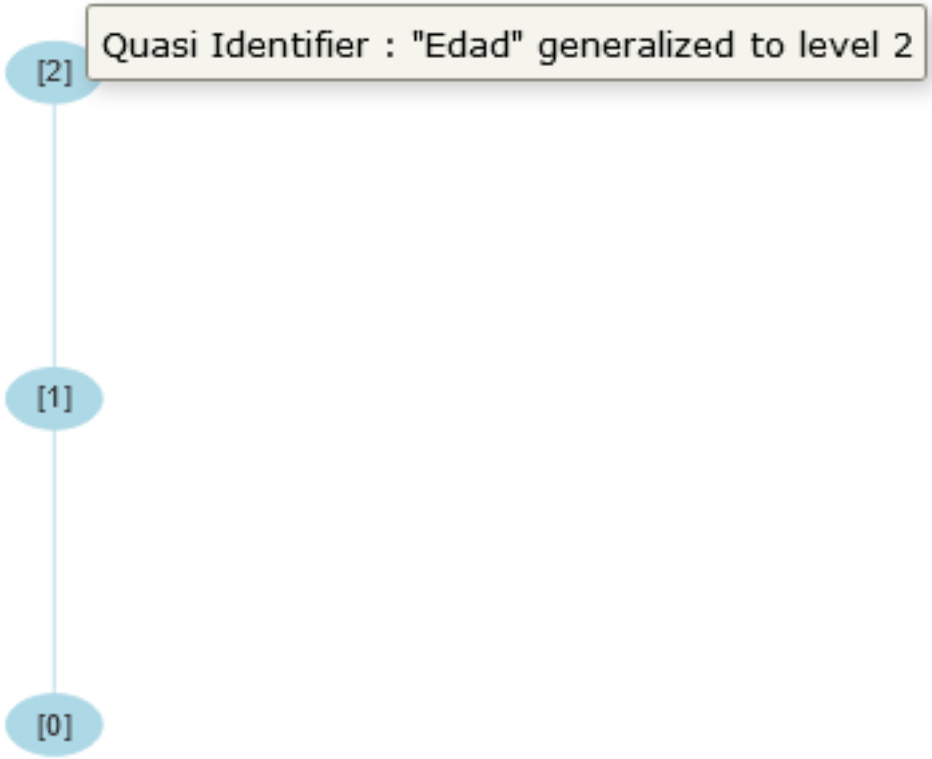
Amnesia

4. Anonimizar los datos



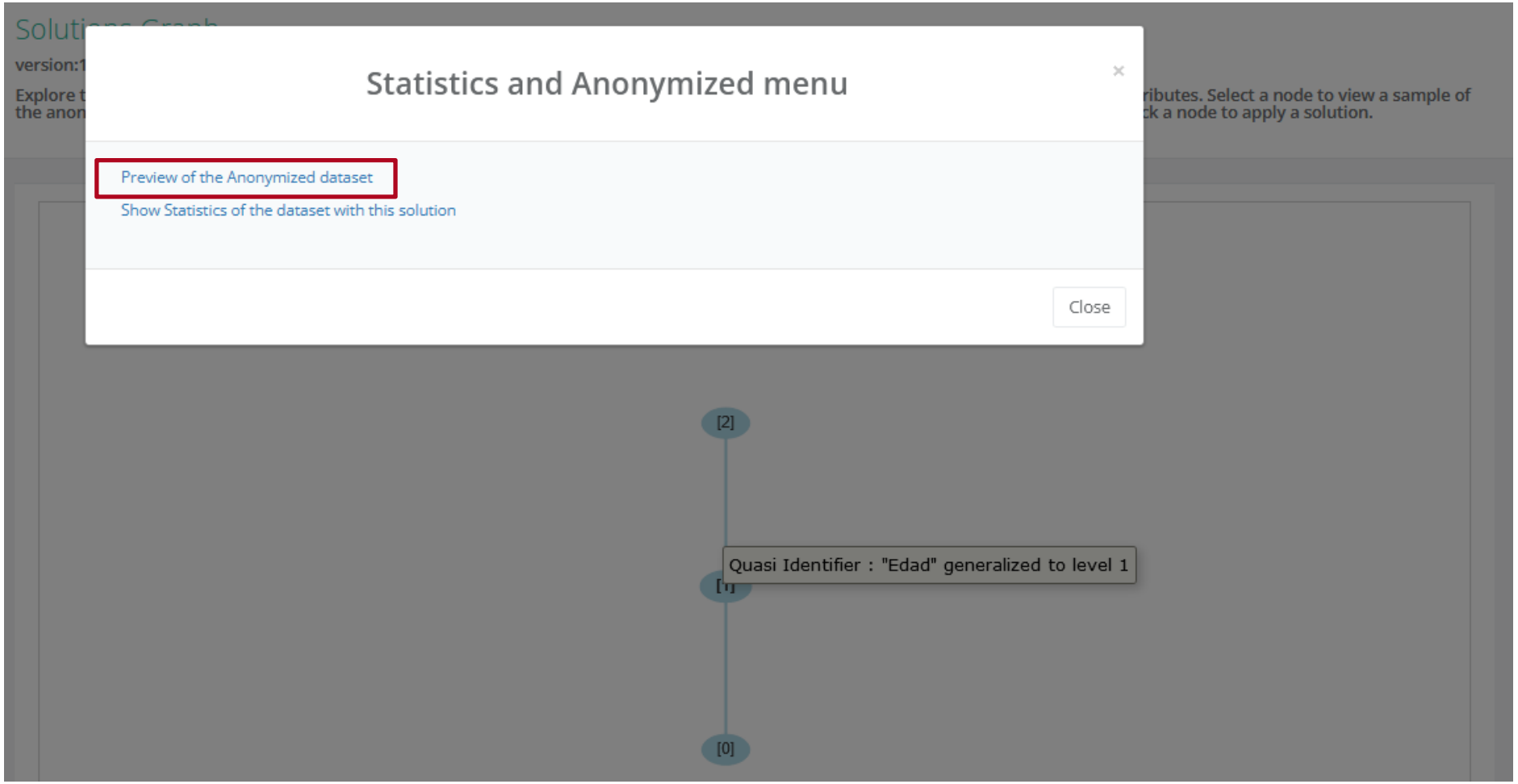
Amnesia

4. Anonimizar los datos



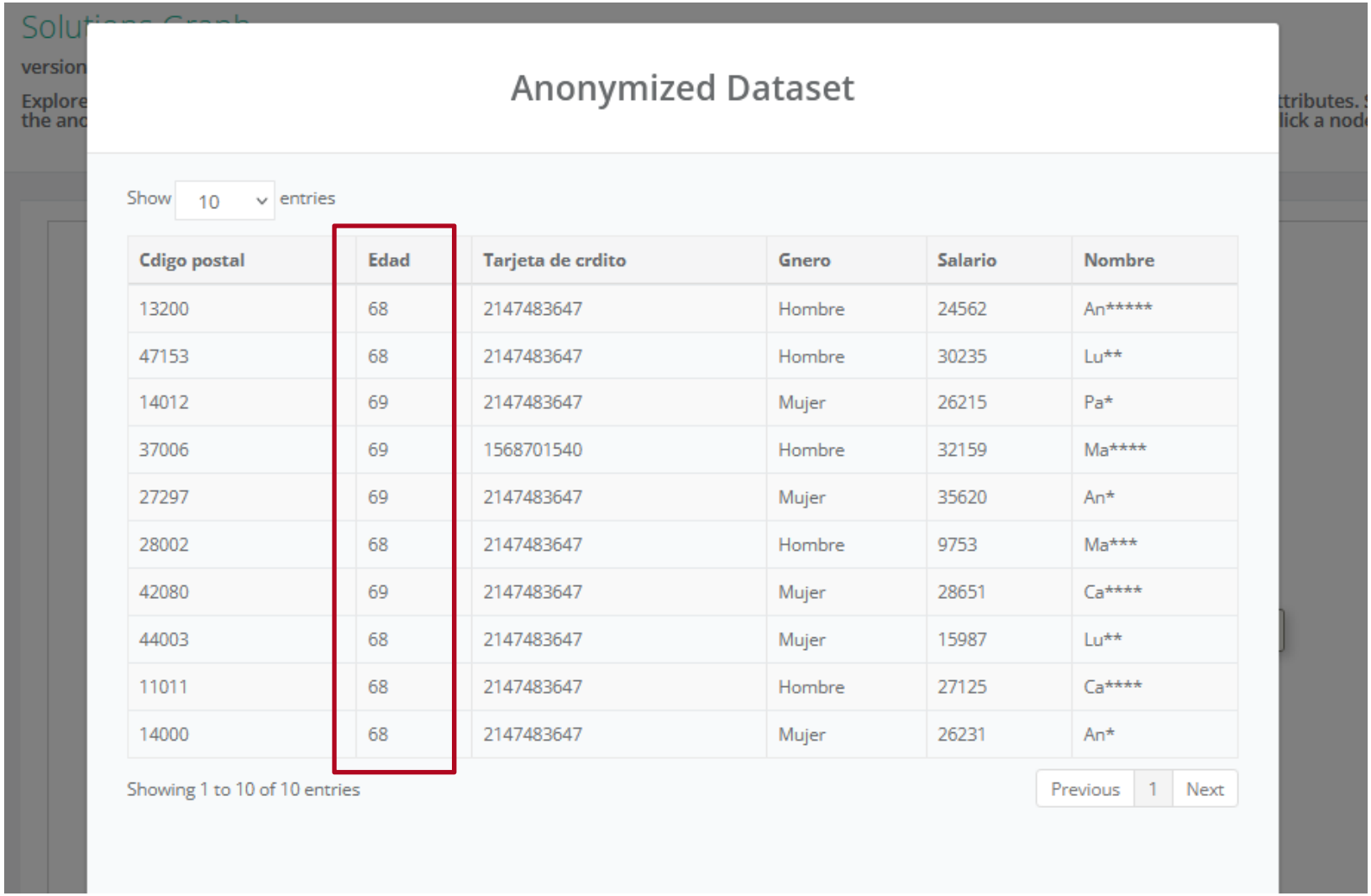
Amnesia

4. Anonimizar los datos



Amnesia

4. Anonimizar los datos



Anonymized Dataset

Show 10 entries

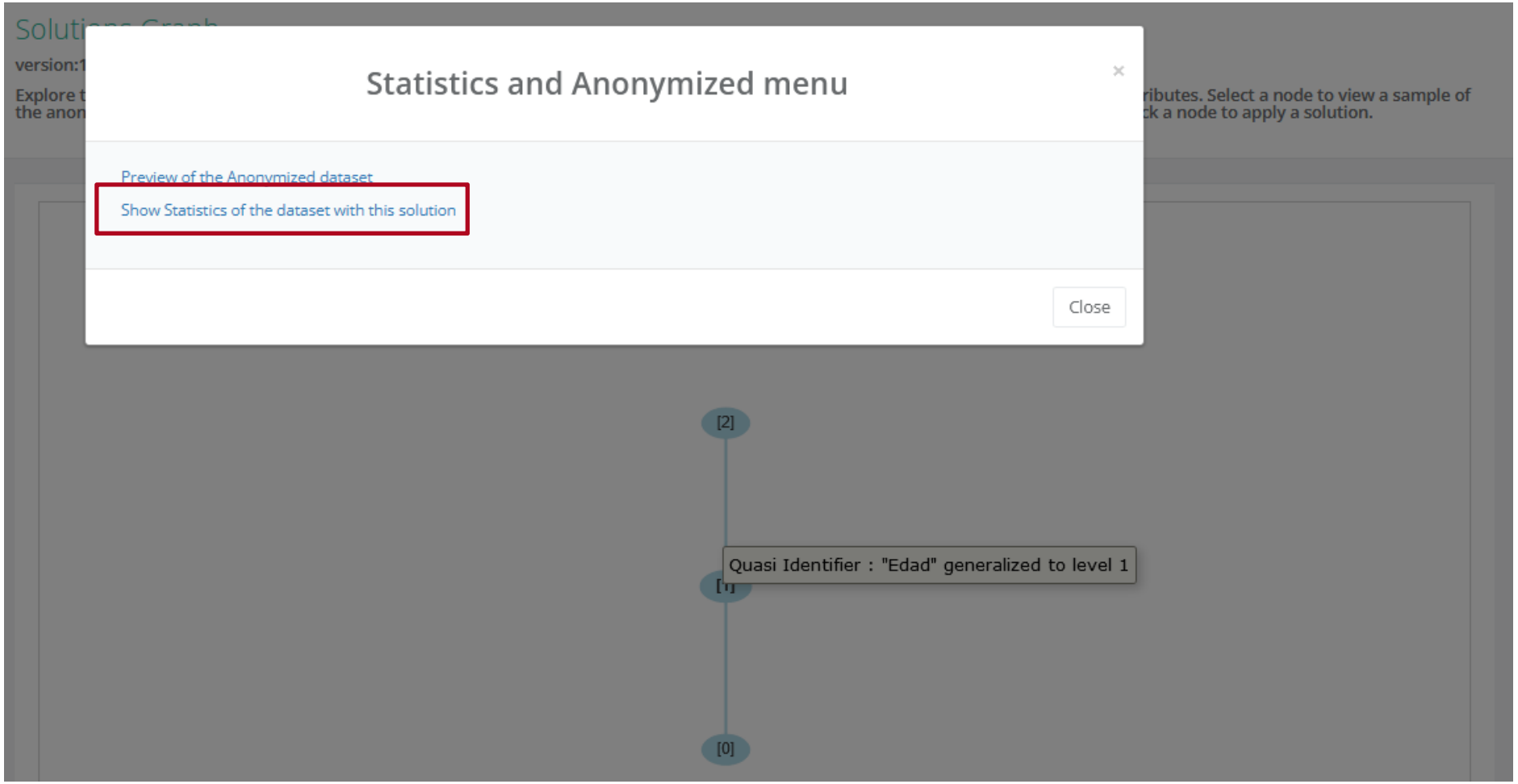
Cdigo postal	Edad	Tarjeta de crdito	Gnero	Salario	Nombre
13200	68	2147483647	Hombre	24562	An*****
47153	68	2147483647	Hombre	30235	Lu**
14012	69	2147483647	Mujer	26215	Pa*
37006	69	1568701540	Hombre	32159	Ma****
27297	69	2147483647	Mujer	35620	An*
28002	68	2147483647	Hombre	9753	Ma***
42080	69	2147483647	Mujer	28651	Ca****
44003	68	2147483647	Mujer	15987	Lu**
11011	68	2147483647	Hombre	27125	Ca****
14000	68	2147483647	Mujer	26231	An*

Showing 1 to 10 of 10 entries

Previous 1 Next

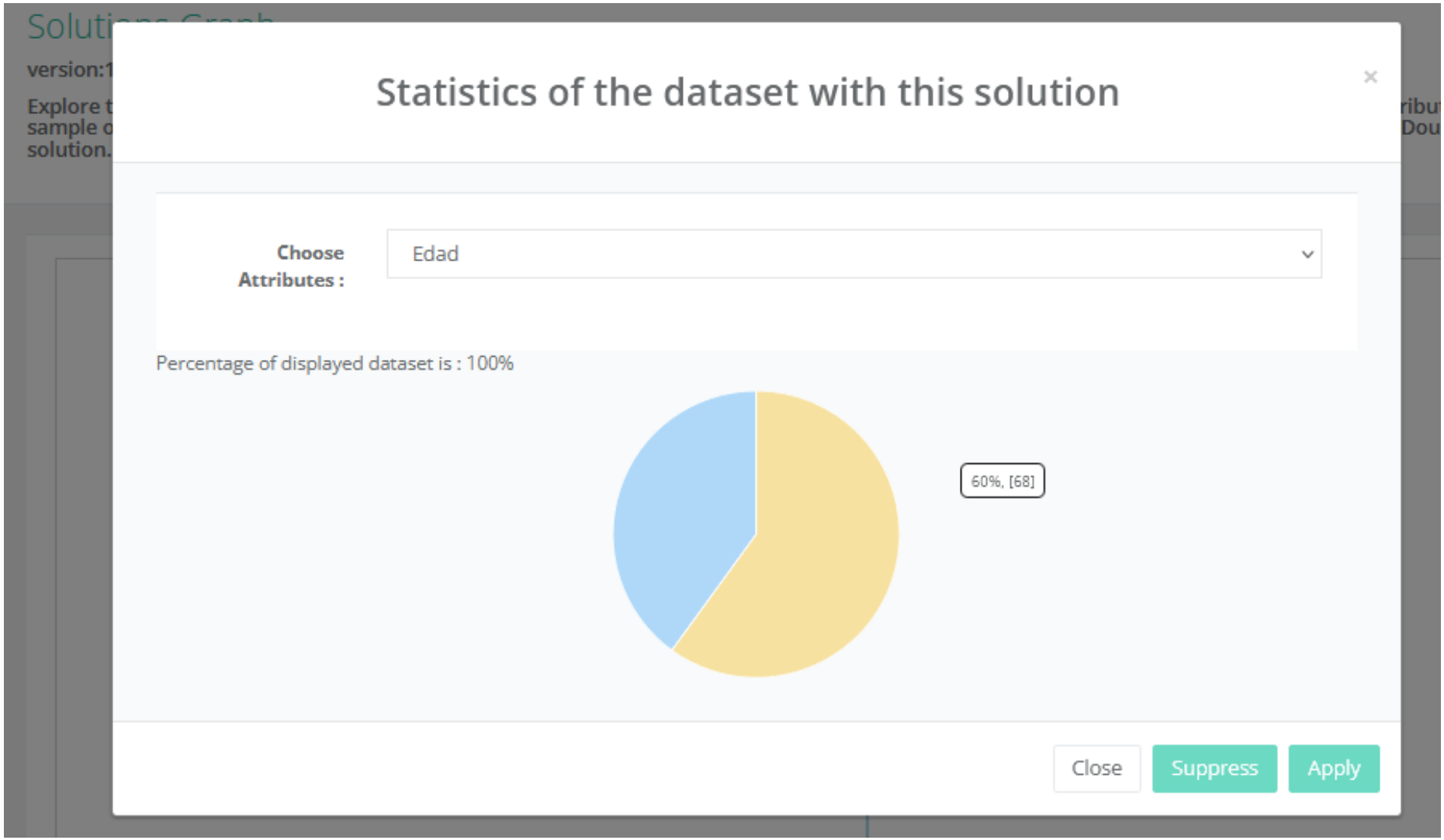
Amnesia

4. Anonimizar los datos



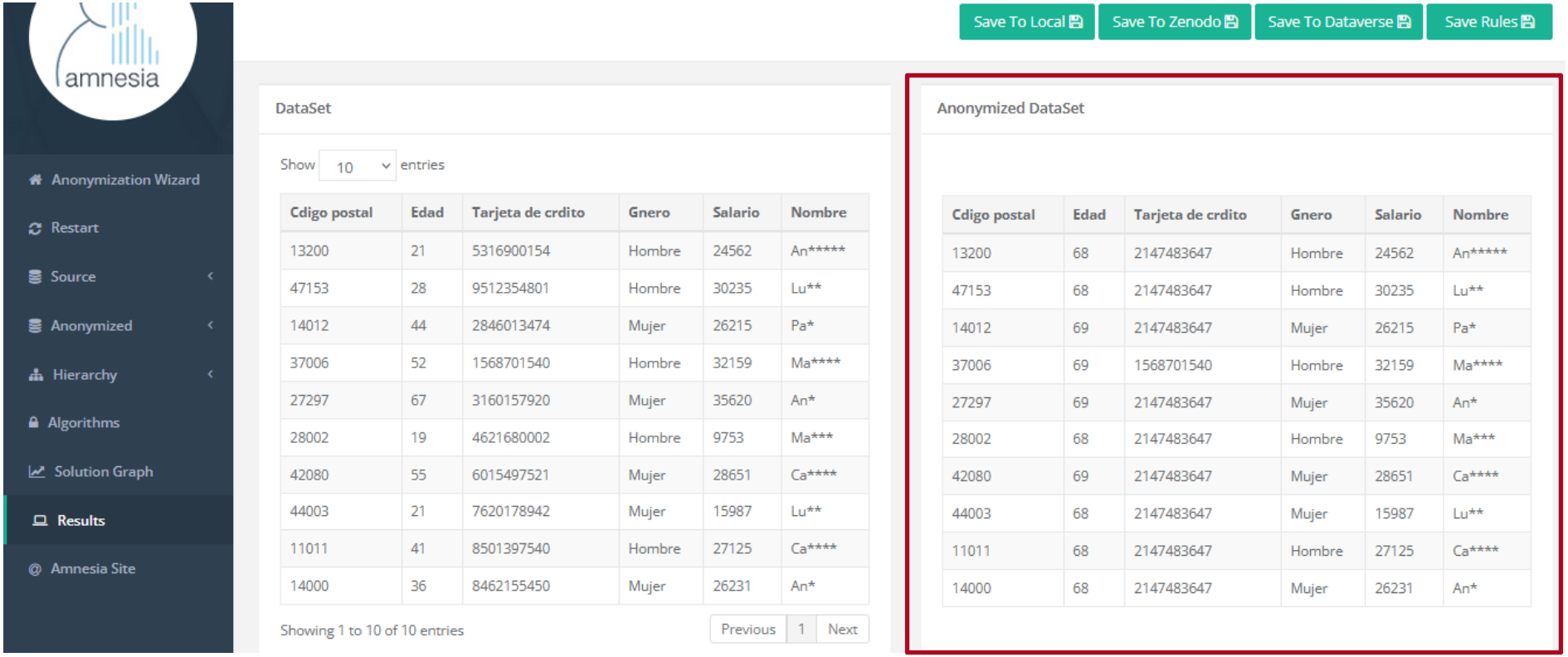
Amnesia

4. Anonimizar los datos



Amnesia

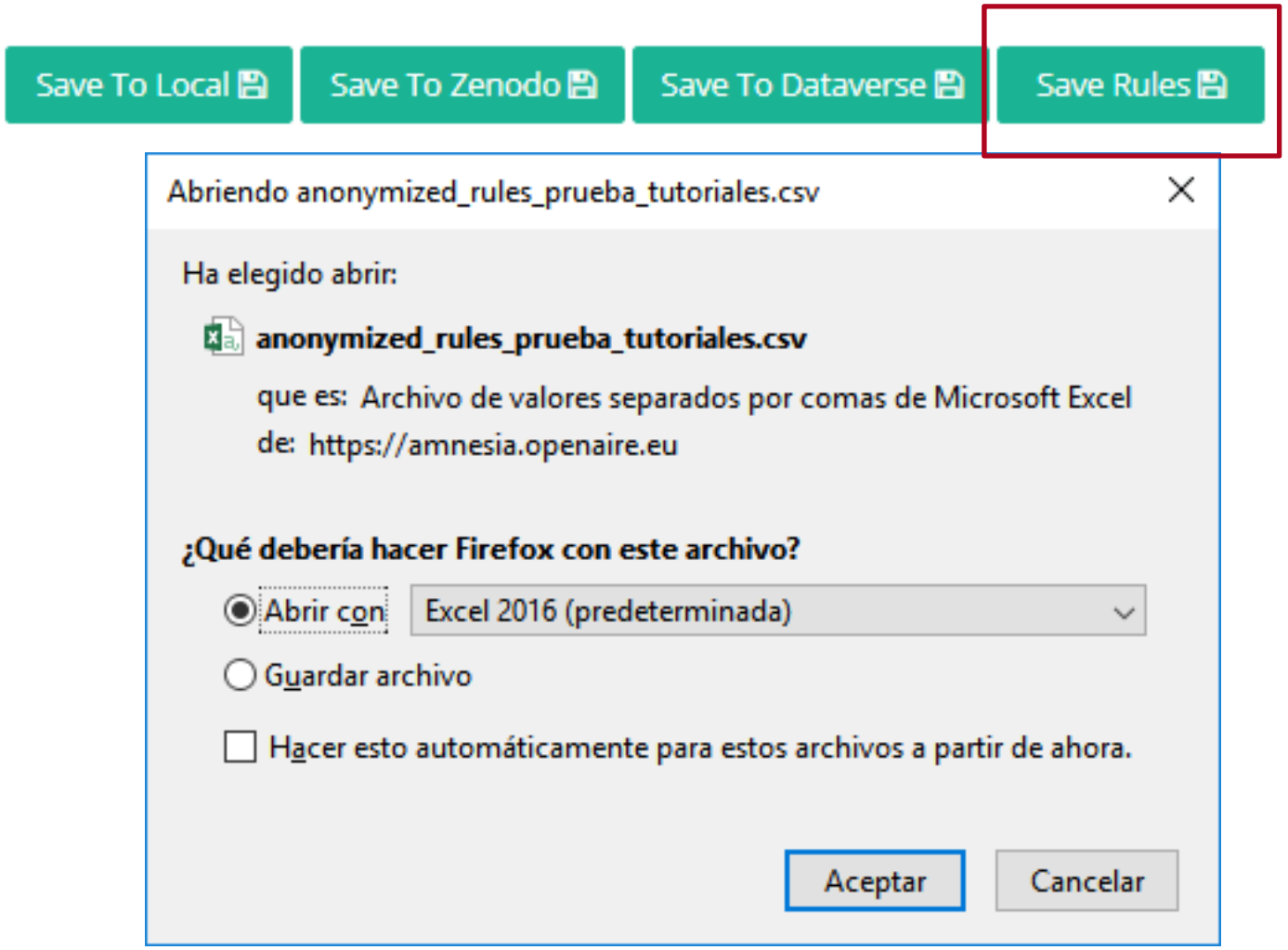
5. Guardar datos anonimizados



DIGITAL.CSIC

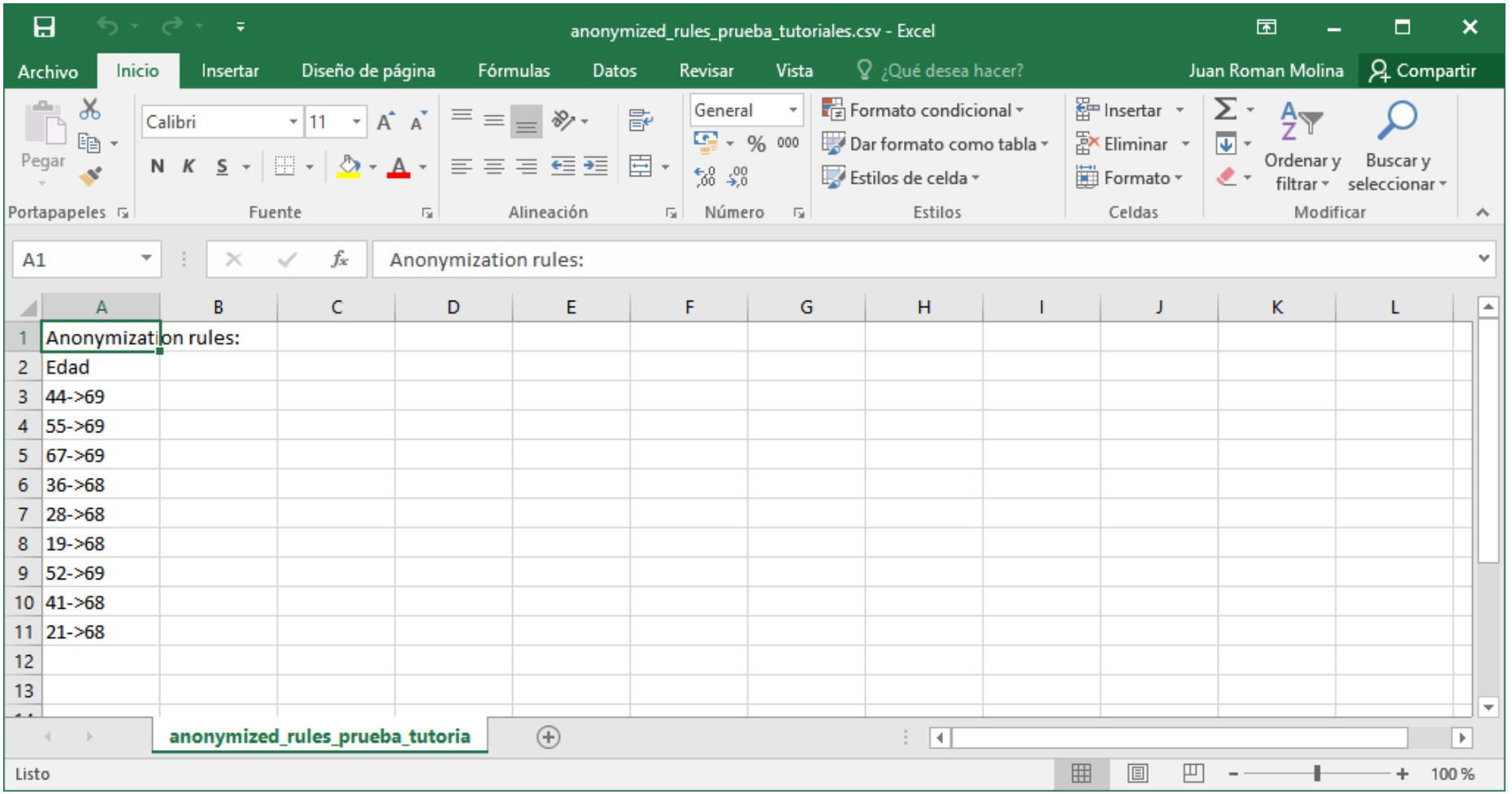
Amnesia

5. Guardar datos anonimizados



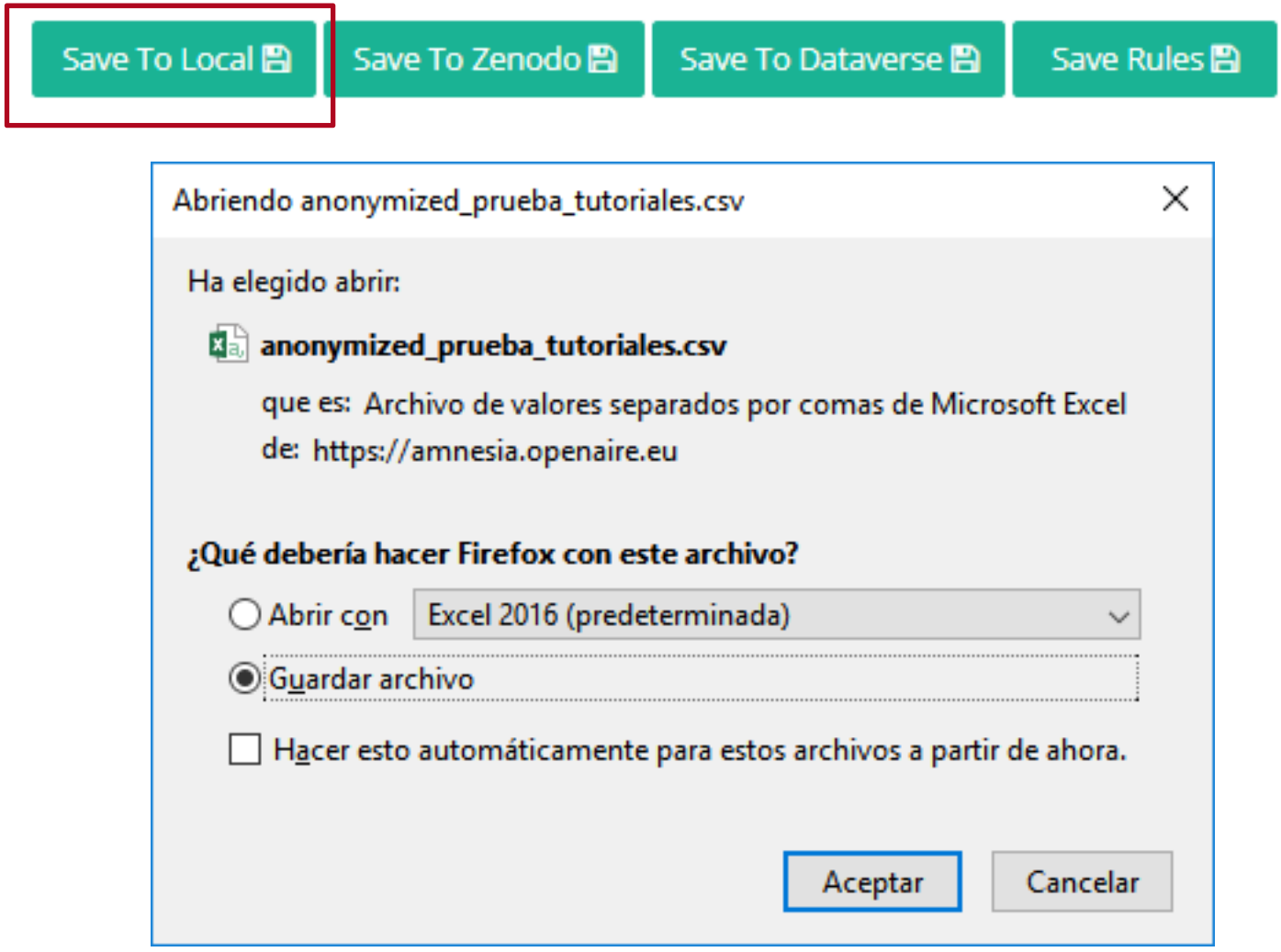
Amnesia

5. Guardar datos anonimizados



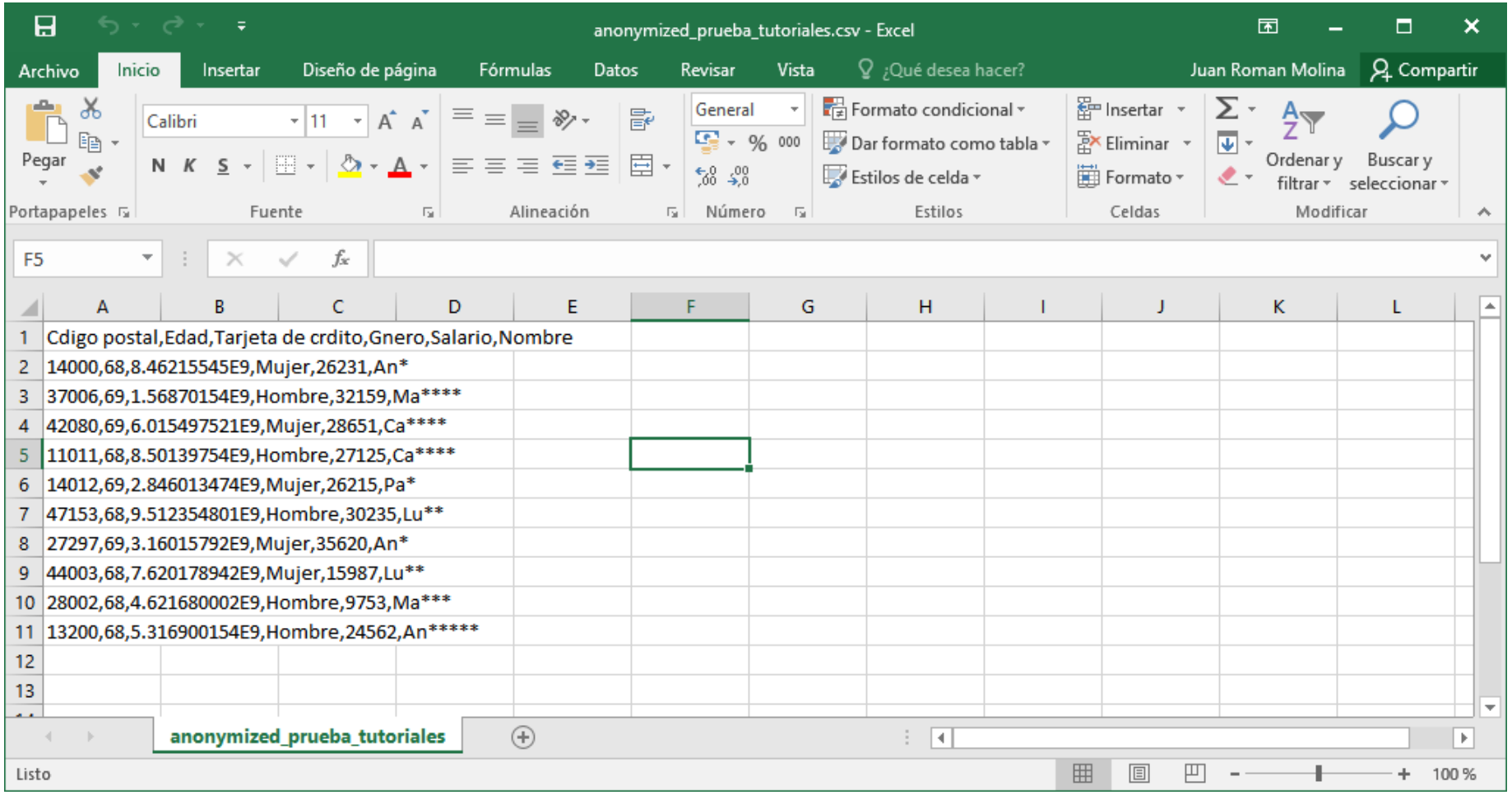
Amnesia

5. Guardar datos anonimizados



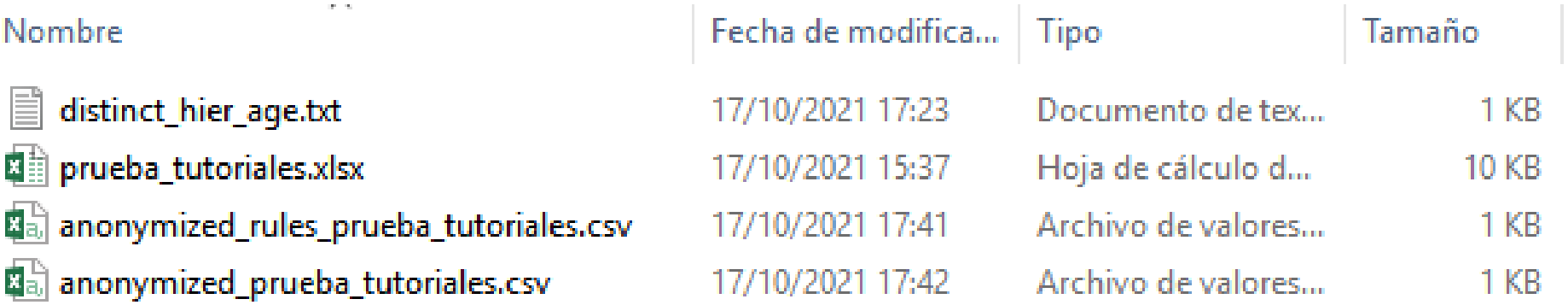
Amnesia

5. Guardar datos anonimizados



Amnesia

5. Guardar datos anonimizados

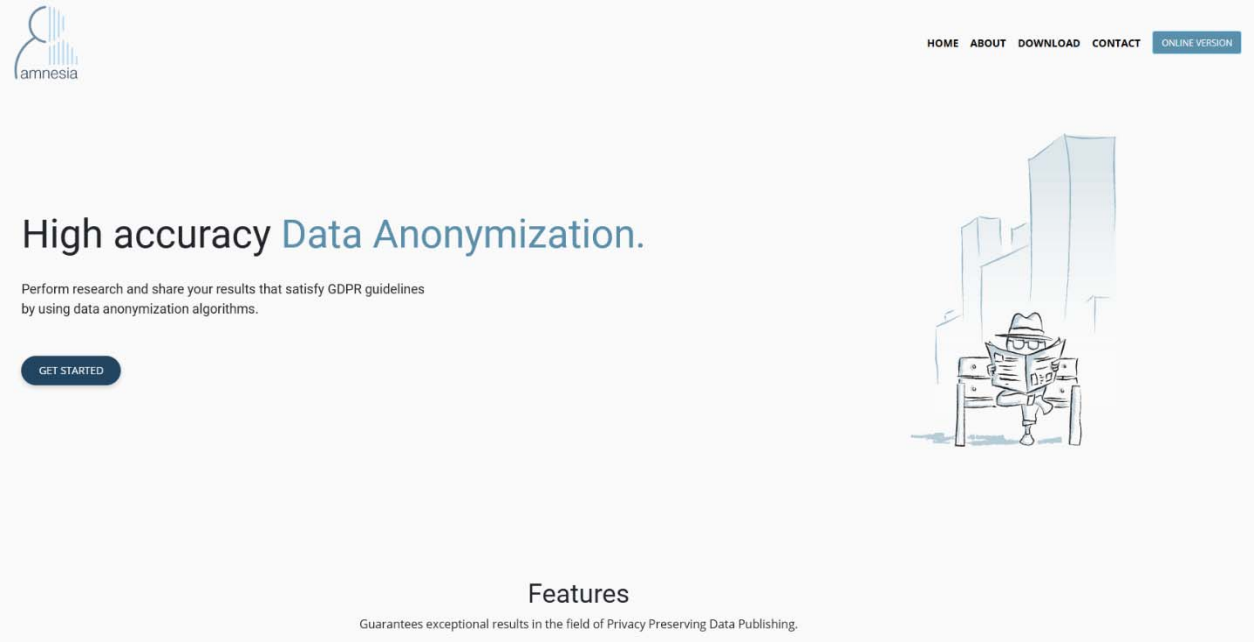


Nombre	Fecha de modifica...	Tipo	Tamaño
distinct_hier_age.txt	17/10/2021 17:23	Documento de tex...	1 KB
prueba_tutoriales.xlsx	17/10/2021 15:37	Hoja de cálculo d...	10 KB
anonymized_rules_prueba_tutoriales.csv	17/10/2021 17:41	Archivo de valores...	1 KB
anonymized_prueba_tutoriales.csv	17/10/2021 17:42	Archivo de valores...	1 KB

DIGITAL.CSIC

Cómo gestionar los datos de investigación: buenas prácticas, estándares, herramientas

PRÁCTICAS



https://amnesia.openaire.eu

DIGITAL.CSIC

Cómo gestionar los datos de investigación: buenas prácticas, estándares, herramientas

1. Primera toma de contacto con la herramienta
2. Trabaja con el ejemplo *simple-table data 1*
3. Sube el dataset
4. Generar jerarquías
5. Anonimiza datos
6. Guarda datos anonimizados

1. Primera toma de contacto con la herramienta
2. Trabaja con el ejemplo *simple-table data 2*
3. Sube el dataset
4. Generar jerarquías
5. Anonimiza datos
6. Guarda datos anonimizados

Cómo gestionar los datos de investigación: buenas prácticas, estándares, herramientas

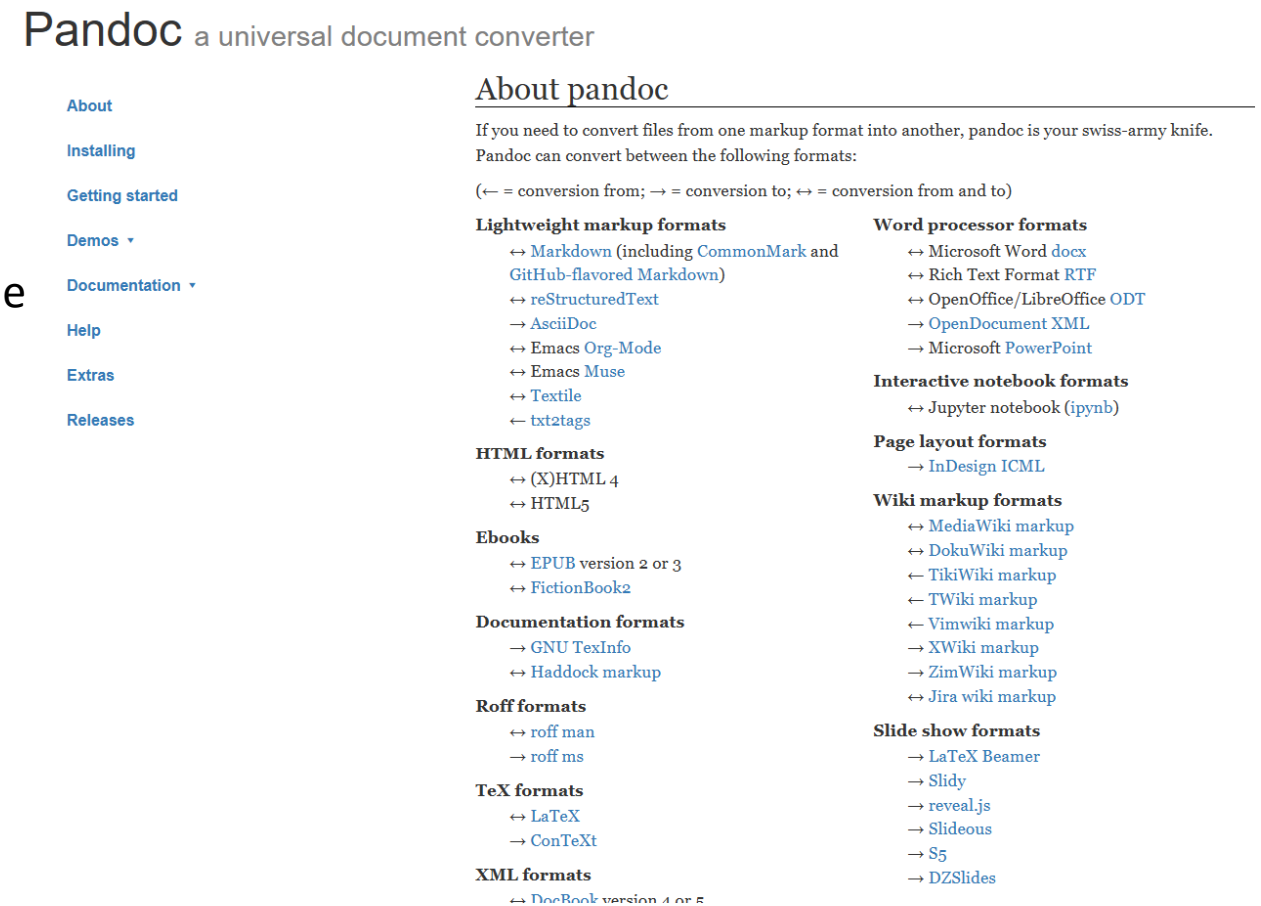
1. Primera toma de contacto con la herramienta
2. Trabaja con el ejemplo *Table with a set-valued attribute data 1*
3. Sube el dataset
4. Generar jerarquías
5. Anonimiza datos
6. Guarda datos anonimizados

Procedure Codes	zipcode	age	creditcard	gender	salary	date
string	int	int	decimal	string	int	date
30|488|860|976	57255	36	5418686973265201	Female	9700	24/8/1966
328	98559	32	5527060358825468	Female	6800	21/11/1971
786|496	28700	58	5312916958971375	Male	4700	11/10/1966
273|365|202|141|522	68925	52	5541858987662877	Male	5700	12/9/1986
318|476|783|917	96338	38	5155271703366251	Female	7100	29/5/2010

DIGITAL.CSIC

Pandoc

1. Panel de trabajo
2. Manual
3. Trabajar con Pandoc online



Pandoc a universal document converter

About
Installing
Getting started
Demos
Documentation
Help
Extras
Releases

About pandoc

If you need to convert files from one markup format into another, pandoc is your swiss-army knife. Pandoc can convert between the following formats:

(← = conversion from; → = conversion to; ↔ = conversion from and to)

Lightweight markup formats
- ↔ Markdown (including CommonMark and GitHub-flavored Markdown)
- ↔ reStructuredText
- → AsciiDoc
- ↔ Emacs Org-Mode
- ↔ Emacs Muse
- ↔ Textile
- ← txt2tags

HTML formats
- ↔ (X)HTML 4
- ↔ HTML5

Ebooks
- ↔ EPUB version 2 or 3
- ↔ FictionBook2

Documentation formats
- → GNU TexInfo
- ↔ Haddock markup

Roff formats
- ↔ roff man
- → roff ms

TeX formats
- ↔ LaTeX
- → ConTeXt

XML formats
- ↔ DocBook version 4 or 5

Word processor formats
- ↔ Microsoft Word docx
- ↔ Rich Text Format RTF
- ↔ OpenOffice/LibreOffice ODT
- → OpenDocument XML
- → Microsoft PowerPoint

Interactive notebook formats
- ↔ Jupyter notebook (ipynb)

Page layout formats
- → InDesign ICML

Wiki markup formats
- ↔ MediaWiki markup
- ↔ DokuWiki markup
- ← TikiWiki markup
- ← TWiki markup
- ← Vimwiki markup
- → XWiki markup
- → ZimWiki markup
- ↔ Jira wiki markup

Slide show formats
- → LaTeX Beamer
- → Slidy
- → reveal.js
- → Slideous
- → S5
- → DZSlides

https://pandoc.org/

Pandoc

Pandoc

1. Panel de trabajo

Pandoc a universal document converter

- About
- Installing
- Getting started
- Demos
  - Try pandoc online
  - Examples
- Documentation
- Help
- Extras
- Releases

About pandoc

If you need to convert files from one markup format into another, pandoc is your swiss-army knife.
Pandoc can convert between the following formats:

(← = conversion from; → = conversion to; ↔ = conversion from and to)

Lightweight markup formats

- ↔ Markdown (including CommonMark and GitHub-flavored Markdown)
- ↔ reStructuredText
- → AsciiDoc
- ↔ Emacs Org-Mode
- ↔ Emacs Muse
- ↔ Textile
- ← txt2tags

HTML formats

- ↔ (X)HTML 4
- ↔ HTML5

Ebooks

- ↔ EPUB version 2 or 3
- ↔ FictionBook2

Documentation formats

- → GNU TexInfo
- ↔ Haddock markup

Roff formats

Word processor formats

- ↔ Microsoft Word docx
- ↔ Rich Text Format RTF
- ↔ OpenOffice/LibreOffice ODT
- → OpenDocument XML
- → Microsoft PowerPoint

Interactive notebook formats

- ↔ Jupyter notebook (ipynb)

Page layout formats

- → InDesign ICML

Wiki markup formats

- ↔ MediaWiki markup
- ↔ DokuWiki markup
- ← TikiWiki markup
- ← TWiki markup
- ← Vimwiki markup
- → XWiki markup
- → ZimWiki markup
- ↔ Jira wiki markup

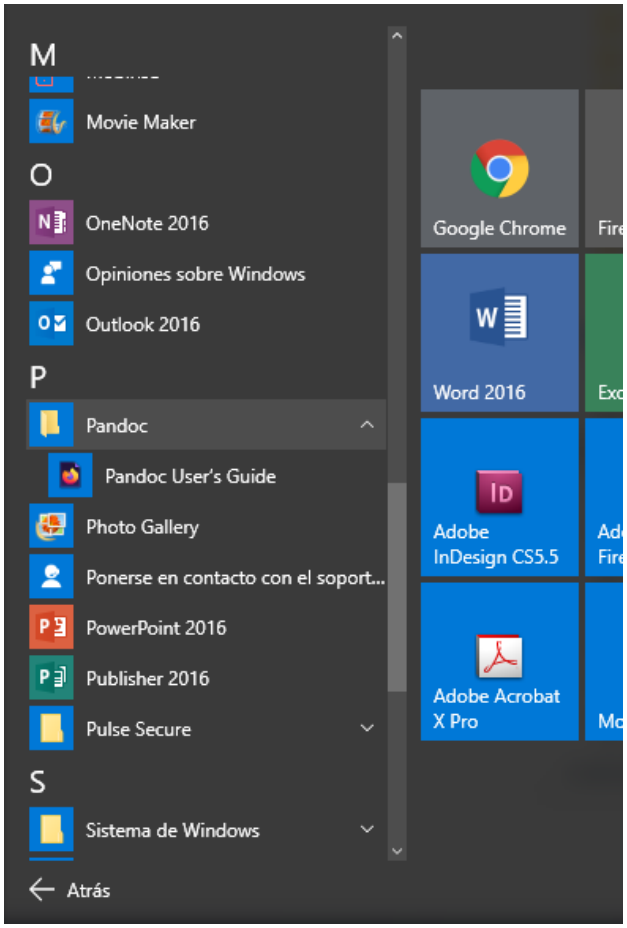
Pandoc

Pandoc

1. Panel de trabajo

Pandoc

2. Manual



Pandoc User's Guide

John MacFarlane

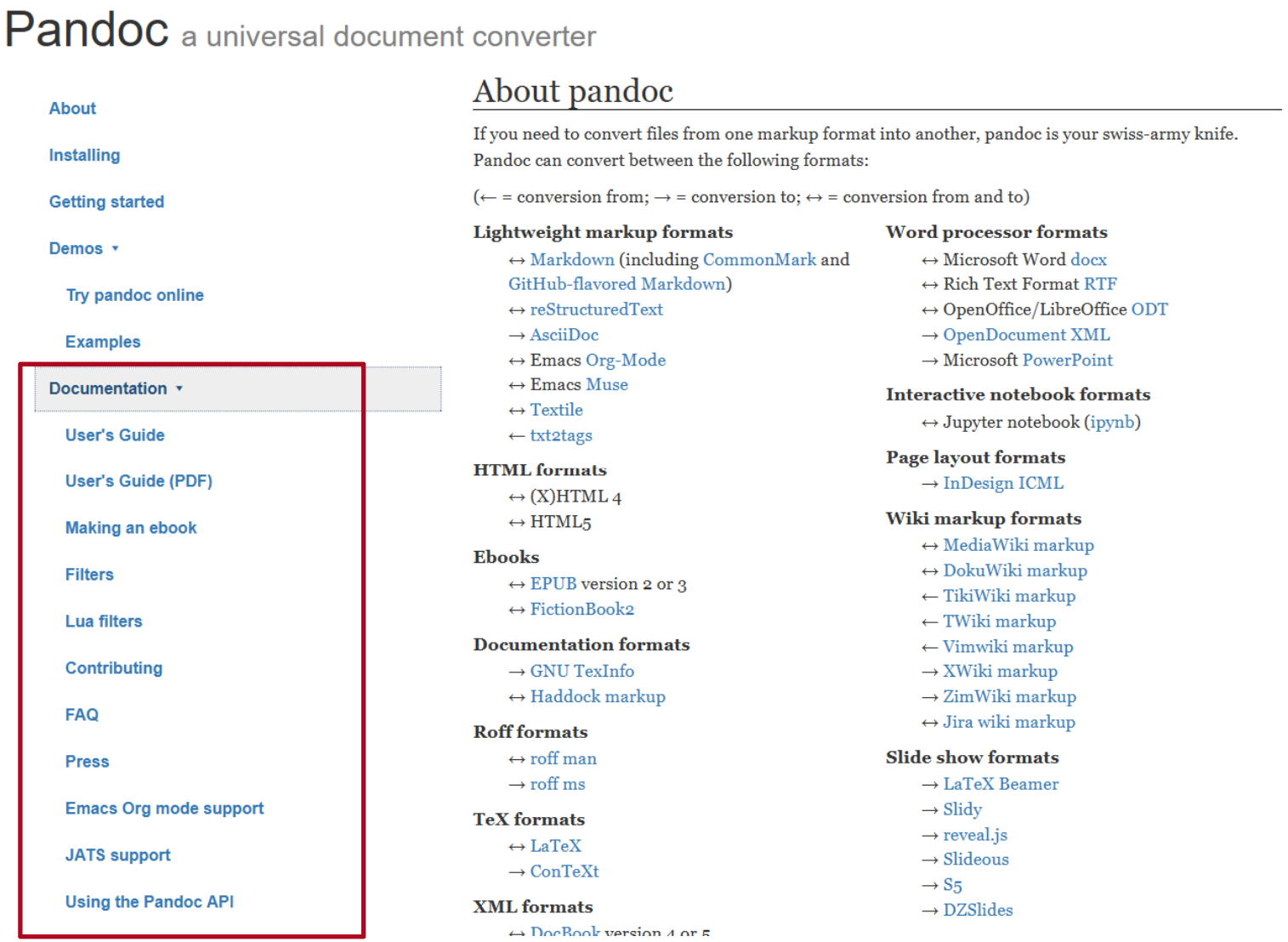
August 20, 2021

Synopsis
Description
- Using pandoc
- Specifying formats
- Character encoding
- Creating a PDF
- Reading from the Web

Options
- General options
- Reader options
- General writer options
- Options affecting specific writers
- Citation rendering
- Math rendering in HTML

Pandoc

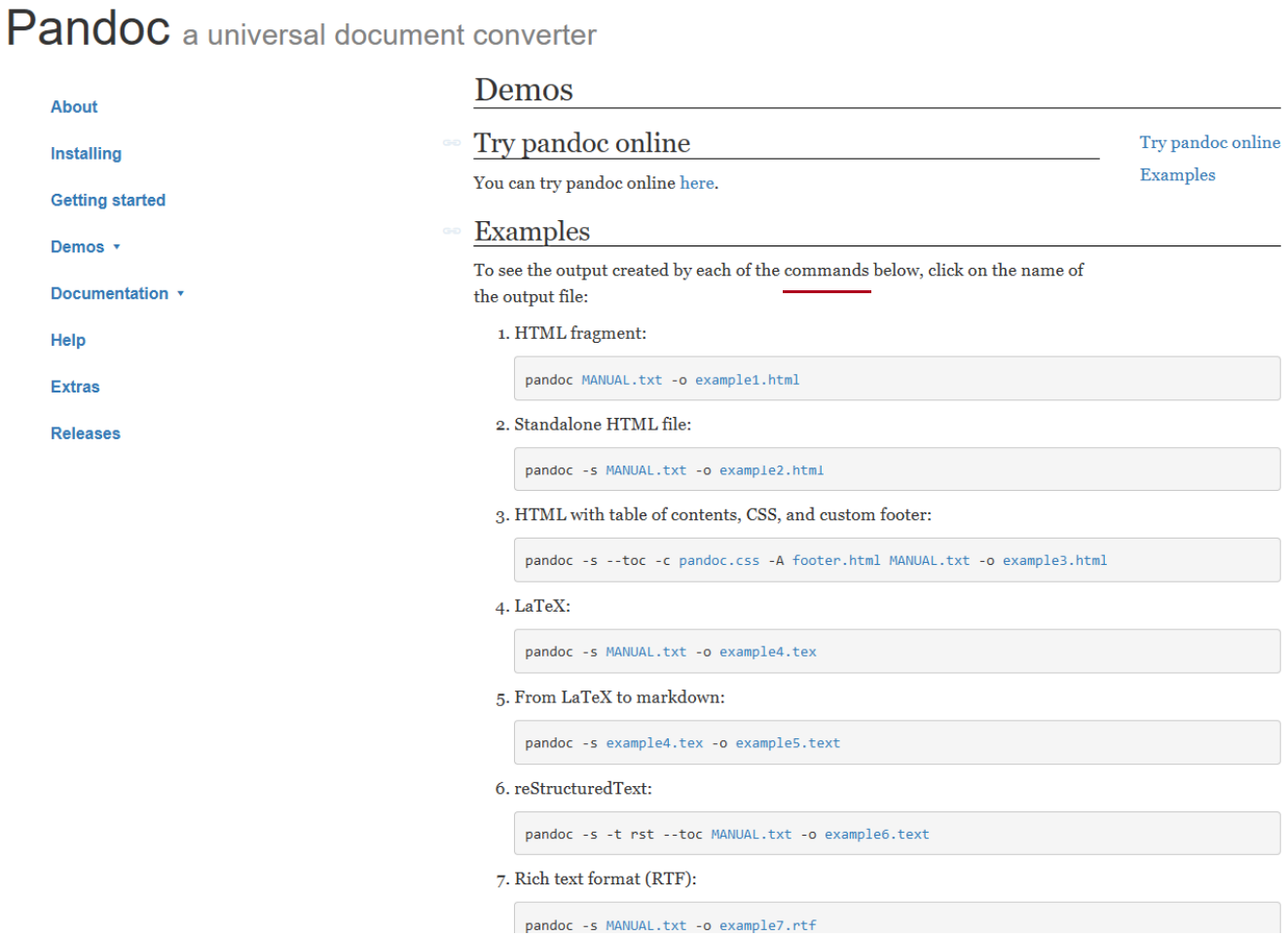
2. Manual



DIGITAL.CSIC

Pandoc

2. Manual



Pandoc a universal document converter

About
Installing
Getting started
Demos
Documentation
Help
Extras
Releases

Demos

Try pandoc online

You can try pandoc online here.

Examples

To see the output created by each of the commands below, click on the name of the output file:

1. HTML fragment:

```
pandoc MANUAL.txt -o example1.html
```

2. Standalone HTML file:

```
pandoc -s MANUAL.txt -o example2.html
```

3. HTML with table of contents, CSS, and custom footer:

```
pandoc -s --toc -c pandoc.css -A footer.html MANUAL.txt -o example3.html
```

4. LaTeX:

```
pandoc -s MANUAL.txt -o example4.tex
```

5. From LaTeX to markdown:

```
pandoc -s example4.tex -o example5.text
```

6. reStructuredText:

```
pandoc -s -t rst --toc MANUAL.txt -o example6.text
```

7. Rich text format (RTF):

```
pandoc -s MANUAL.txt -o example7.rtf
```

Try pandoc online
Examples

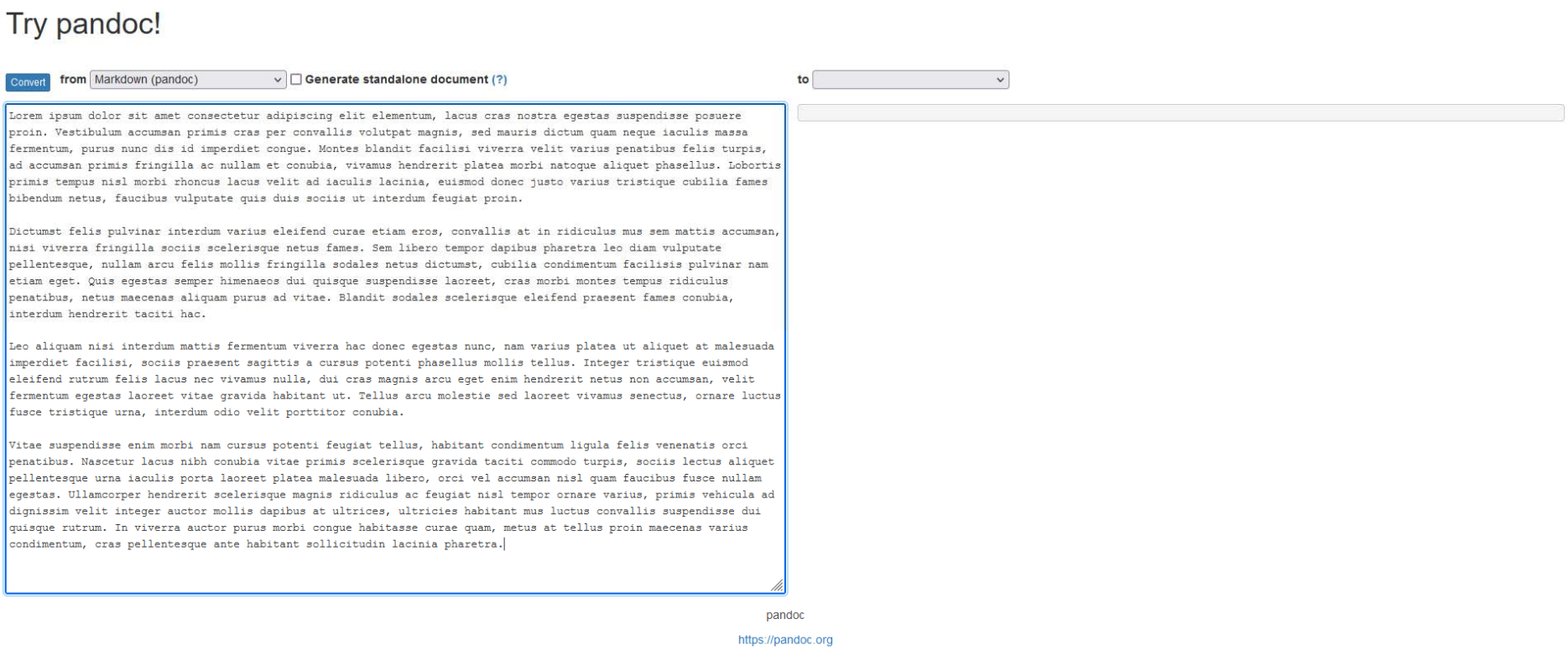
https://pandoc.org/demos.html

Pandoc

3. Trabajar con Pandoc online

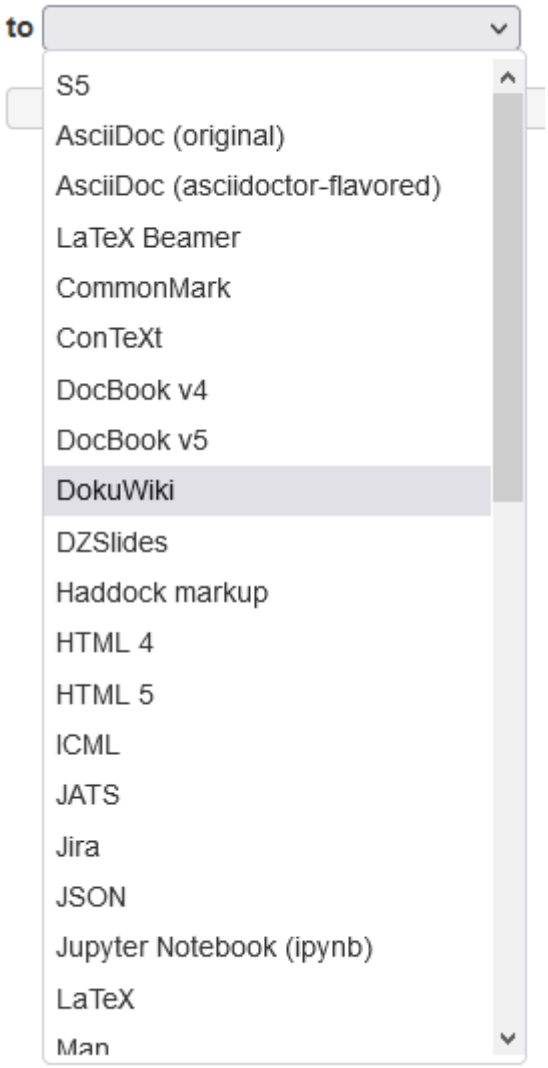
Pandoc

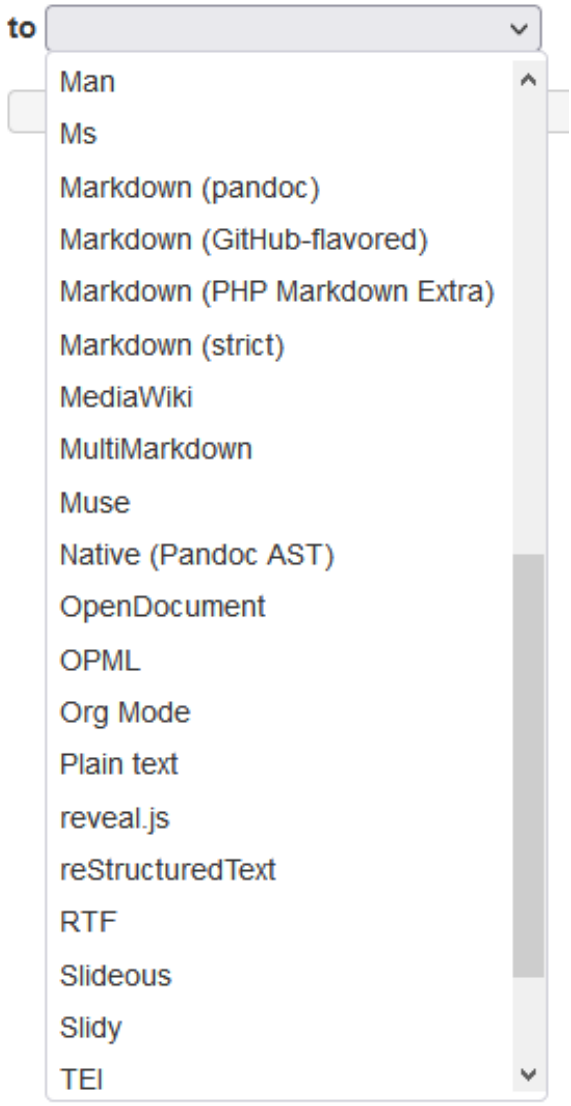
3. Trabajar con Pandoc online

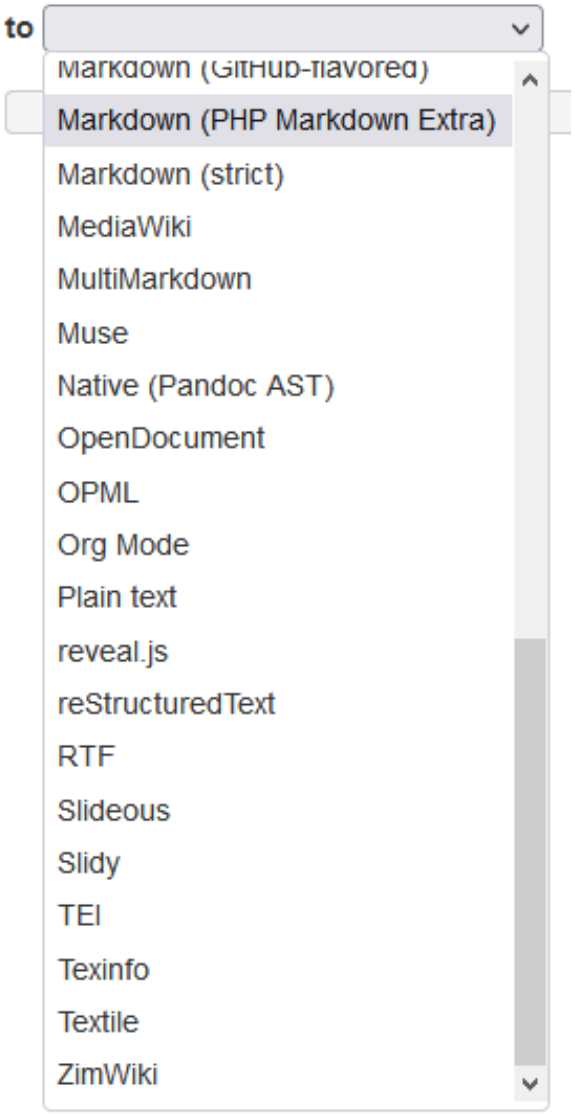


Pandoc

3. Trabajar con Pandoc online

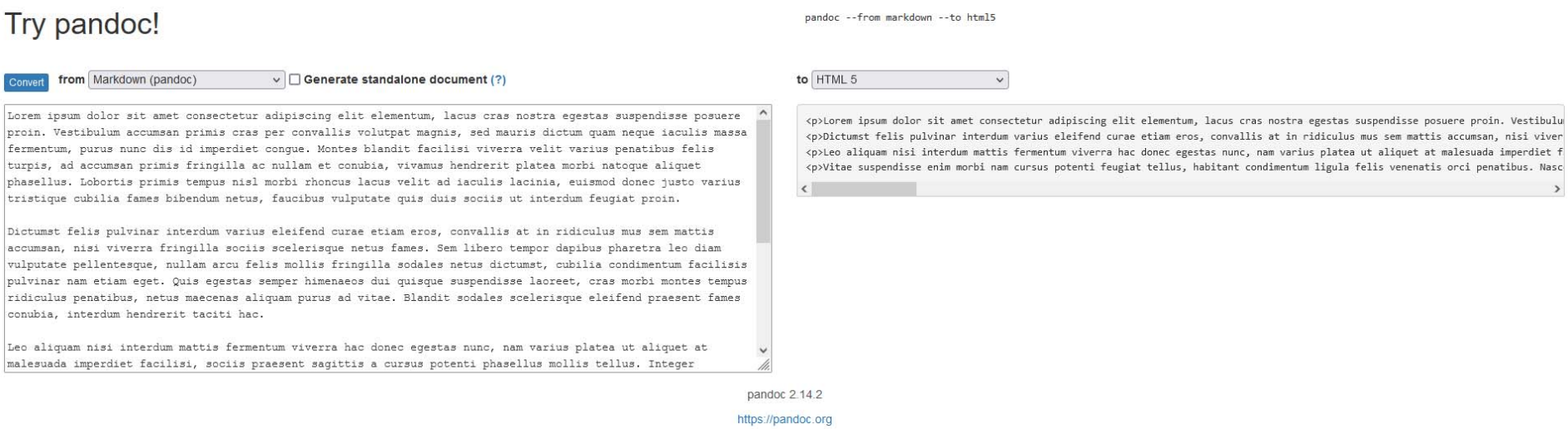






Pandoc

3. Trabajar con Pandoc online



Pandoc

3. Trabajar con Pandoc online

Try pandoc!

```
pandoc --from markdown --to docbook5
```

Convert **from** Markdown (pandoc) ☐ **Generate standalone document** (?) **to** DocBook v5

```
Lorem ipsum dolor sit amet consectetur adipiscing elit elementum, lacus cras nostra egestas suspendisse posuere
proin. Vestibulum accumsan primis cras per convallis volutpat magnis, sed mauris dictum quam neque iaculis massa
fermentum, purus nunc dis id imperdiet congue. Montes blandit facilisi viverra velit varius penatibus felis
turpis, ad accumsan primis fringilla ac nullam et conubia, vivamus hendrerit platea morbi natoque aliquet
phasellus. Lobortis primis tempus nisl morbi rhoncus lacus velit ad iaculis lacinia, euismod donec justo varius
tristique cubilia fames bibendum netus, faucibus vulputate quis duis sociis ut interdum feugiat proin.

Dictumst felis pulvinar interdum varius eleifend curae etiam eros, convallis at in ridiculus mus sem mattis
accumsan, nisi viverra fringilla sociis scelerisque netus fames. Sem libero tempor dapibus pharetra leo diam
vulputate pellentesque, nullam arcu felis mollis fringilla sodales netus dictumst, cubilia condimentum facilisis
pulvinar nam etiam eget. Quis egestas semper himenaeos dui quisque suspendisse laoreet, cras morbi montes tempus
ridiculus penatibus, netus maecenas aliquam purus ad vitae. Blandit sodales scelerisque eleifend praesent fames
conubia, interdum hendrerit taciti hac.

Leo aliquam nisi interdum mattis fermentum viverra hac donec egestas nunc, nam varius platea ut aliquet at
malesuada imperdiet facilisi, sociis praesent sagittis a cursus potenti phasellus mollis tellus. Integer
```

```
<para>
  Lorem ipsum dolor sit amet consectetur adipiscing elit elementum,
  lacus cras nostra egestas suspendisse posuere proin. Vestibulum
  accumsan primis cras per convallis volutpat magnis, sed mauris dictum
  quam neque iaculis massa fermentum, purus nunc dis id imperdiet
  congue. Montes blandit facilisi viverra velit varius penatibus felis
  turpis, ad accumsan primis fringilla ac nullam et conubia, vivamus
  hendrerit platea morbi natoque aliquet phasellus. Lobortis primis
  tempus nisl morbi rhoncus lacus velit ad iaculis lacinia, euismod
  donec justo varius tristique cubilia fames bibendum netus, faucibus
  vulputate quis duis sociis ut interdum feugiat proin.
</para>
<para>
  Dictumst felis pulvinar interdum varius eleifend curae etiam eros,
  convallis at in ridiculus mus sem mattis accumsan, nisi viverra
  fringilla sociis scelerisque netus fames. Sem libero tempor dapibus
  pharetra leo diam vulputate pellentesque, nullam arcu felis mollis
  fringilla sodales netus dictumst, cubilia condimentum facilisis
  pulvinar nam etiam eget. Quis egestas semper himenaeos dui quisque
  suspendisse laoreet, cras morbi montes tempus ridiculus penatibus,
  netus maecenas aliquam purus ad vitae. Blandit sodales scelerisque
  eleifend praesent fames conubia, interdum hendrerit taciti hac.
</para>
<para>
  Leo aliquam nisi interdum mattis fermentum viverra hac donec egestas
  nunc, nam varius platea ut aliquet at malesuada imperdiet facilisi,
  sociis praesent sagittis a cursus potenti phasellus mollis tellus.
  Integer tristique euismod eleifend rutrum felis lacus nec vivamus
  nulla, dui cras magnis arcu eget enim hendrerit netus non accumsan,
  velit fermentum egestas laoreet vitae gravida habitant ut. Tellus arcu
  molestie sed laoreet vivamus senectus, ornare luctus fusce tristique
  urna, interdum odio velit porttitor conubia.
</para>
<para>
  Vitae suspendisse enim morbi nam cursus potenti feugiat tellus,
  habitant condimentum ligula felis venenatis orci penatibus. Nascetur
  lacus nibh conubia vitae primis scelerisque gravida taciti commodo
  turpis, sociis lectus aliquet pellentesque urna iaculis porta laoreet
  platea malesuada libero, orci vel accumsan nisl quam faucibus fusce
  nullam egestas. Ullamcorper hendrerit scelerisque magnis ridiculus ac
  feugiat nisl tempor ornare varius, primis vehicula ad dignissim velit
  integer auctor mollis dapibus at ultrices, ultricies habitant mus
```

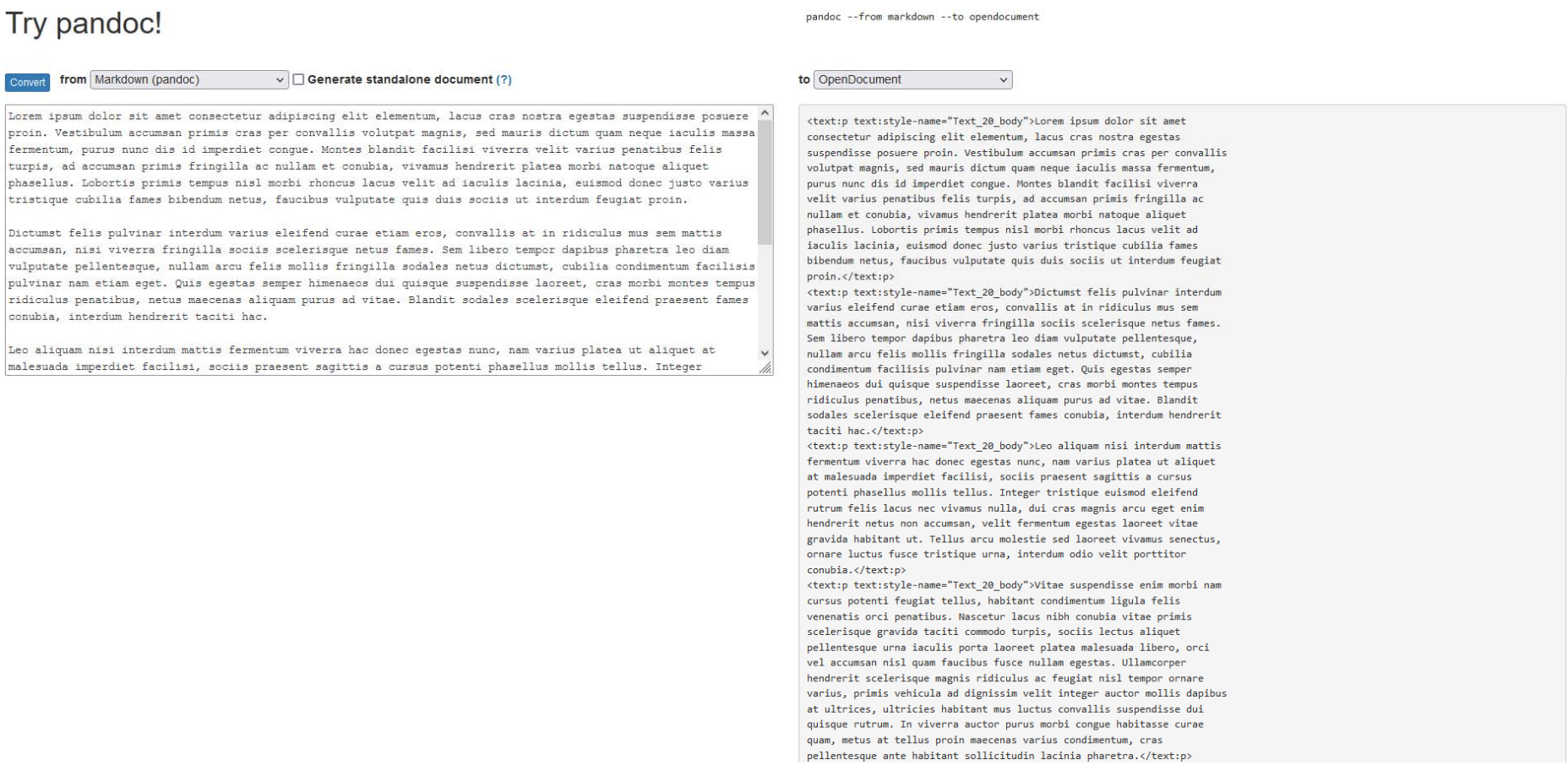
Pandoc

3. Trabajar con Pandoc online

DIGITAL.CSIC

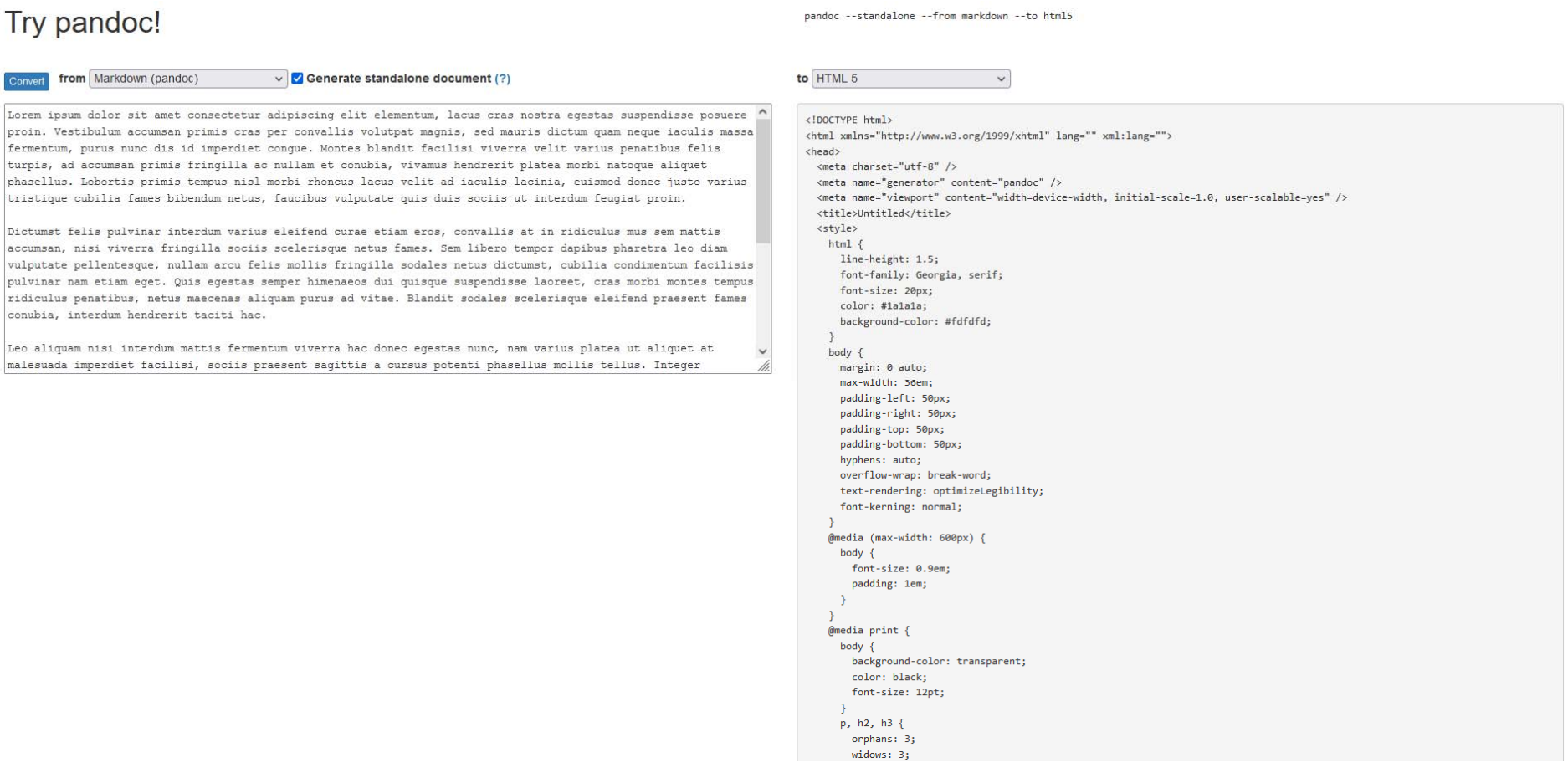
Pandoc

3. Trabajar con Pandoc online



Pandoc

3. Trabajar con Pandoc online



Pandoc

3. Trabajar con Pandoc online

`-s` , `--standalone`

Produce output with an appropriate header and footer (e.g. a standalone HTML, LaTeX, TEI, or RTF file, not a fragment). This option is set automatically for `pdf` , `epub` , `epub3` , `fb2` , `docx` , and `odt` output. For `native` output, this option causes metadata to be included; otherwise, metadata is suppressed.

`--template=` *FILE*|*URL*

Use the specified file as a custom template for the generated document. Implies `--standalone` . See Templates, below, for a description of template syntax. If no extension is specified, an extension corresponding to the writer will be added, so that `--template=special` looks for `special.html` for HTML output. If the template is not found, pandoc will search for it in the `templates` subdirectory of the user data directory (see `--data-dir`). If this option is not used, a default template appropriate for the output format will be used (see `-D/--print-default-template`).

`-V` *KEY*[`=`*VAL*], `--variable=` *KEY*[`:`*VAL*]

Set the template variable *KEY* to the value *VAL* when rendering the document in standalone mode. If no *VAL* is specified, the key will be given the value `true` .

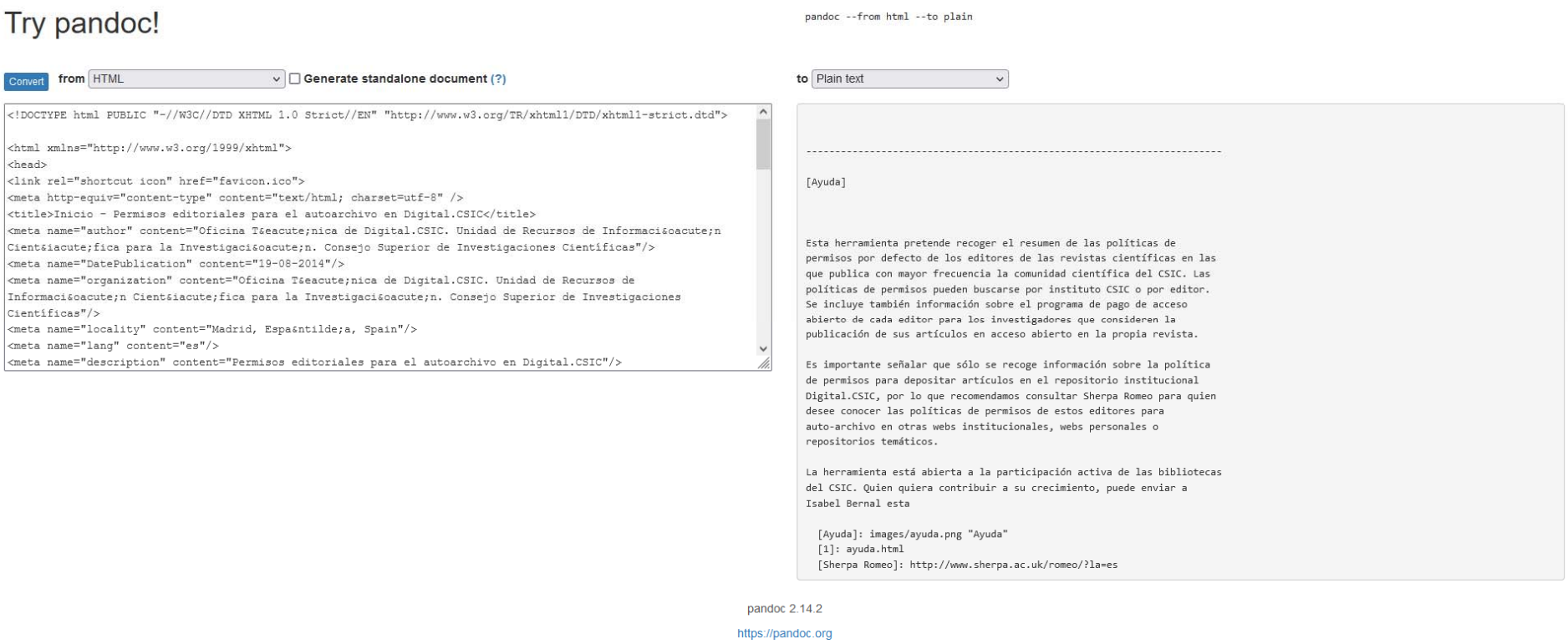
`-D` *FORMAT*, `--print-default-template=` *FORMAT*

Print the system default template for an output *FORMAT*. (See `-t` for a list of possible *FORMAT*s.) Templates in the user data directory are ignored. This option may be used with `-o` / `--output` to redirect output to a file, but `-o` / `--output` must come before `--print-default-template` on the command line.

Note that some of the default templates use partials, for example `styles.html` . To print the partials, use `--print-default-data-file` : for example, `--print-default-data-file=templates/styles.html` .

Pandoc

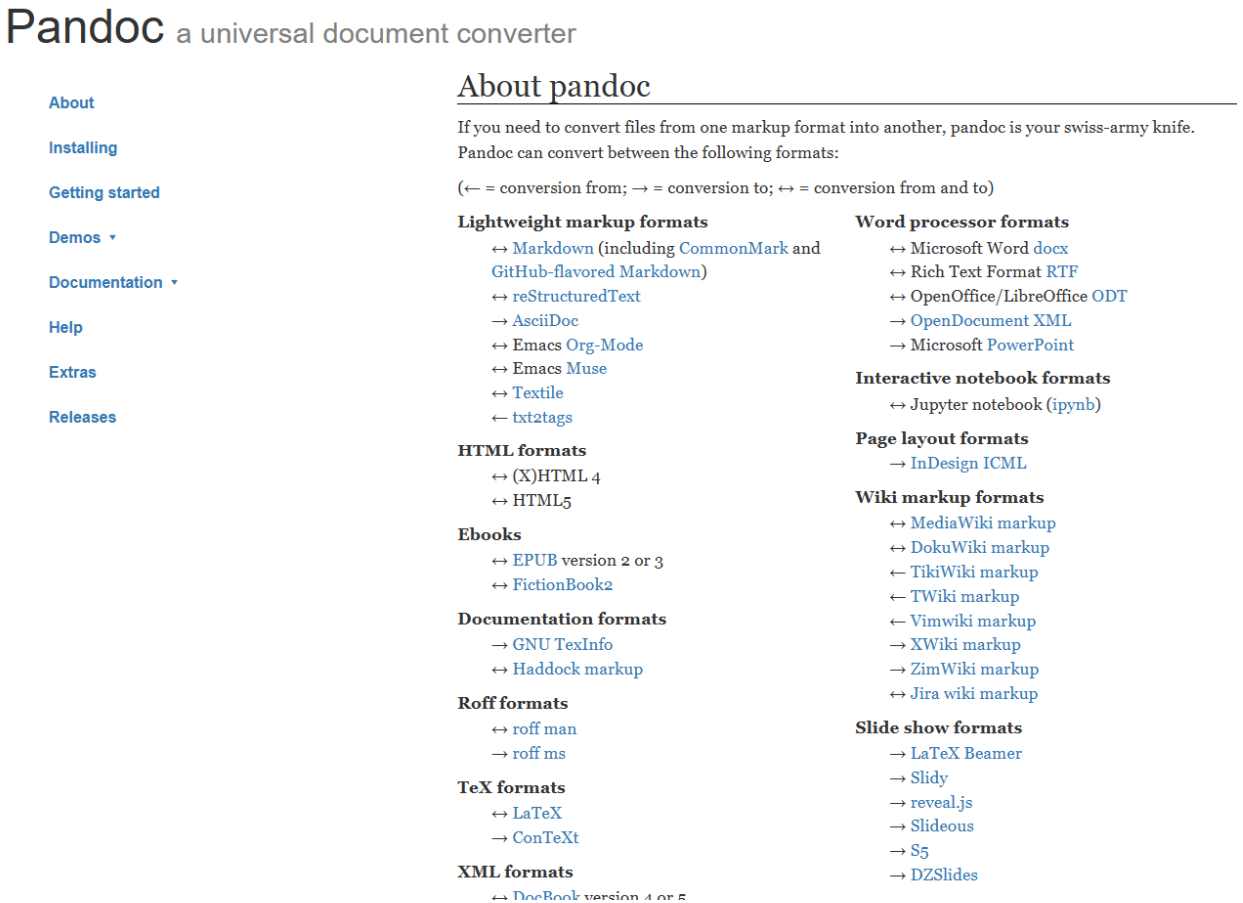
3. Trabajar con Pandoc online



DIGITAL.CSIC

Cómo gestionar los datos de investigación: buenas prácticas, estándares, herramientas

PROGRAMA



Pandoc a universal document converter

About
Installing
Getting started
Demos
Documentation
Help
Extras
Releases

About pandoc

If you need to convert files from one markup format into another, pandoc is your swiss-army knife.
Pandoc can convert between the following formats:

(← = conversion from; → = conversion to; ↔ = conversion from and to)

Lightweight markup formats
- ↔ Markdown (including CommonMark and GitHub-flavored Markdown)
- ↔ reStructuredText
- → AsciiDoc
- ↔ Emacs Org-Mode
- ↔ Emacs Muse
- ↔ Textile
- ← txt2tags

HTML formats
- ↔ (X)HTML 4
- ↔ HTML5

Ebooks
- ↔ EPUB version 2 or 3
- ↔ FictionBook2

Documentation formats
- → GNU TexInfo
- ↔ Haddock markup

Roff formats
- ↔ roff man
- → roff ms

TeX formats
- ↔ LaTeX
- → ConTeXt

XML formats
- ↔ DocBook version 4 or 5

Word processor formats
- ↔ Microsoft Word docx
- ↔ Rich Text Format RTF
- ↔ OpenOffice/LibreOffice ODT
- → OpenDocument XML
- → Microsoft PowerPoint

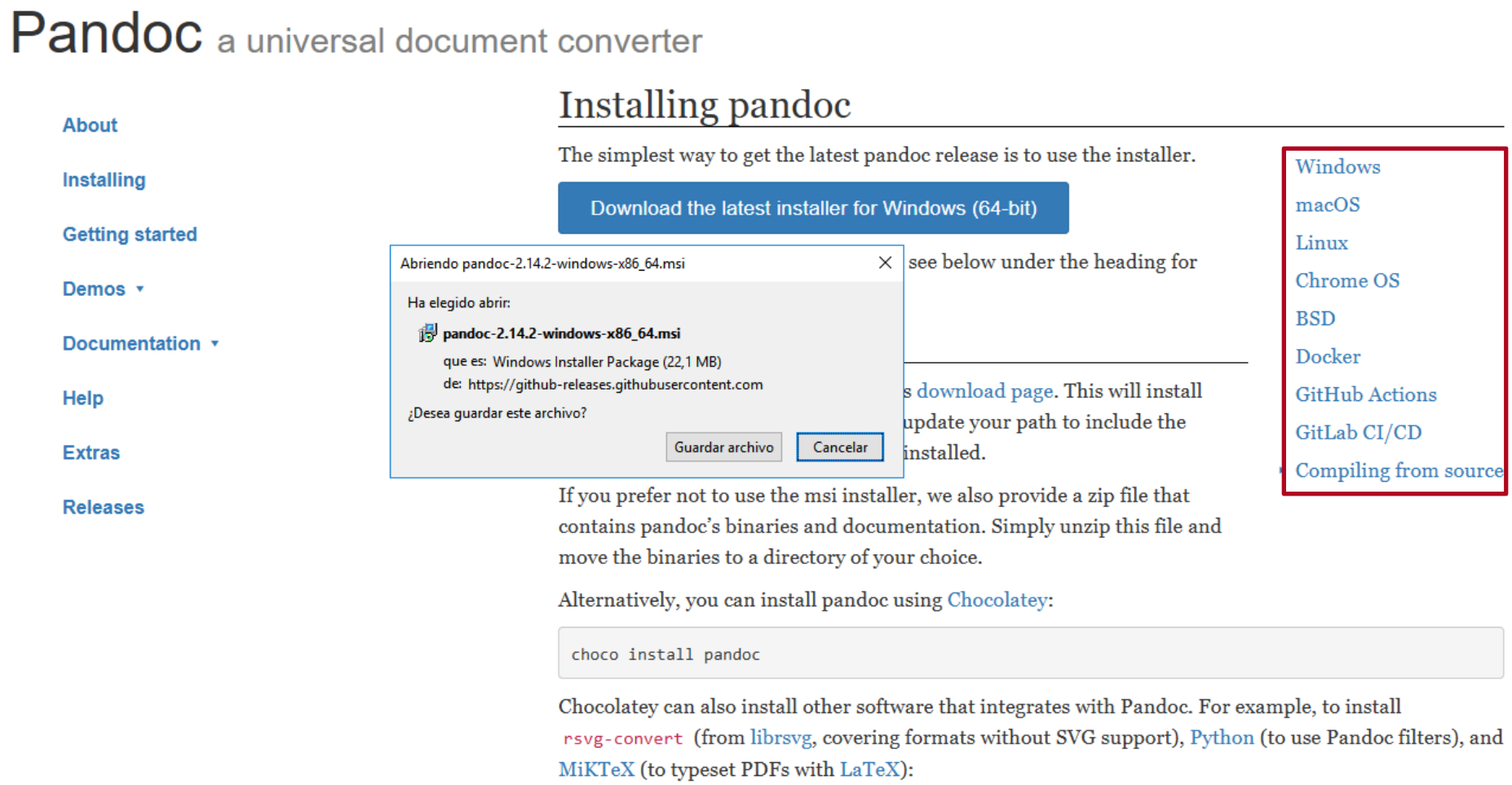
Interactive notebook formats
- ↔ Jupyter notebook (ipynb)

Page layout formats
- → InDesign ICML

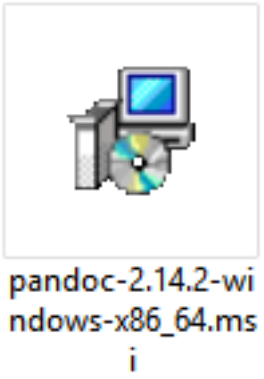
Wiki markup formats
- ↔ MediaWiki markup
- ↔ DokuWiki markup
- ← TikiWiki markup
- ← TWiki markup
- ← Vimwiki markup
- → XWiki markup
- → ZimWiki markup
- ↔ Jira wiki markup

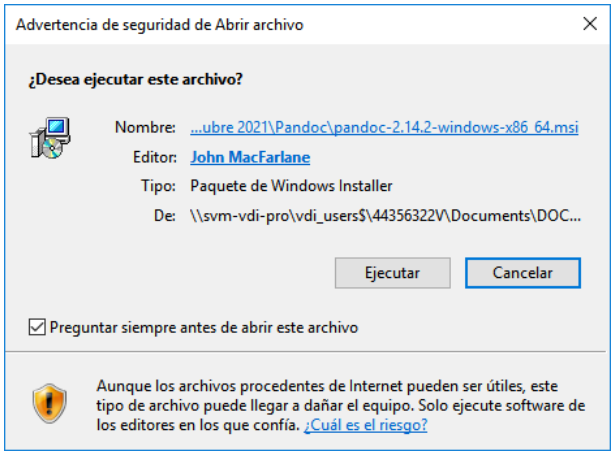
Slide show formats
- → LaTeX Beamer
- → Slidy
- → reveal.js
- → Slideous
- → S5
- → DZSlides

https://pandoc.org/



https://pandoc.org/installing.html





Step 1: Install pandoc

First, install pandoc, following the instructions for your platform.

Step 2: Open a terminal

Pandoc is a command-line tool. There is no graphic user interface.
So, to use it, you'll need to open a terminal window:

https://pandoc.org/getting-started.html

Pandoc a universal document converter

About
Installing
Getting started
Demos ▾
Try pandoc online
Examples
Documentation ▾
Help
Extras
Releases

About pandoc

If you need to convert files from one markup format into another, pandoc is your swiss-army knife.
Pandoc can convert between the following formats:

(← = conversion from; → = conversion to; ↔ = conversion from and to)

Lightweight markup formats
- ↔ Markdown (including CommonMark and GitHub-flavored Markdown)
- ↔ reStructuredText
- → AsciiDoc
- ↔ Emacs Org-Mode
- ↔ Emacs Muse
- ↔ Textile
- ← txt2tags

HTML formats
- ↔ (X)HTML 4
- ↔ HTML5

Ebooks
- ↔ EPUB version 2 or 3
- ↔ FictionBook2

Documentation formats
- → GNU TexInfo
- ↔ Haddock markup

Roff formats

Word processor formats
- ↔ Microsoft Word docx
- ↔ Rich Text Format RTF
- ↔ OpenOffice/LibreOffice ODT
- → OpenDocument XML
- → Microsoft PowerPoint

Interactive notebook formats
- ↔ Jupyter notebook (ipynb)

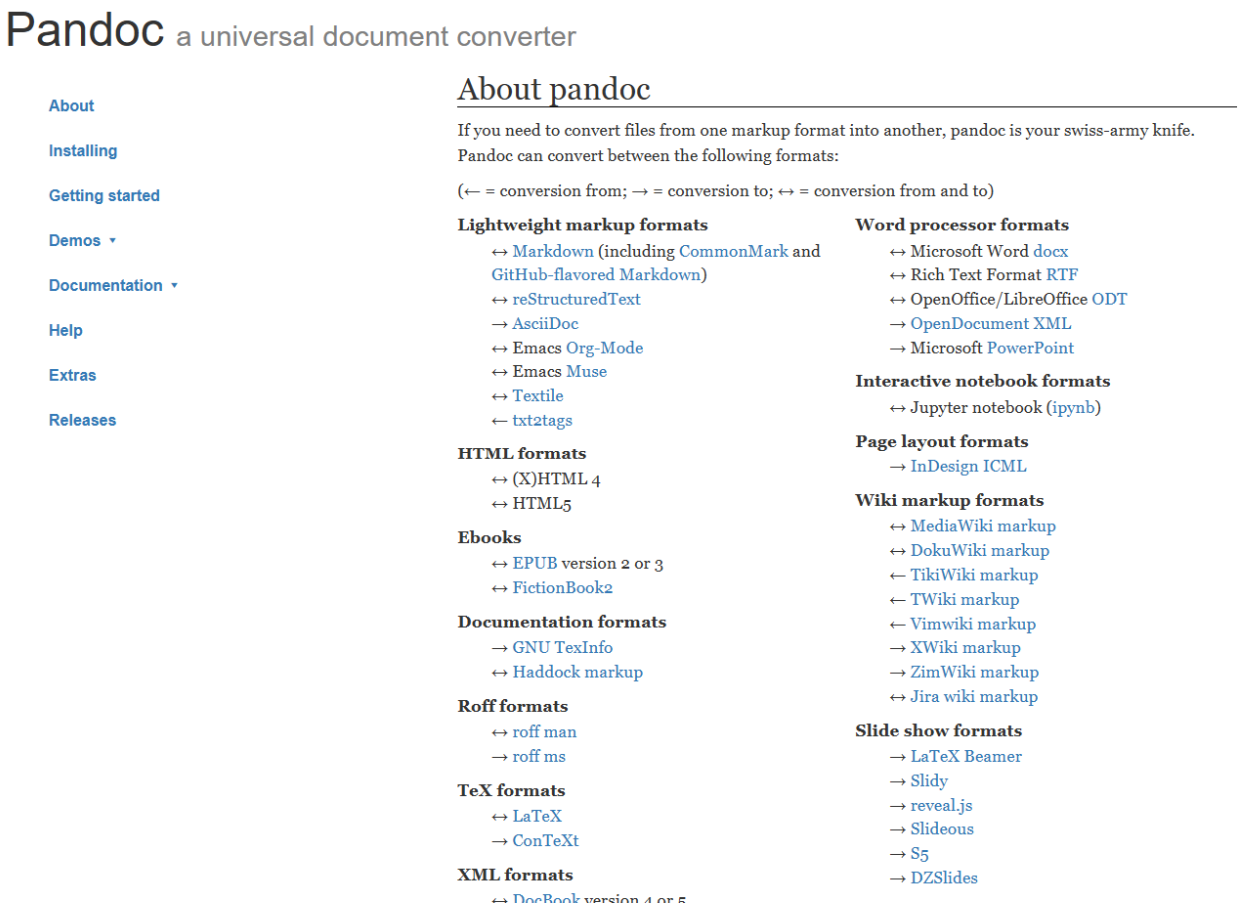
Page layout formats
- → InDesign ICML

Wiki markup formats
- ↔ MediaWiki markup
- ↔ DokuWiki markup
- ← TikiWiki markup
- ← TWiki markup
- ← Vimwiki markup
- → XWiki markup
- → ZimWiki markup
- ↔ Jira wiki markup

DIGITAL.CSIC

Cómo gestionar los datos de investigación: buenas prácticas, estándares, herramientas

PRÁCTICAS



Pandoc a universal document converter

About
Installing
Getting started
Demos
Documentation
Help
Extras
Releases

About pandoc

If you need to convert files from one markup format into another, pandoc is your swiss-army knife.
Pandoc can convert between the following formats:

(← = conversion from; → = conversion to; ↔ = conversion from and to)

Lightweight markup formats
- ↔ Markdown (including CommonMark and GitHub-flavored Markdown)
- ↔ reStructuredText
- → AsciiDoc
- ↔ Emacs Org-Mode
- ↔ Emacs Muse
- ↔ Textile
- ← txt2tags

HTML formats
- ↔ (X)HTML 4
- ↔ HTML5

Ebooks
- ↔ EPUB version 2 or 3
- ↔ FictionBook2

Documentation formats
- → GNU TexInfo
- ↔ Haddock markup

Roff formats
- ↔ roff man
- → roff ms

TeX formats
- ↔ LaTeX
- → ConTeXt

XML formats
- ↔ DocBook version 4 or 5

Word processor formats
- ↔ Microsoft Word docx
- ↔ Rich Text Format RTF
- ↔ OpenOffice/LibreOffice ODT
- → OpenDocument XML
- → Microsoft PowerPoint

Interactive notebook formats
- ↔ Jupyter notebook (ipynb)

Page layout formats
- → InDesign ICML

Wiki markup formats
- ↔ MediaWiki markup
- ↔ DokuWiki markup
- ← TikiWiki markup
- ← TWiki markup
- ← Vimwiki markup
- → XWiki markup
- → ZimWiki markup
- ↔ Jira wiki markup

Slide show formats
- → LaTeX Beamer
- → Slidy
- → reveal.js
- → Slideous
- → S5
- → DZSlides

https://pandoc.org/

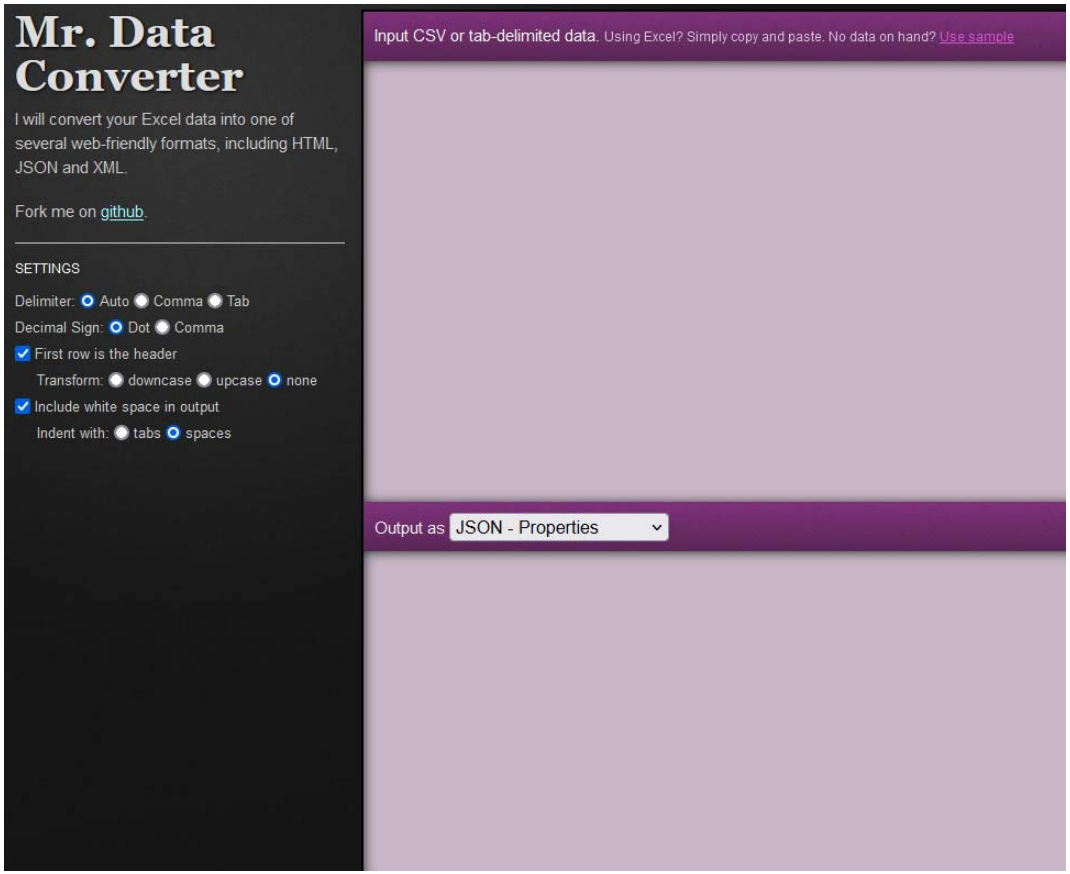
Pandoc

1. Consulta formatos con los que trabaja Pandoc https://pandoc.org/
2. Consulta la demo de ejemplos: https://pandoc.org/demos.html
3. Trabaja con Try Pandoc online https://pandoc.org/try/

DIGITAL.CSIC

Mr. Data Converter

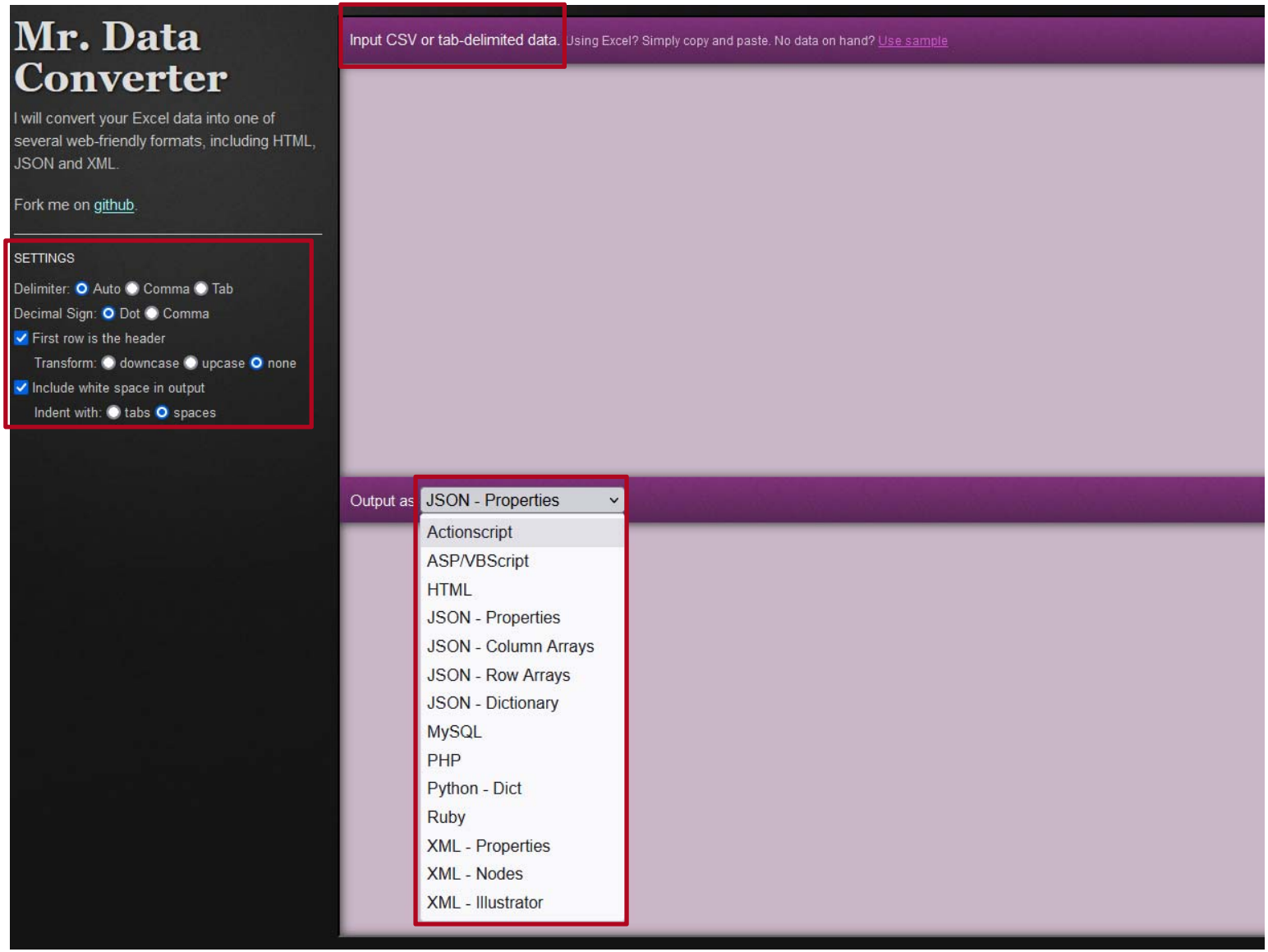
1. Panel de trabajo
2. Ejemplos de uso



https://shancarter.github.io/mr-data-converter/

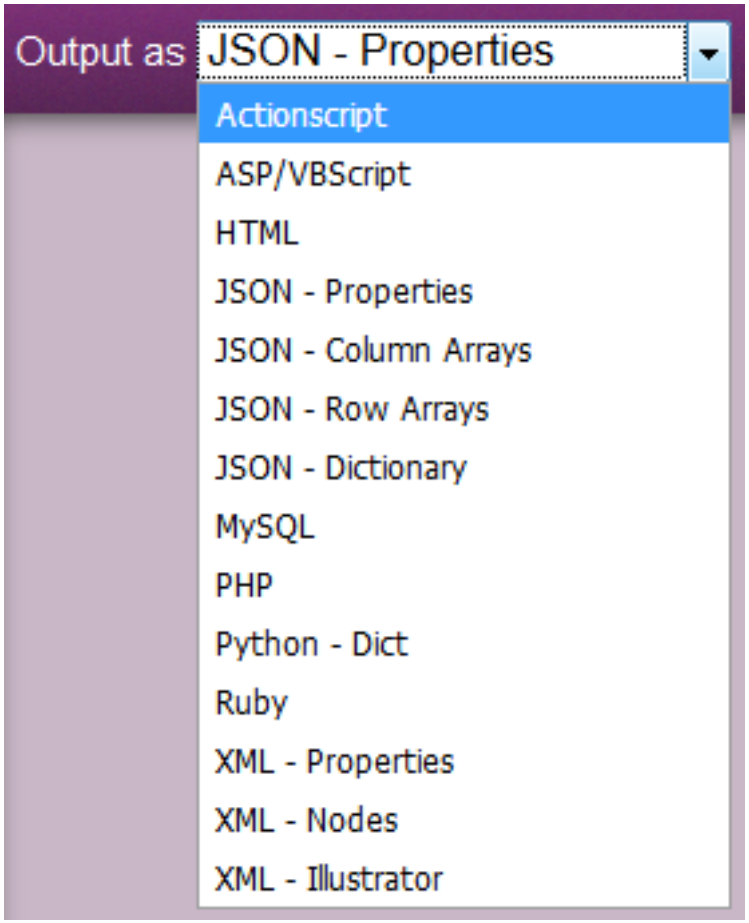
Mr. Data Converter

1. Panel de trabajo



Mr. Data Converter

1. Panel de trabajo



Mr. Data Converter

1. Panel de trabajo

ActionScript: lenguaje de programación de la plataforma Adobe Flash. Desarrollado como una forma para que los desarrolladores programen de forma más interactiva. La programación permite mucha más eficiencia en las aplicaciones de la plataforma Flash para construir animaciones de todo tipo, desde simples a complejas, ricas en datos e interfaces interactivas.

ASP/VBScript (abreviatura de ***Visual Basic Script Edition***): lenguaje interpretado por el Windows Scripting Host de Microsoft. Su sintaxis refleja su origen como variación del lenguaje de programación Visual Basic. El crecimiento del uso de las tecnologías de Internet ha supuesto un significativo avance para este lenguaje, dado que es parte fundamental de la ejecución de aplicaciones de servidor programadas en ASP (*Active Server Pages*).

HTML (***HyperText Markup Language***): lenguaje de marcado para la elaboración de páginas web. Es un estándar que sirve de referencia del software que conecta con la elaboración de páginas web en sus diferentes versiones, define una estructura básica y un código (denominado código HTML) para la definición de contenido de una página web, como texto, imágenes, videos, juegos, entre otros. Está a cargo del *World Wide Web Consortium* (W3C).

JSON (***JavaScript Object Notation***): formato de texto ligero para el intercambio de datos. JSON es un subconjunto de la notación literal de objetos de JavaScript aunque hoy, debido a su amplia adopción como alternativa a XML, se considera un formato de lenguaje independiente. Formato muy extendido para el intercambio de datos.

Fuente: Wikipedia

Mr. Data Converter

1. Panel de trabajo

MySQL: sistema de gestión de bases de datos relacional desarrollado bajo licencia dual (licencia pública general / licencia comercial por Oracle Corporation, Considerada la base datos de código abierto más popular del mundo y una de las más populares en general sobre todo para entornos de desarrollo web.

PHP (***Hypertext Preprocessor***): un lenguaje de programación de propósito general de código del lado del servidor originalmente diseñado para el desarrollo web de contenido dinámico. Fue uno de los primeros lenguajes de programación del lado del servidor que se podían incorporar directamente en un documento HTML en lugar de llamar a un archivo externo que procese los datos. El código es interpretado por un servidor web con un módulo de procesador de PHP que genera el HTML resultante

Phyton: lenguaje de programación interpretado cuya filosofía hace hincapié en una sintaxis que favorezca un código legible. A dictionary is a collection which is unordered, changeable and indexed. In Python dictionaries are written with curly brackets, and they have keys and values.

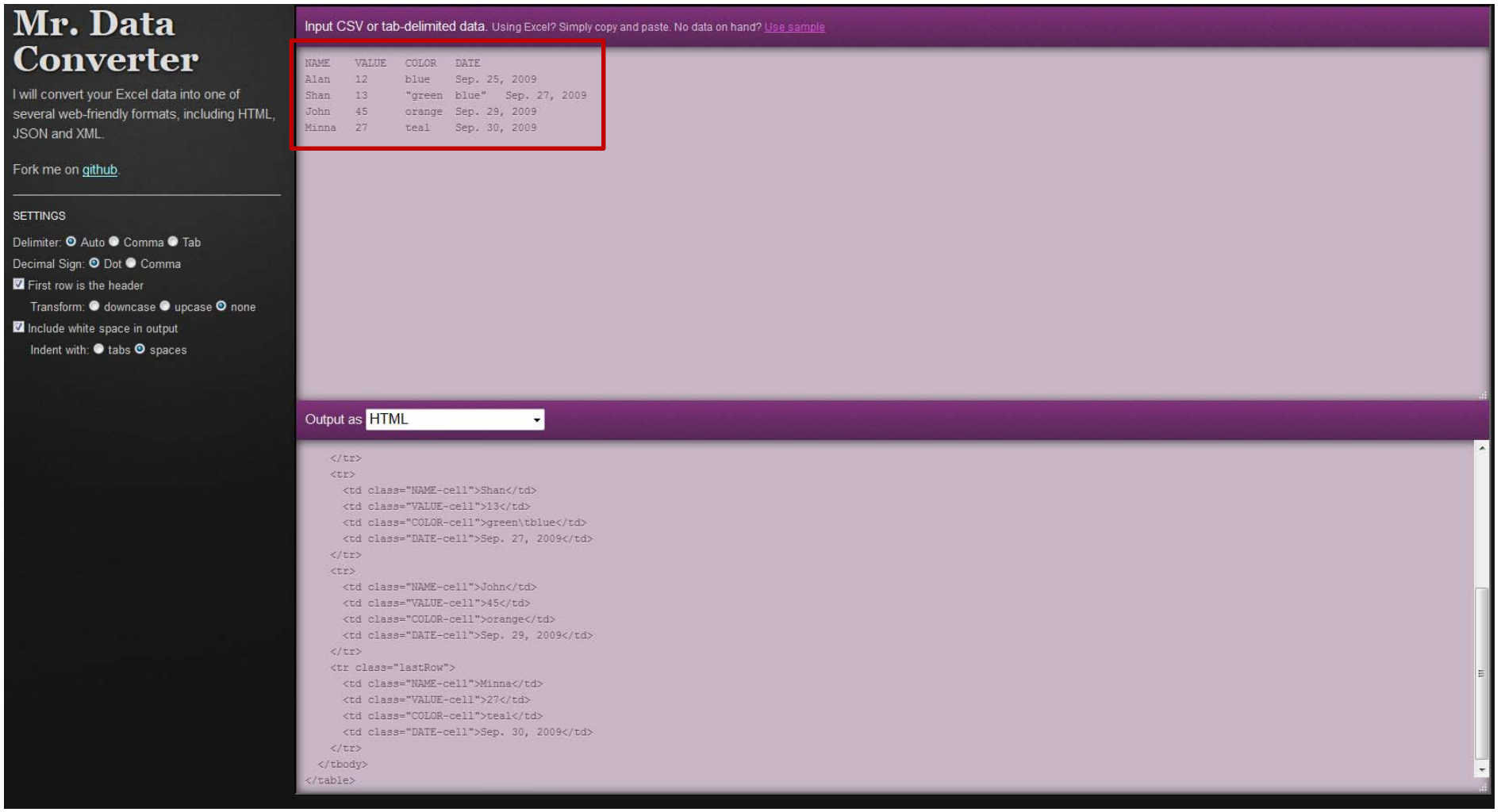
Ruby: lenguaje de programación interpretado, reflexivo y orientado a objetos.

XML (***eXtensible Markup Language***)*:* meta-lenguaje que permite definir lenguajes de marcas desarrollado por W3C y utilizado para almacenar datos en forma legible.

Fuente: Wikipedia

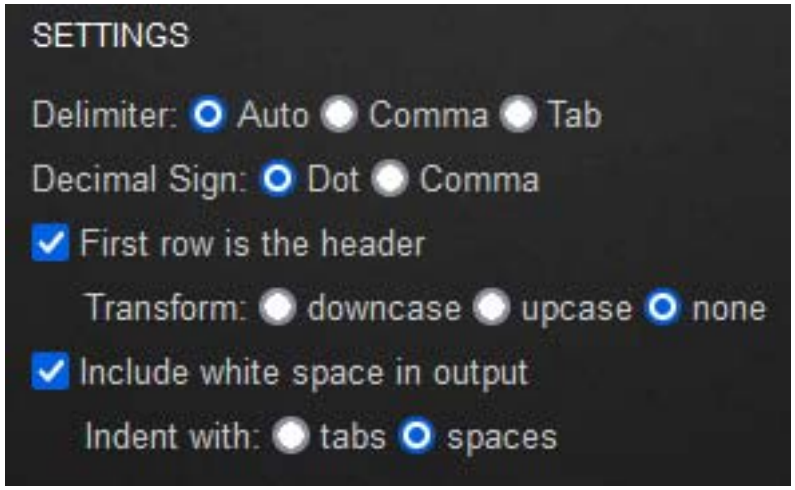
Mr. Data Converter

2. Ejemplos de uso



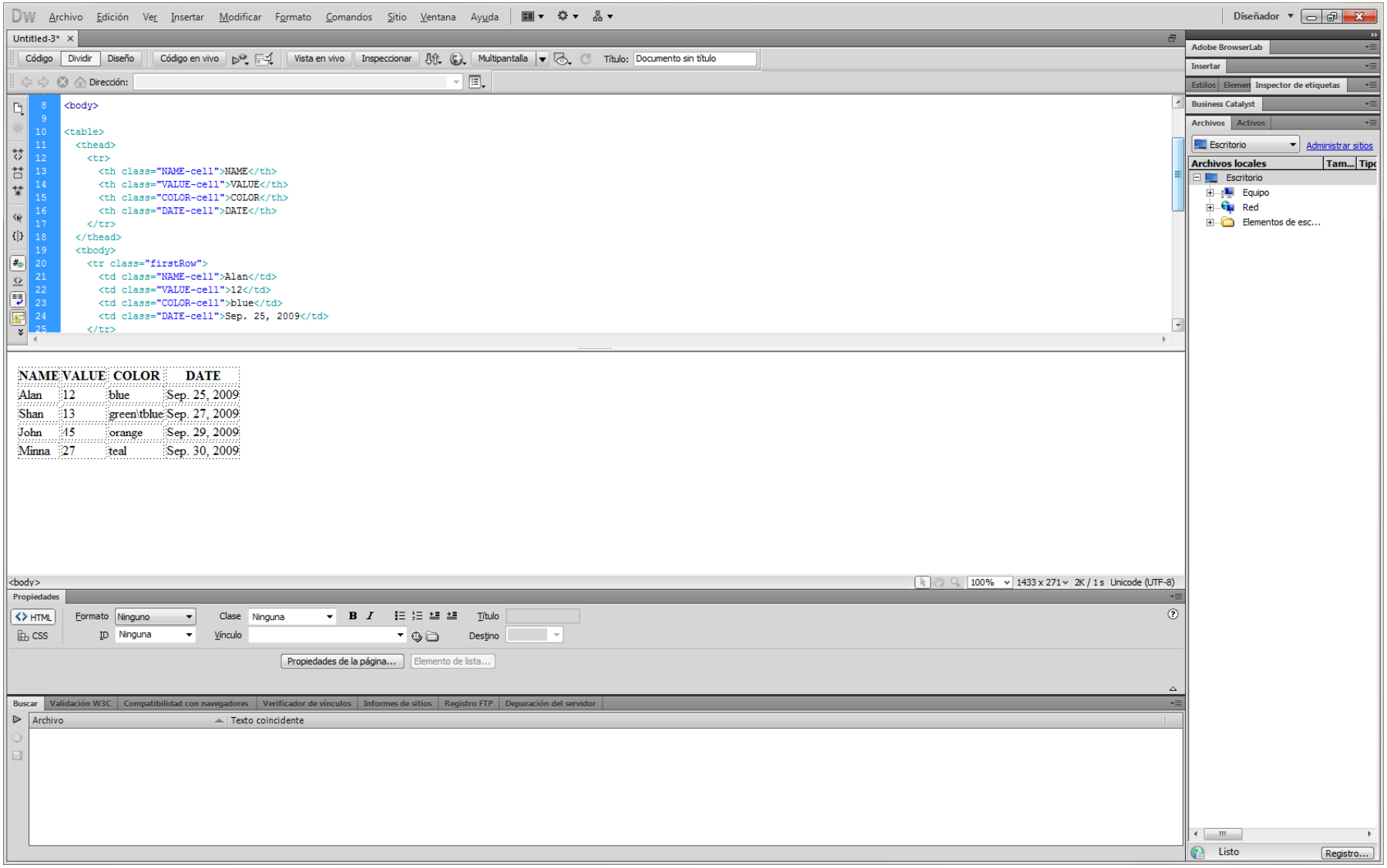
Mr. Data Converter

2. Ejemplos de uso



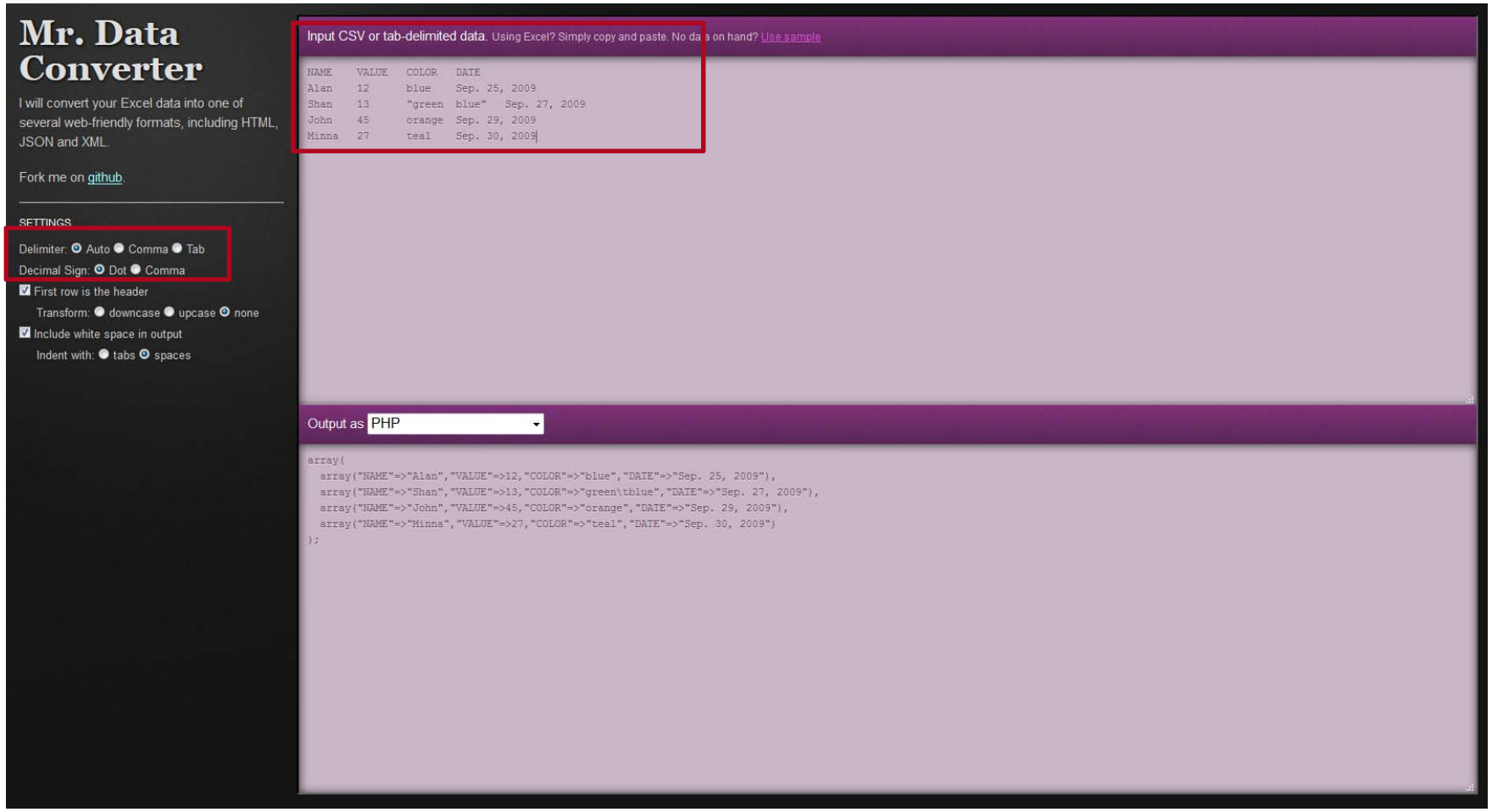
Mr. Data Converter

2. Ejemplos de uso



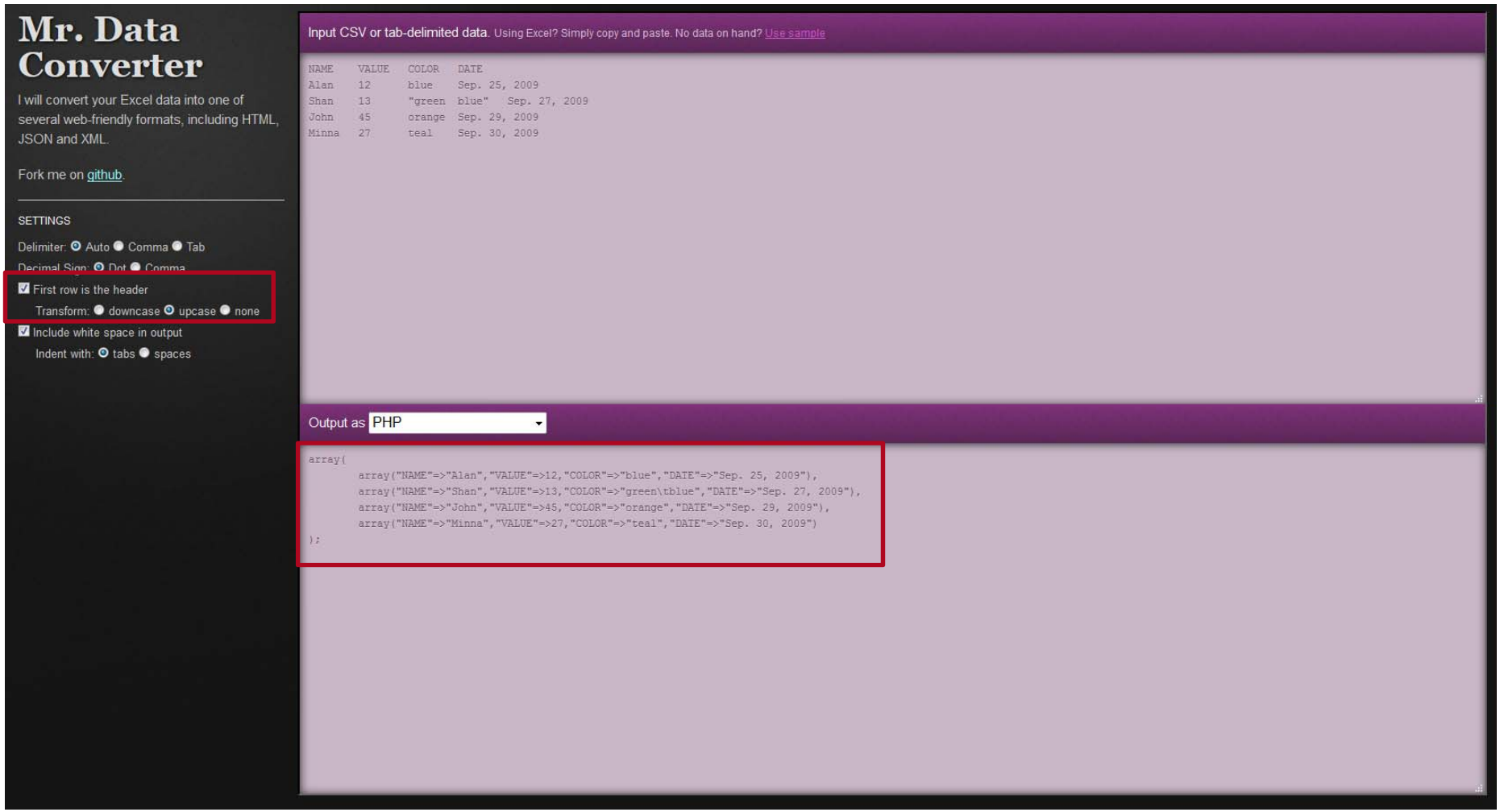
Mr. Data Converter

2. Ejemplos de uso



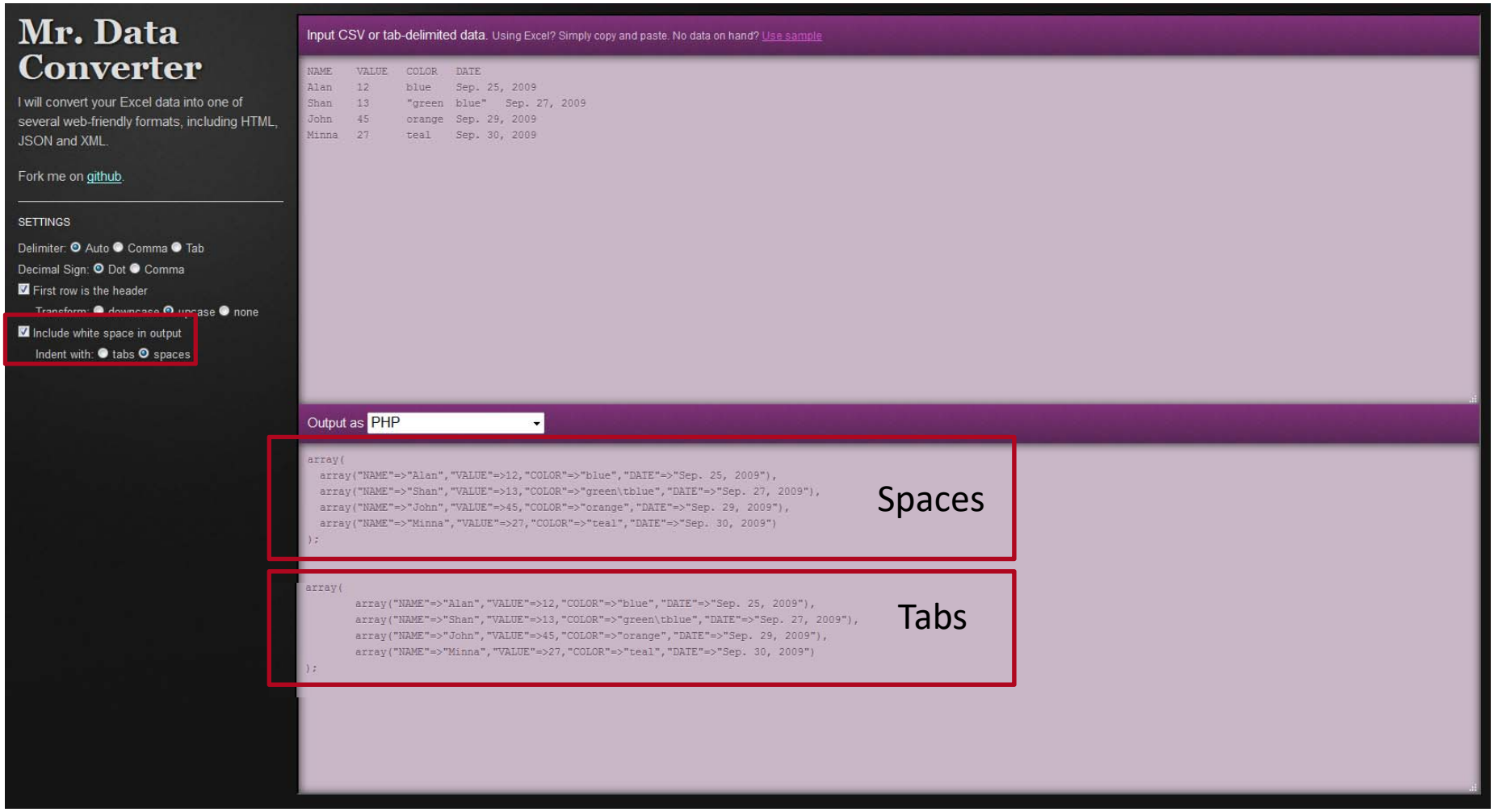
Mr. Data Converter

2. Ejemplos de uso



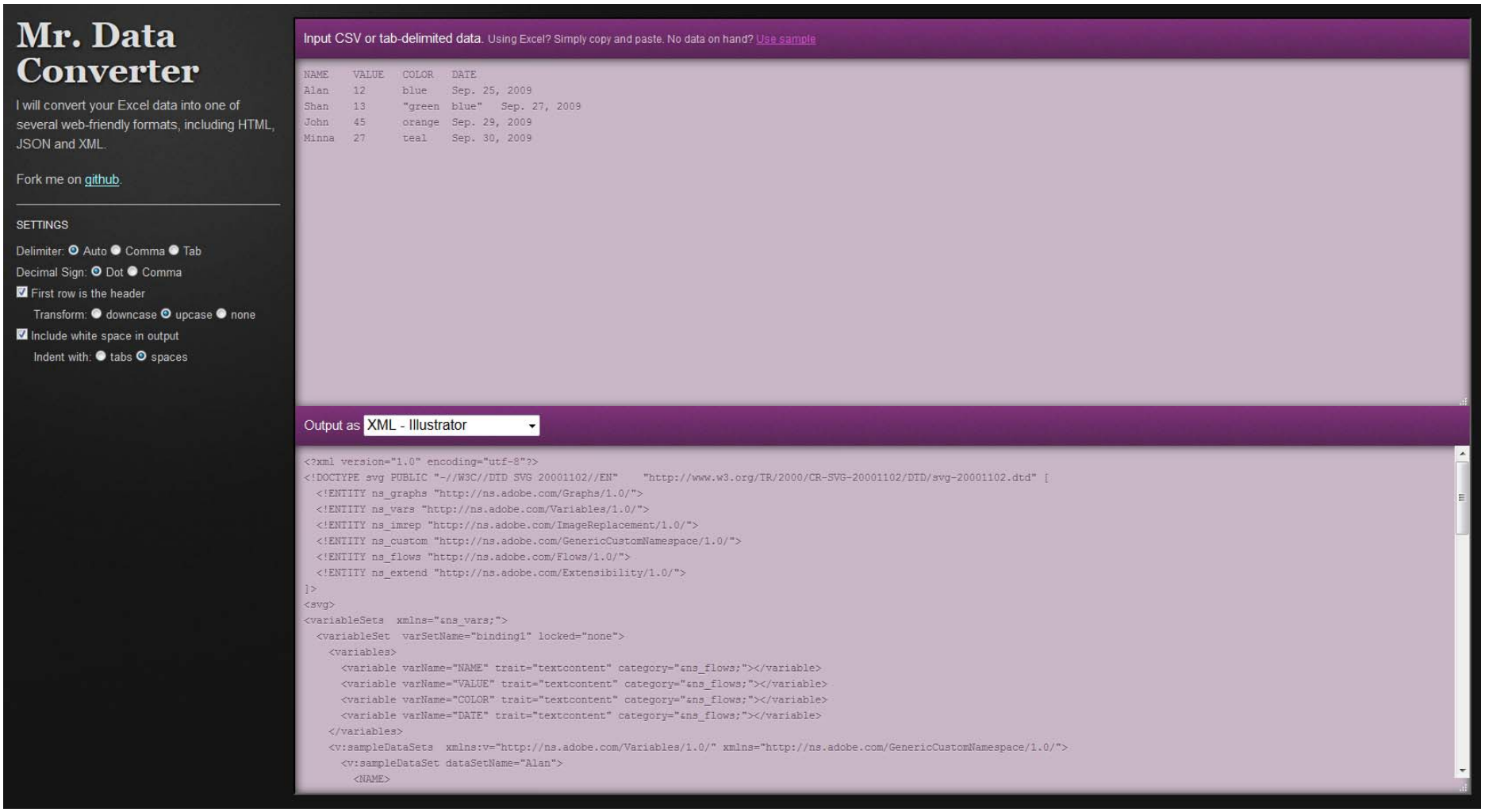
Mr. Data Converter

2. Ejemplos de uso



Mr. Data Converter

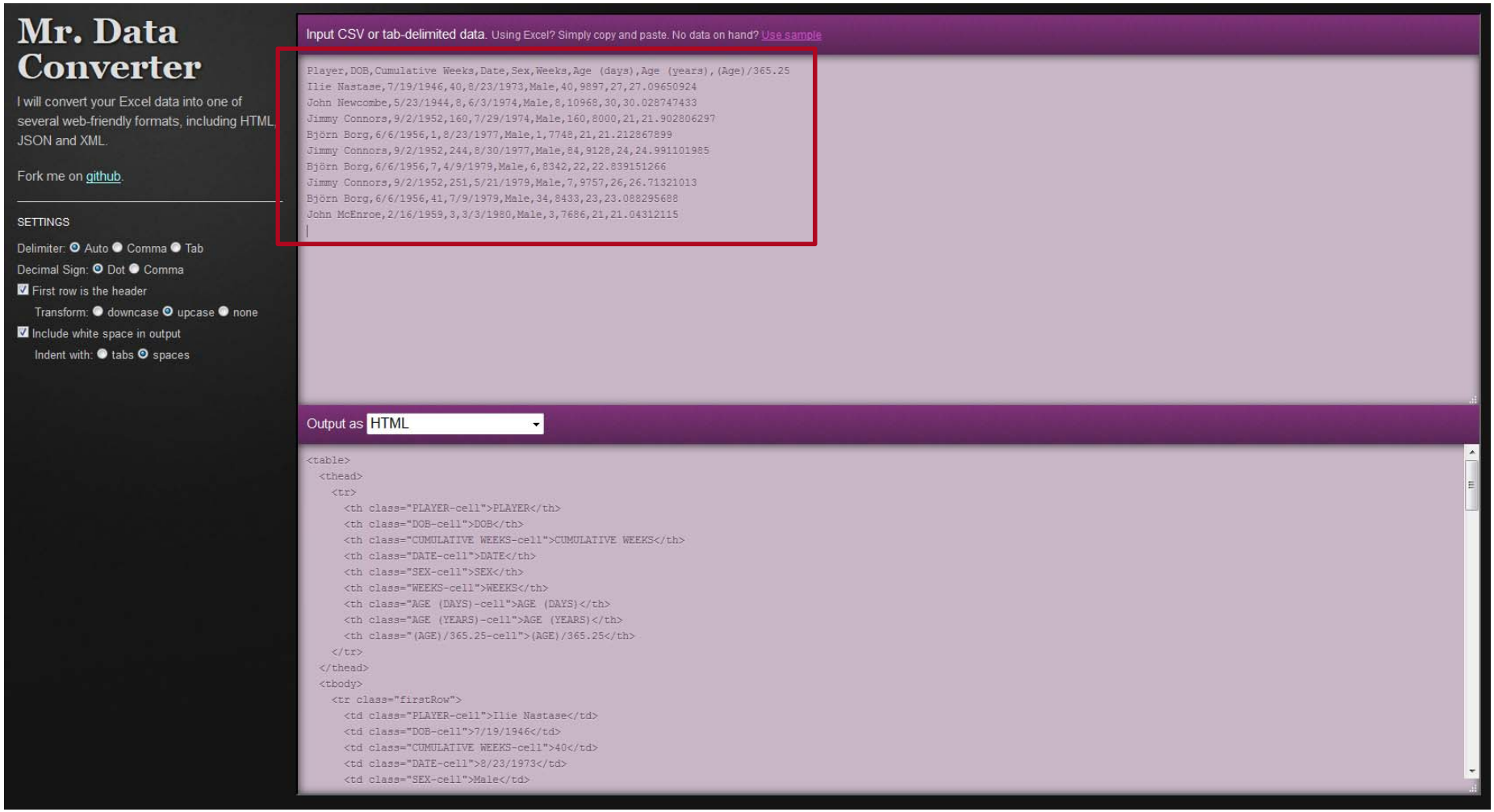
2. Ejemplos de uso



DIGITAL.CSIC

Mr. Data Converter

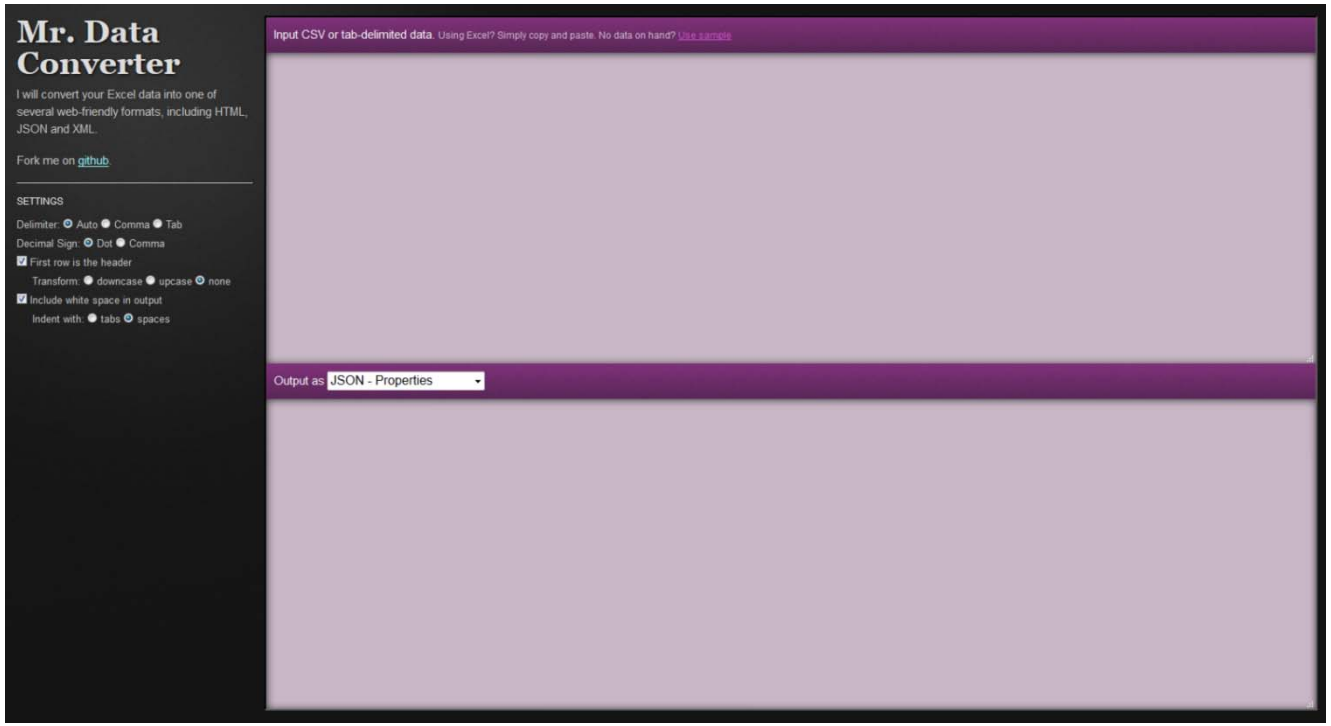
2. Ejemplos de uso



DIGITAL.CSIC

PRÁCTICAS

Mr. Data Converter



https://shancarter.github.io/mr-data-converter/

Mr. Data Converter

airbnb.xlsx
significantvolcanoeruptions.xlsx
atp_number_1.csv
wimbledons_champions.csv

1. Trabaja con los ficheros xlsx aportados para el curso
2. Modifica el apartado de ajustes (*settings*)

3. Convierte a distintos formatos con los ficheros csv
4. Lleva a cabo cambios en el apartado de ajustes (*settings*)

Mr. Data Converter

5. Lleva a cabo algunas conversiones a HTML escribiendo a mano los valores (separados por comas)

Pos.	País	PIB nominal (millones de USD)
	Mundo	**93 889 577**
1	Estados Unidos	22 939 580
	Unión Europea	17 127 535
2	China	16 862 979
3	Japón	5 103 110
4	Alemania	4 230 172
5	Reino Unido	3 108 416
6	India	2 946 061
7	Francia	2 940 428
8	Italia	2 120 232
9	Canadá	2 015 983
10	Corea del Sur	1 823 852
11	Rusia	1 647 568
12	Brasil	1 645 837
13	Australia	1 610 556
14	España	1 439 958
15	México	1 285 518
16	Indonesia	1 150 245
17	Irán	1 081 383
18	Países Bajos	1 007 562
19	Arabia Saudita	842 588
20	Suiza	810 830

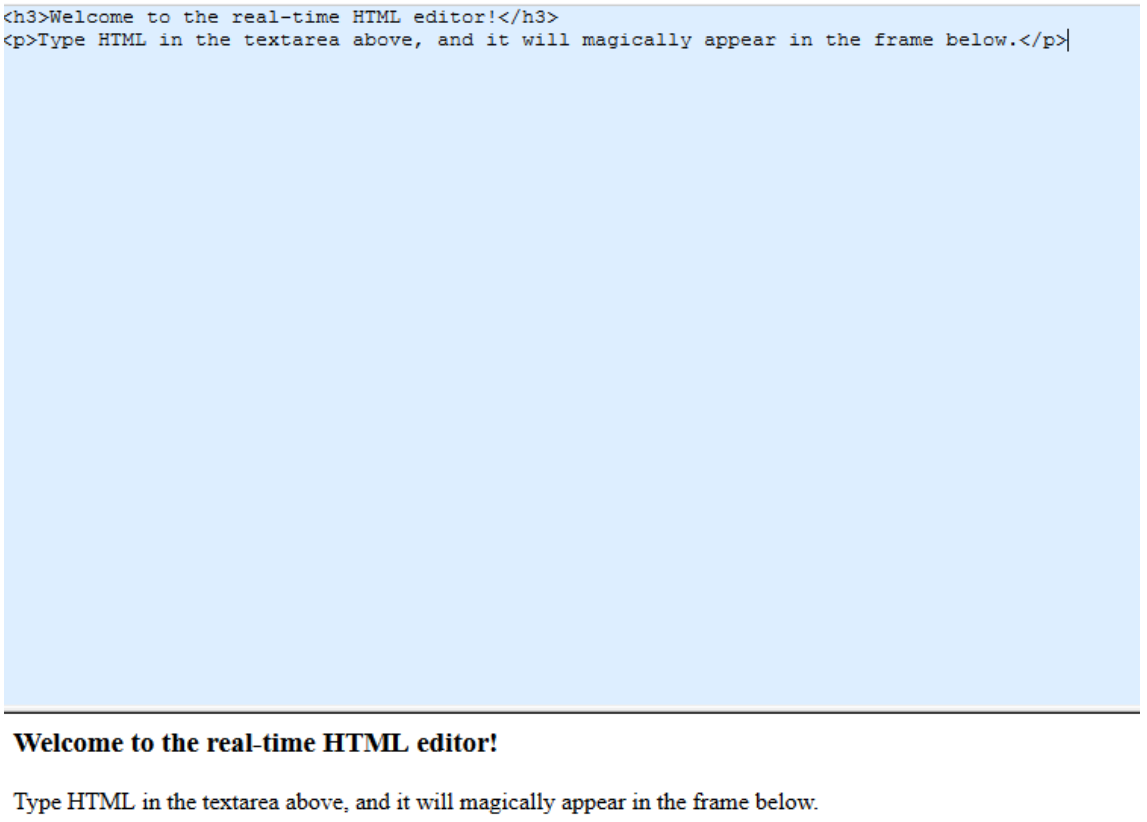
TOP 20 de países por producto interno bruto (PIB) según el Fondo Monetario Internacional (octubre 2021). Fuente Wikipedia

Mr. Data Converter

Si la conversión la haces a HTML puedes ver el resultado en un editor de HTML:

Real-time HTML Editor

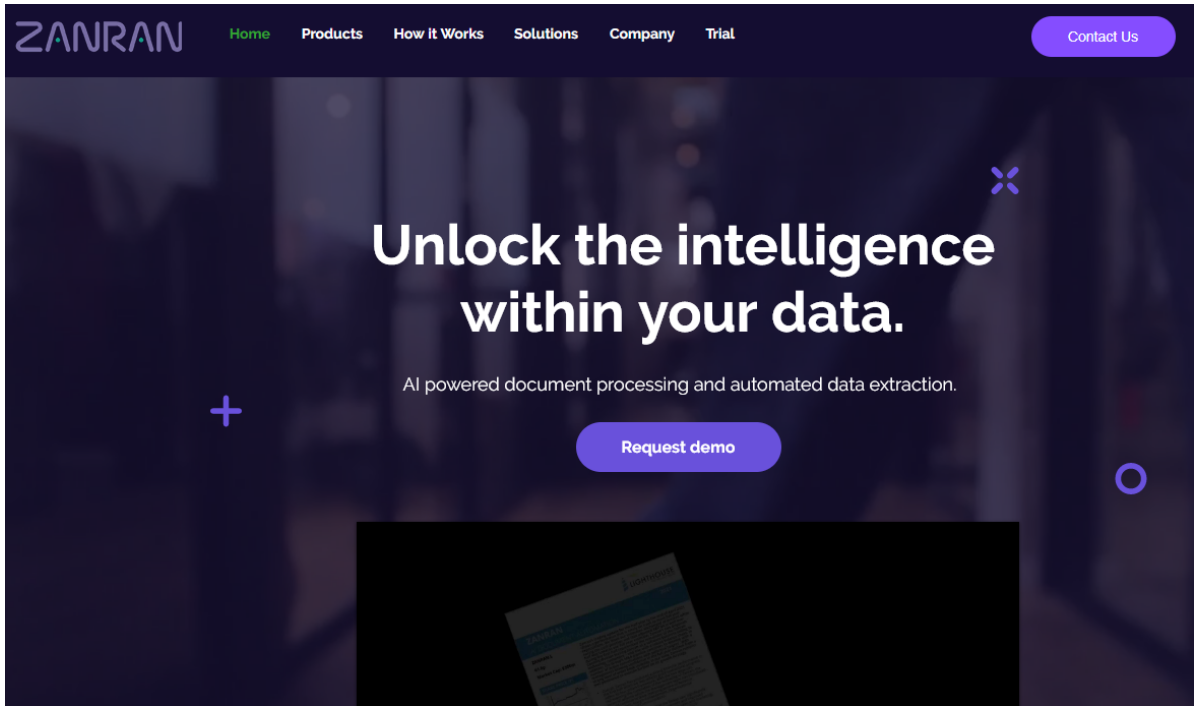
https://htmledit.squarefree.com/



DIGITAL.CSIC

Zanran

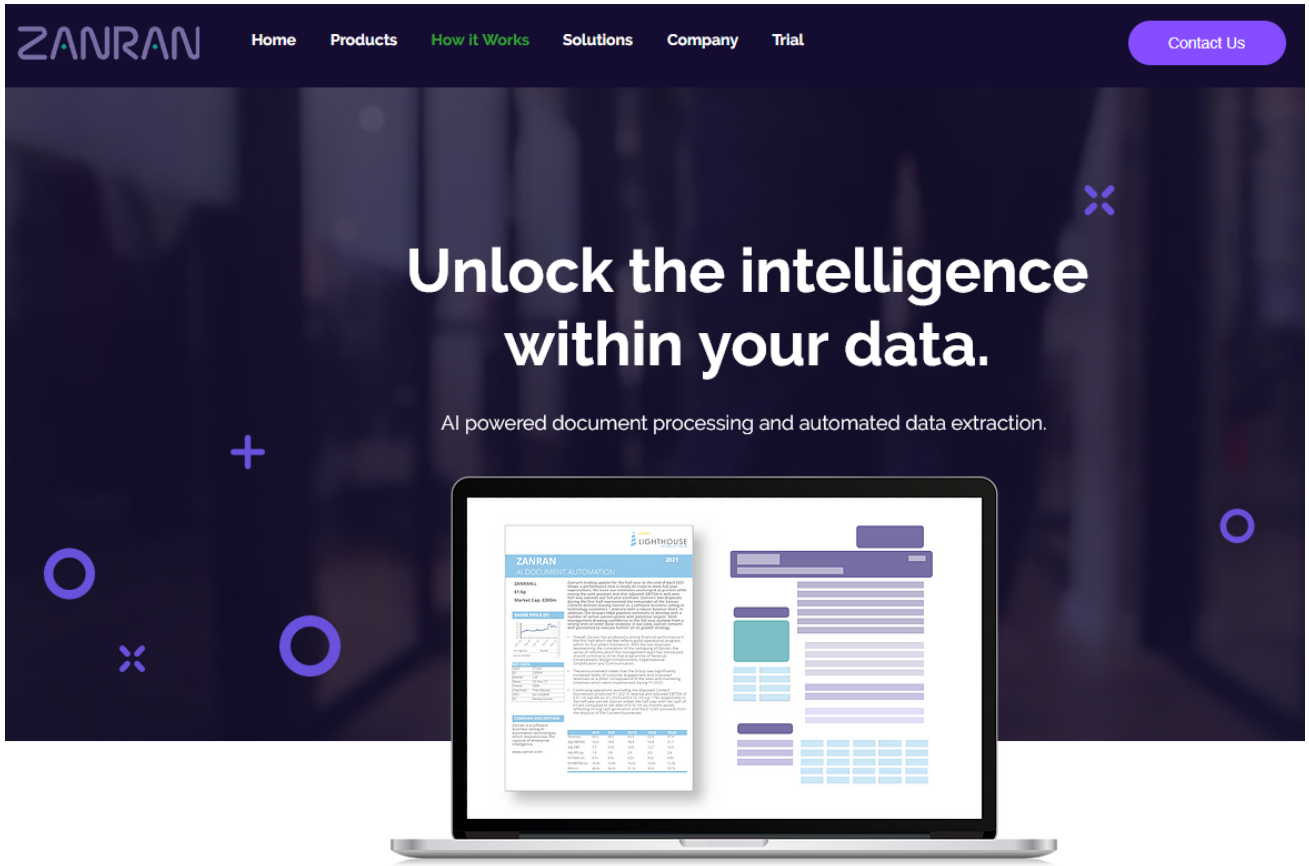
1. Panel de trabajo / guías
2. Subir fichero
3. Resultados
4. Descargar resultados
5. Ver resultados



https://www.zanran.com/

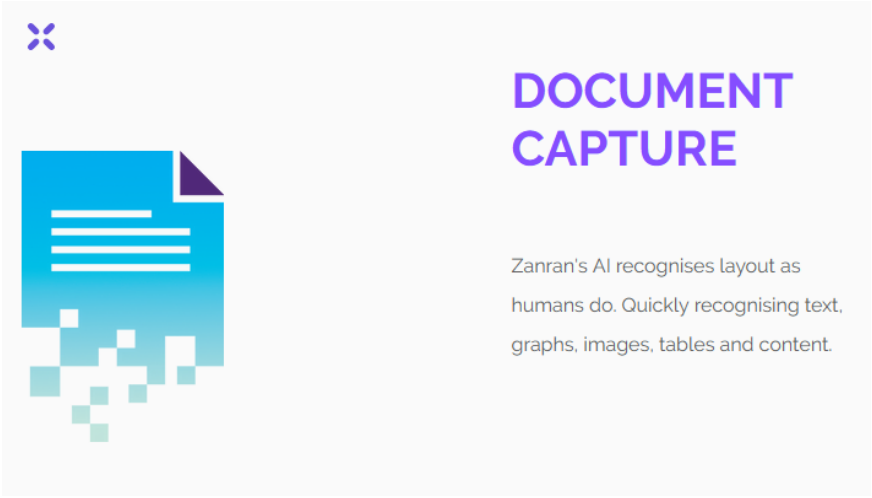
Zanran

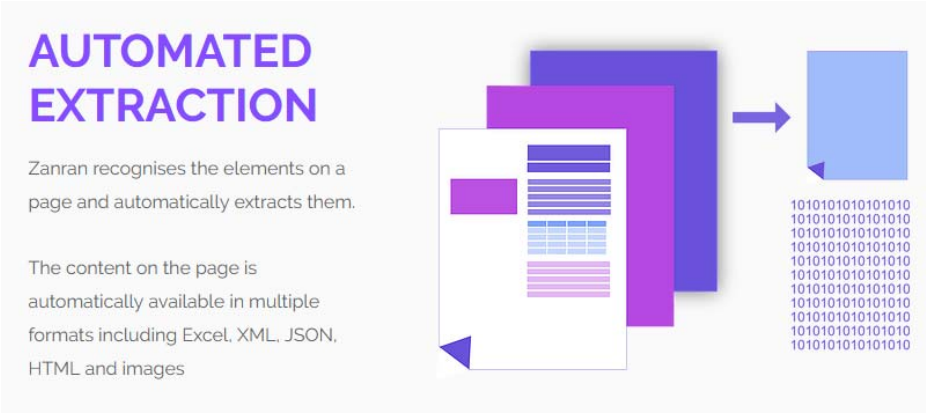
1. Panel de trabajo / guías

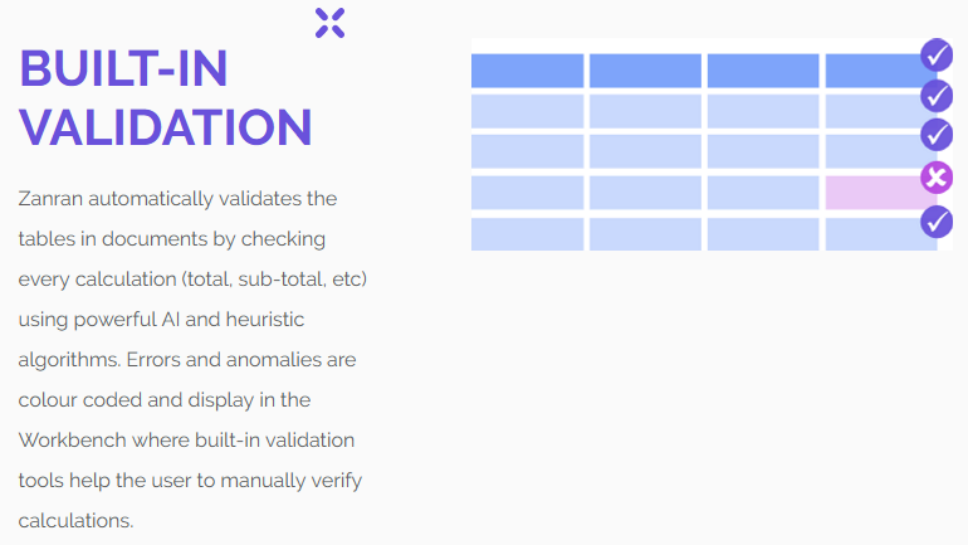


Zanran

1. Panel de trabajo / guías



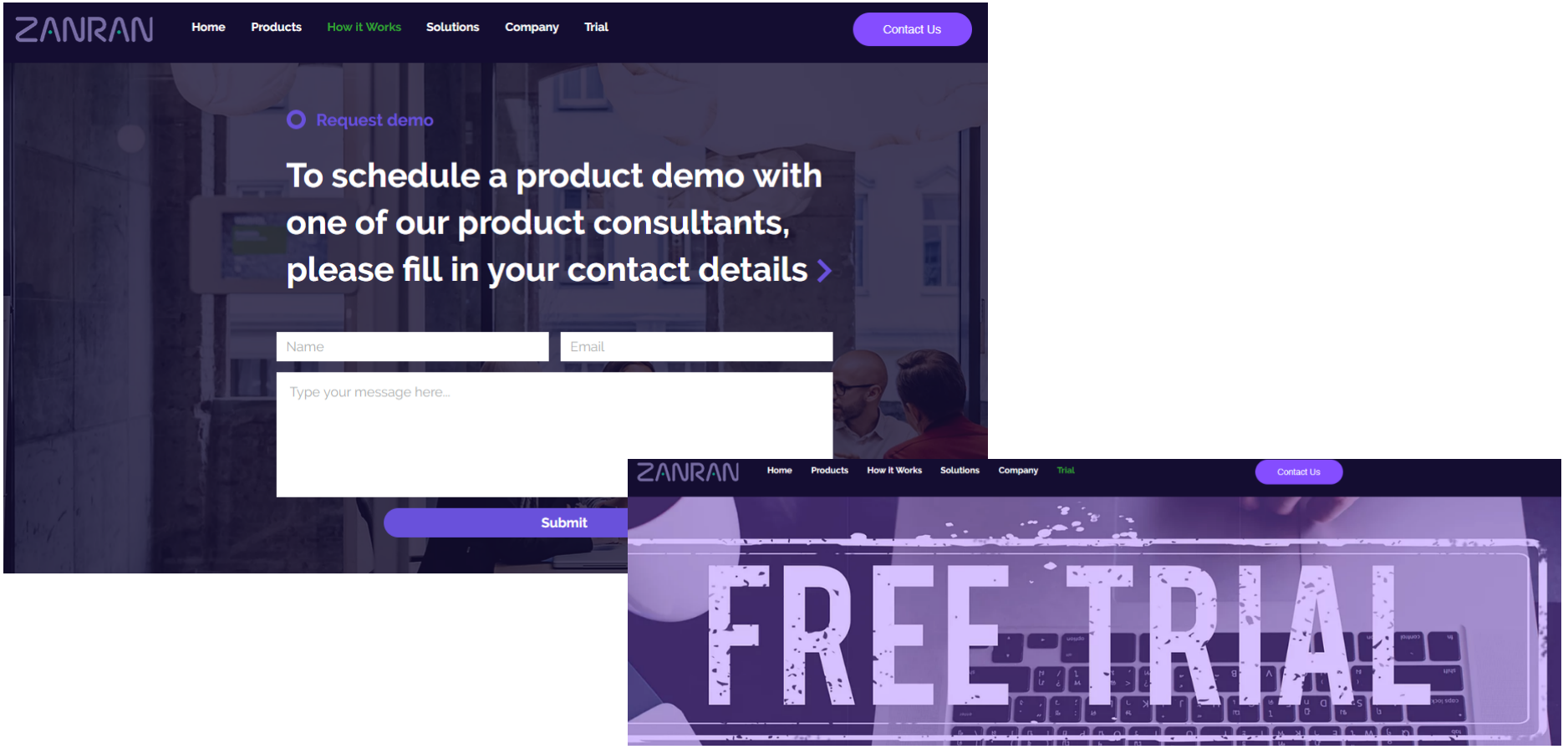




DIGITAL.CSIC

Zanran

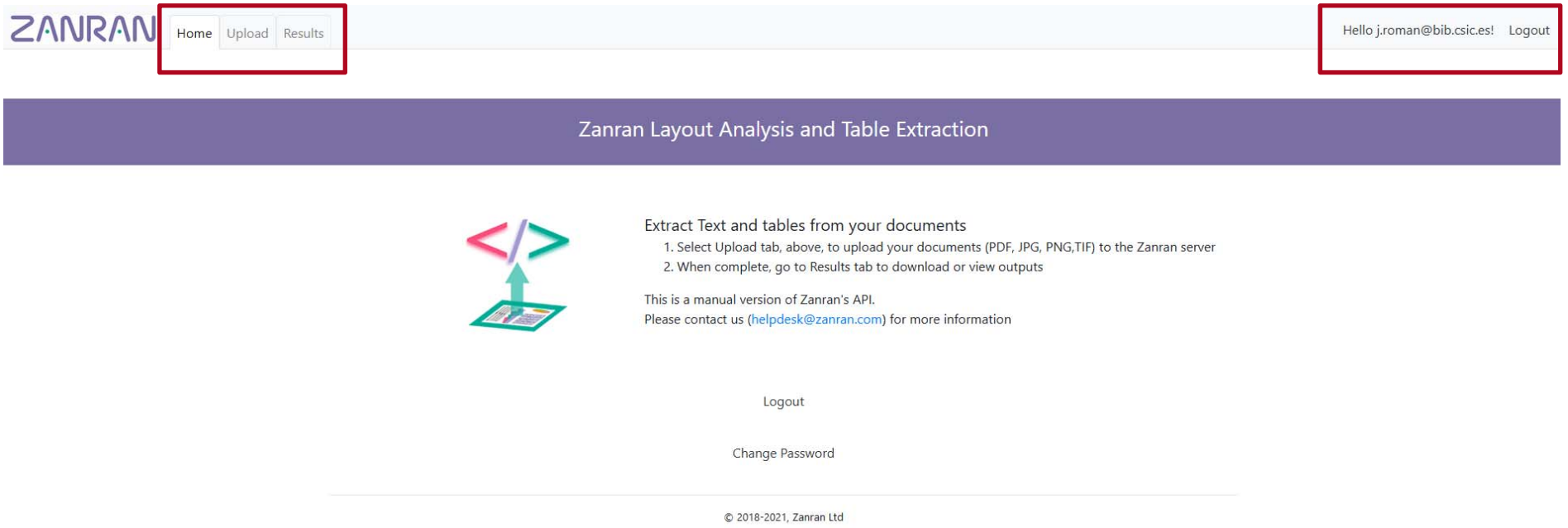
1. Panel de trabajo / guías



DIGITAL.CSIC

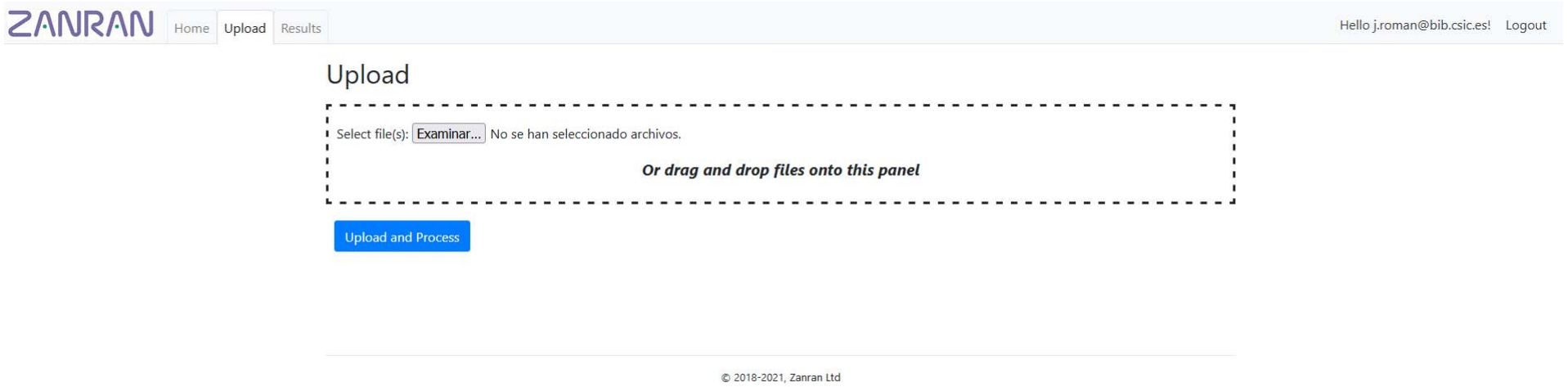
Zanran

1. Panel de trabajo / guías



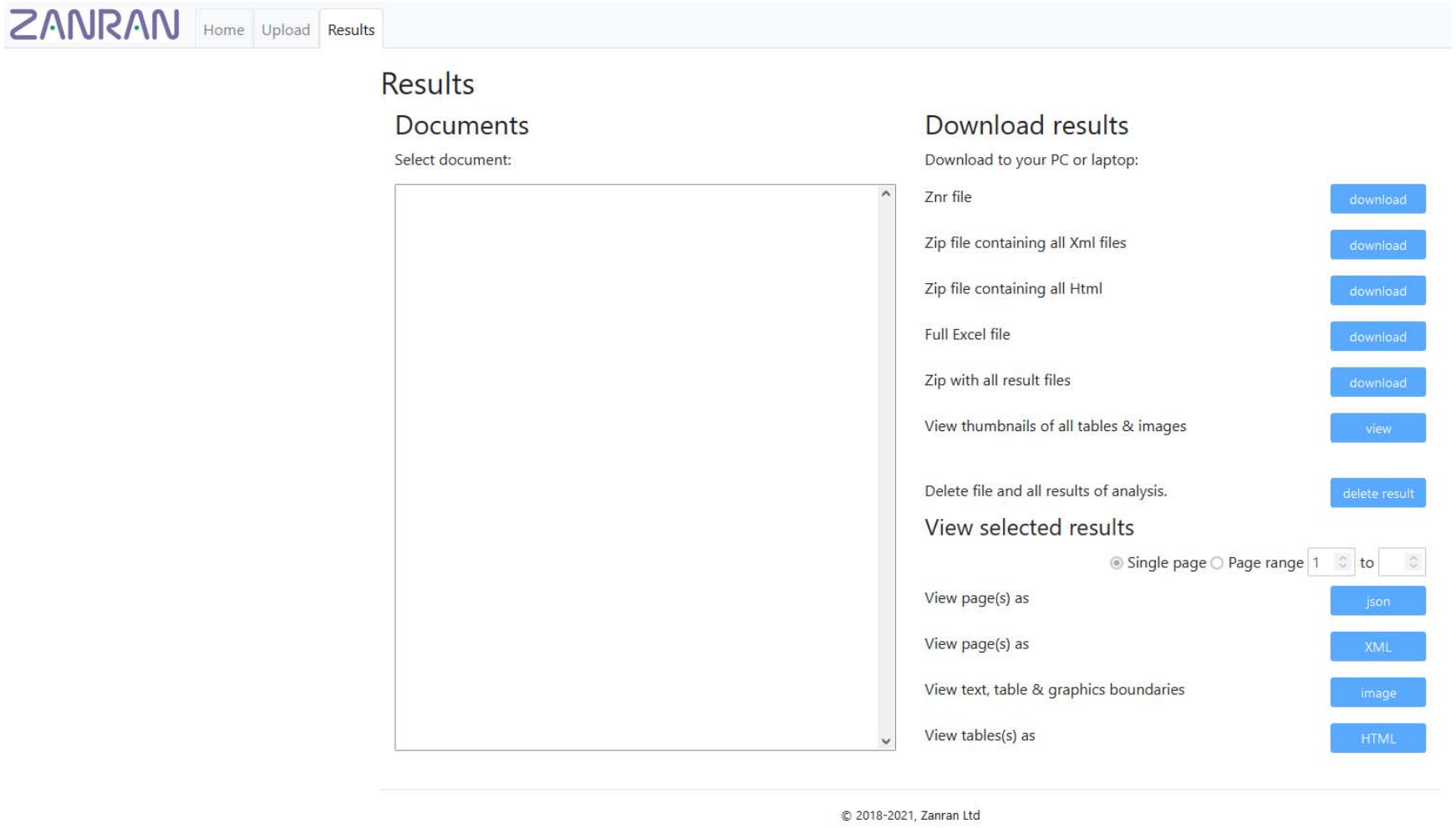
Zanran

1. Panel de trabajo / guías



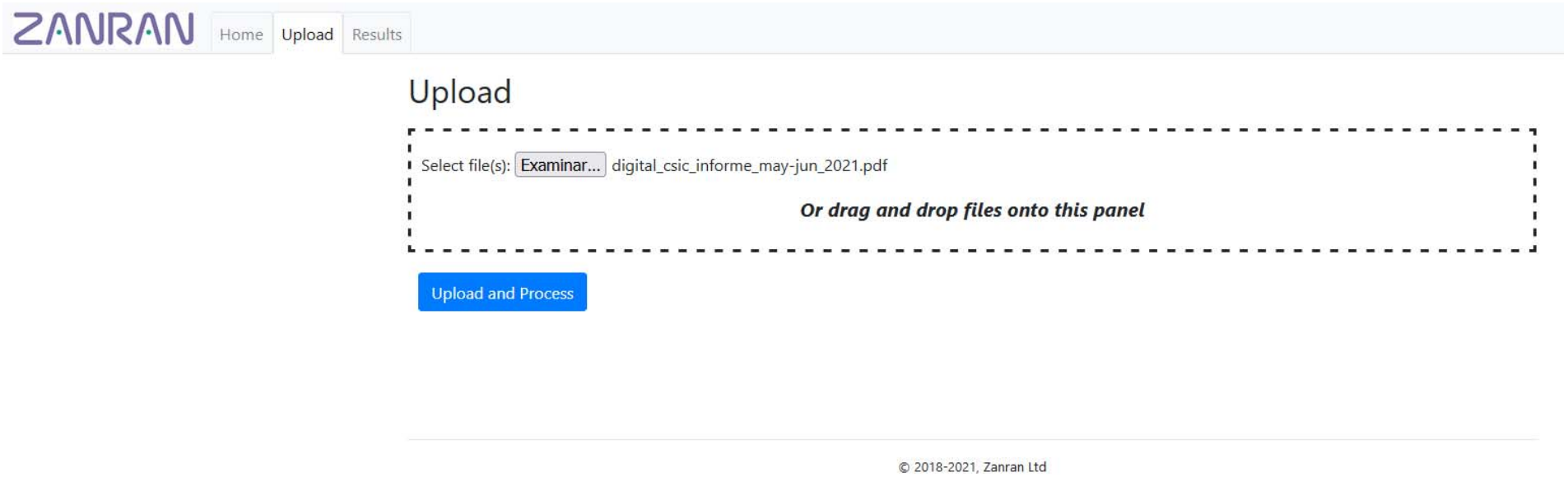
Zanran

1. Panel de trabajo / guías



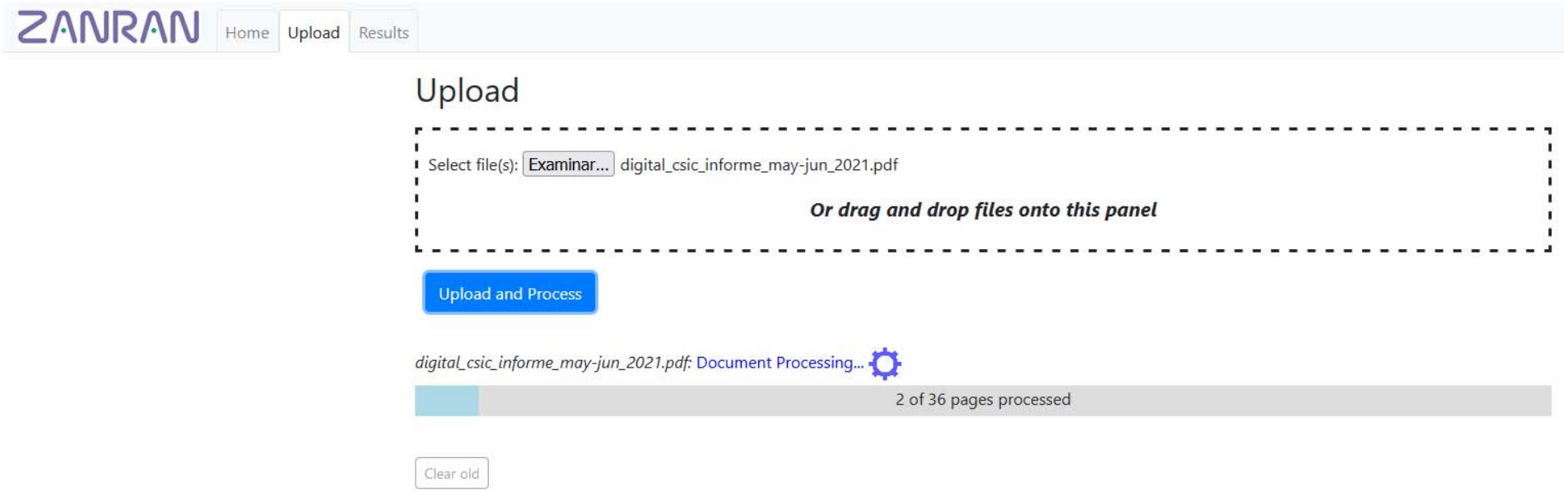
Zanran

2. Subir fichero/s



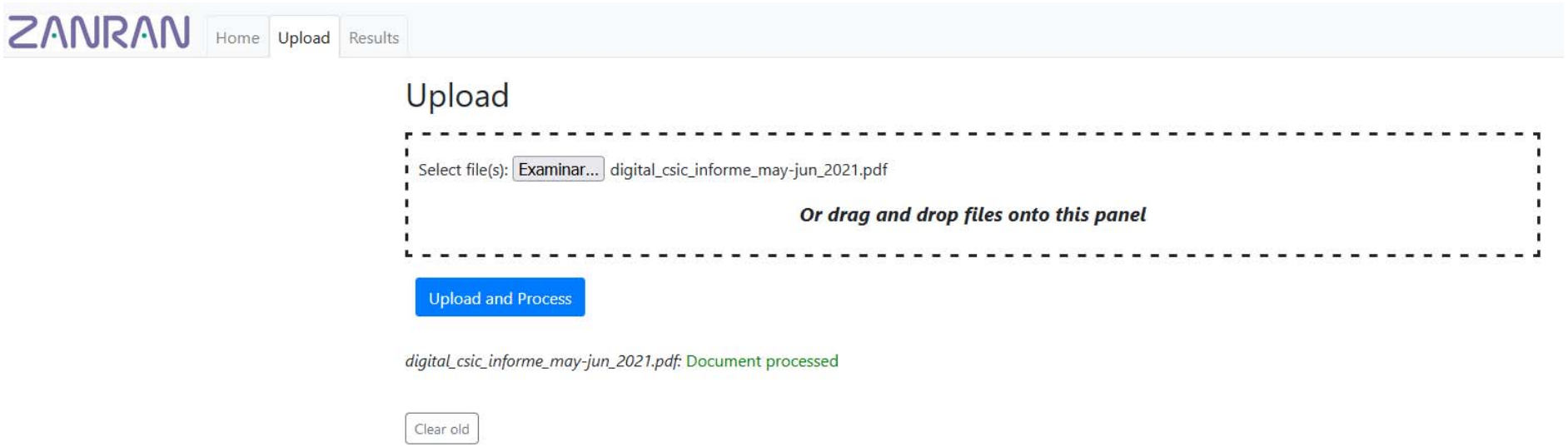
Zanran

2. Subir fichero/s



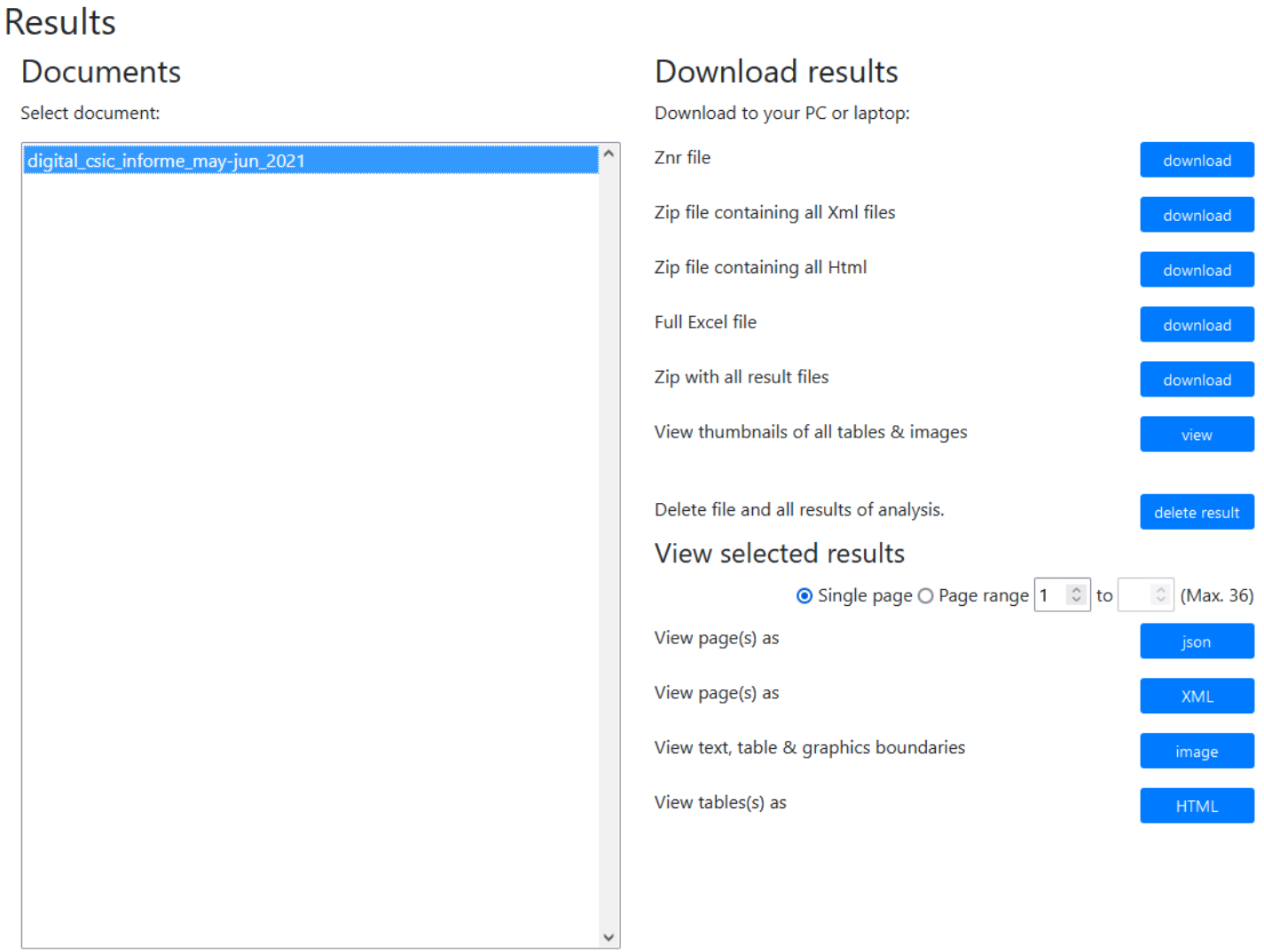
Zanran

2. Subir fichero/s



Zanran

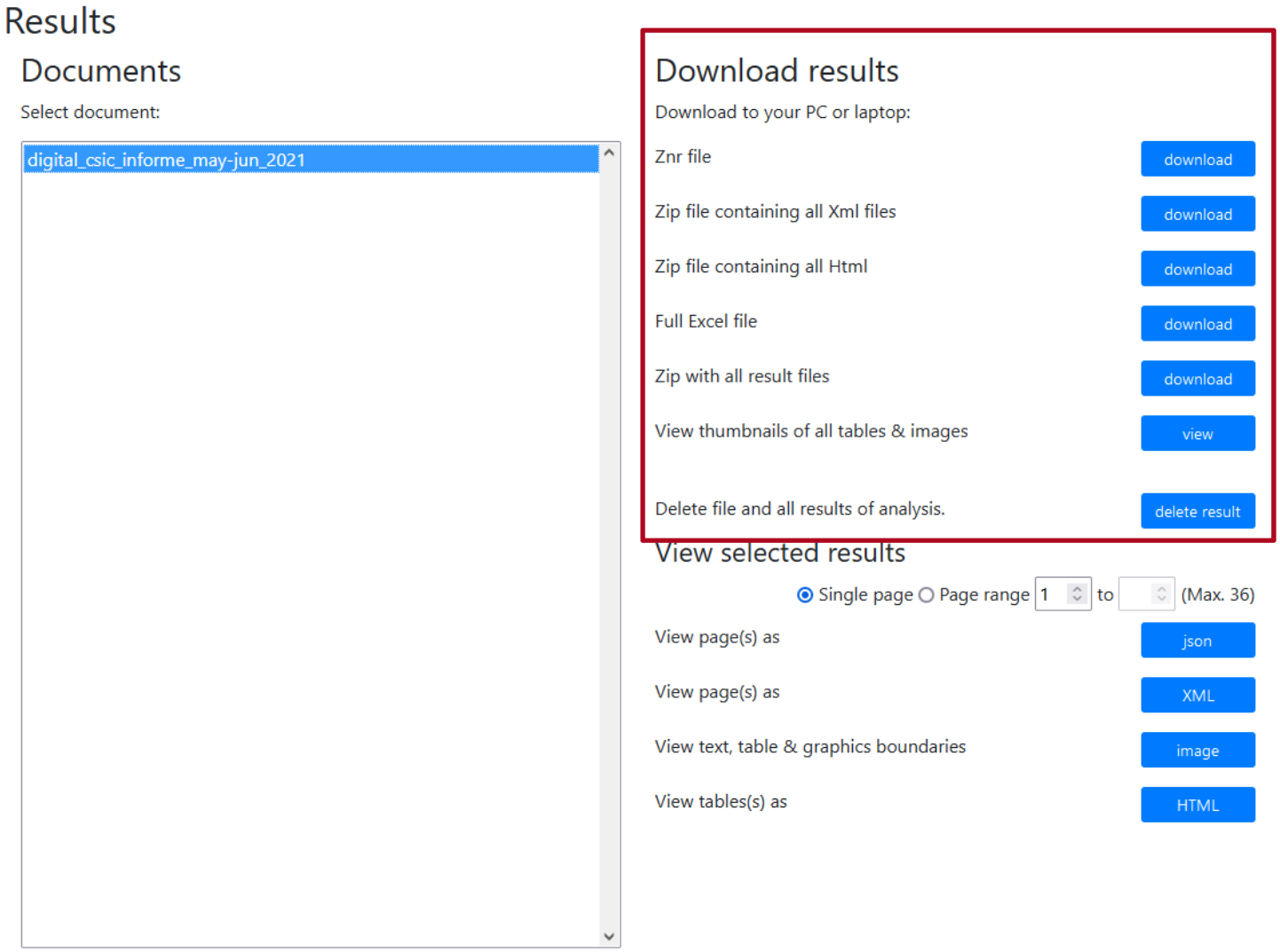
3. Resultados



DIGITAL.CSIC

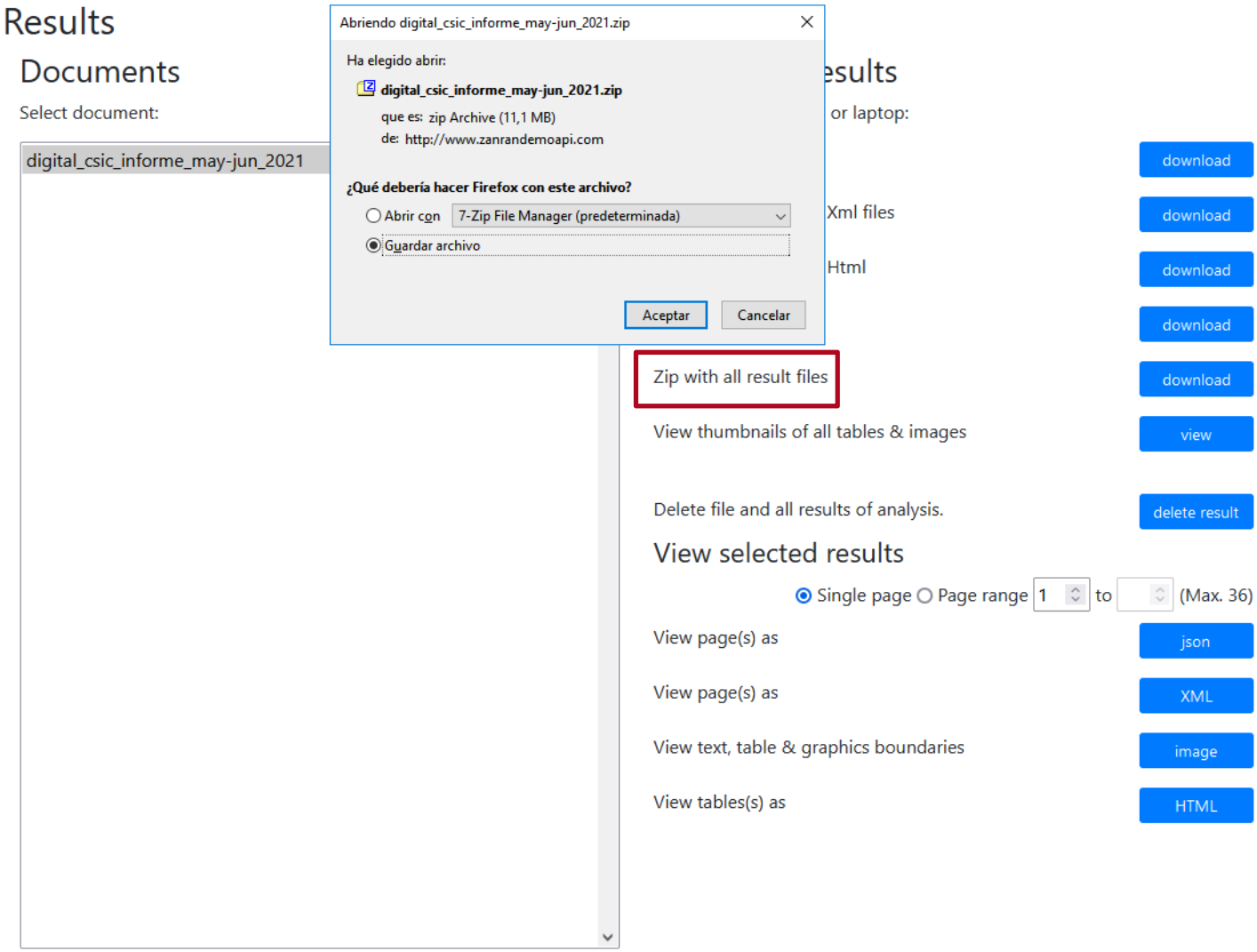
Zanran

4. Descargar resultados



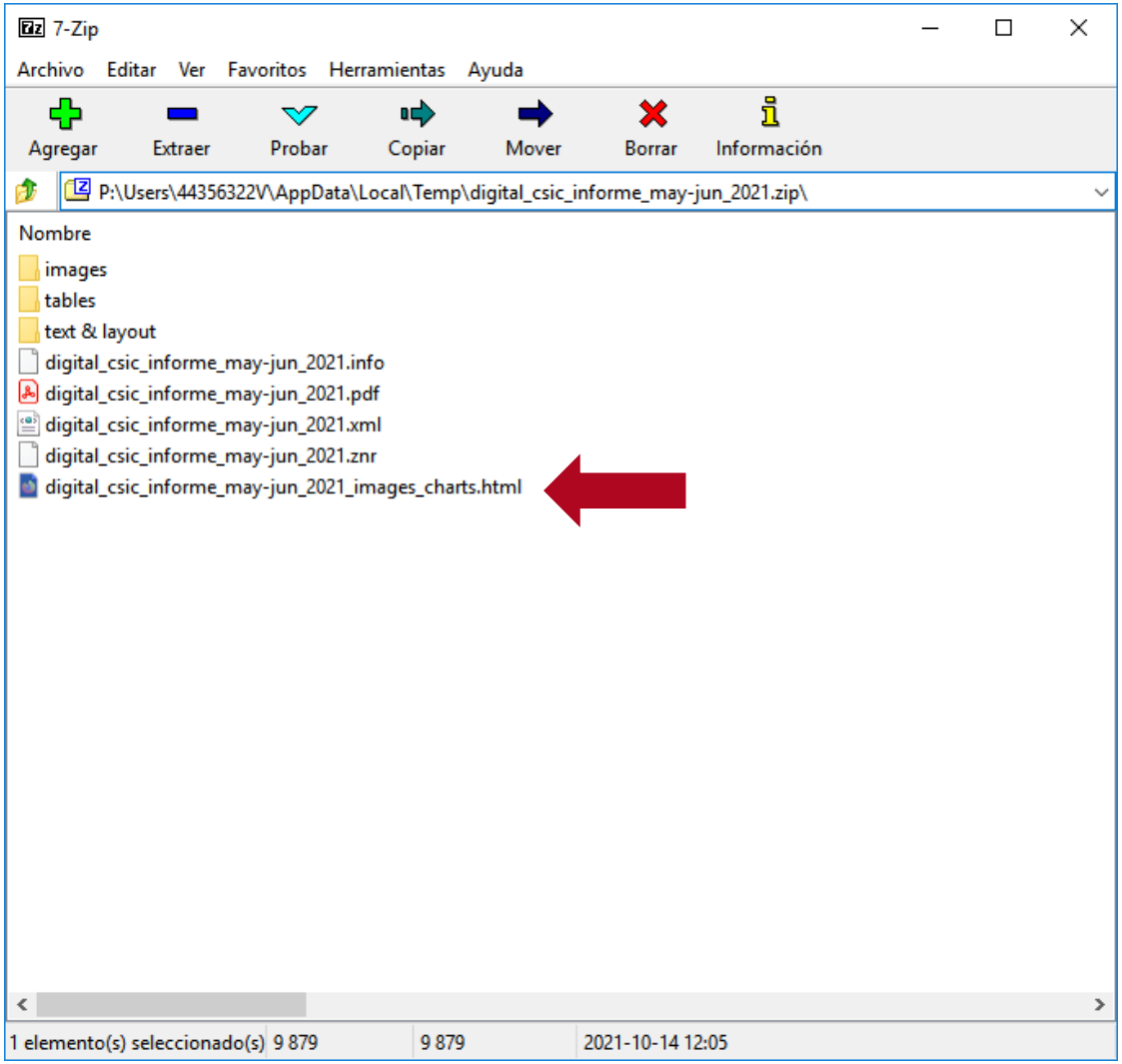
Zanran

4. Descargar resultados



Zanran

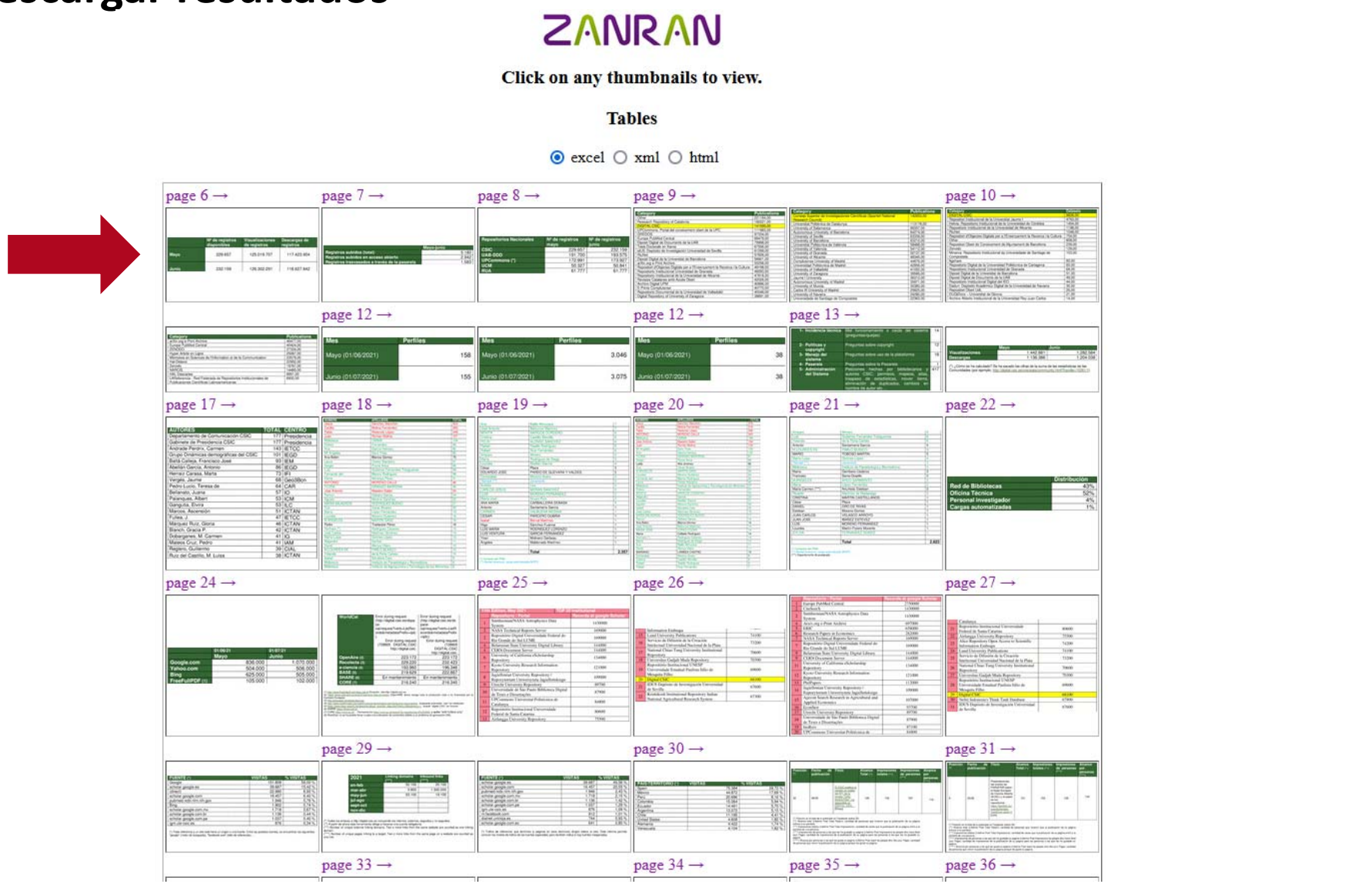
4. Descargar resultados



DIGITAL.CSIC

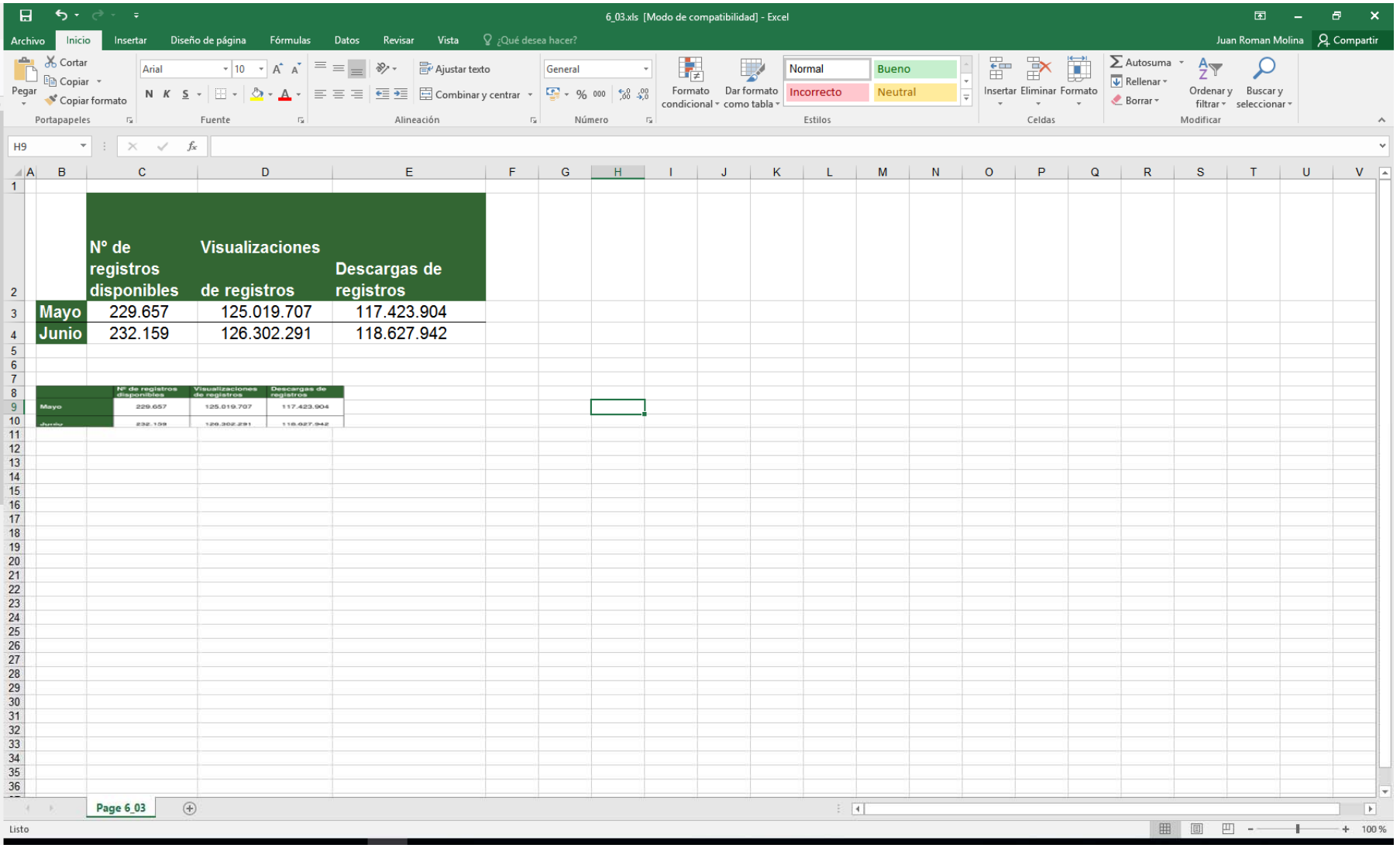
Zanran

4. Descargar resultados



Zanran

4. Descargar resultados



	Nº de registros disponibles	Visualizaciones de registros	Descargas de registros
Mayo	229.657	125.019.707	117.423.904
Junio	232.159	126.302.291	118.627.942

	Nº de registros disponibles	Visualizaciones de registros	Descargas de registros
Mayo	229.657	125.019.707	117.423.904
Junio	232.159	126.302.291	118.627.942

Zanran

4. Descargar resultados

digital_csic_informe_may-jun_2021 - page 6

	Nº de registros disponibles	Visualizaciones de registros	Descargas de registros
Mayo	229.657	125.019.707	117.423.904
Junio	232.159	126.302.291	118.627.942

Caution: The extraction process can occasionally result in data corruption. Please compare table data against the original image below to ensure accuracy.

	Nº de registros disponibles	Visualizaciones de registros	Descargas de registros
Mayo	229.657	125.019.707	117.423.904
Junio	232.159	126.302.291	118.627.942

Zanran

4. Descargar resultados

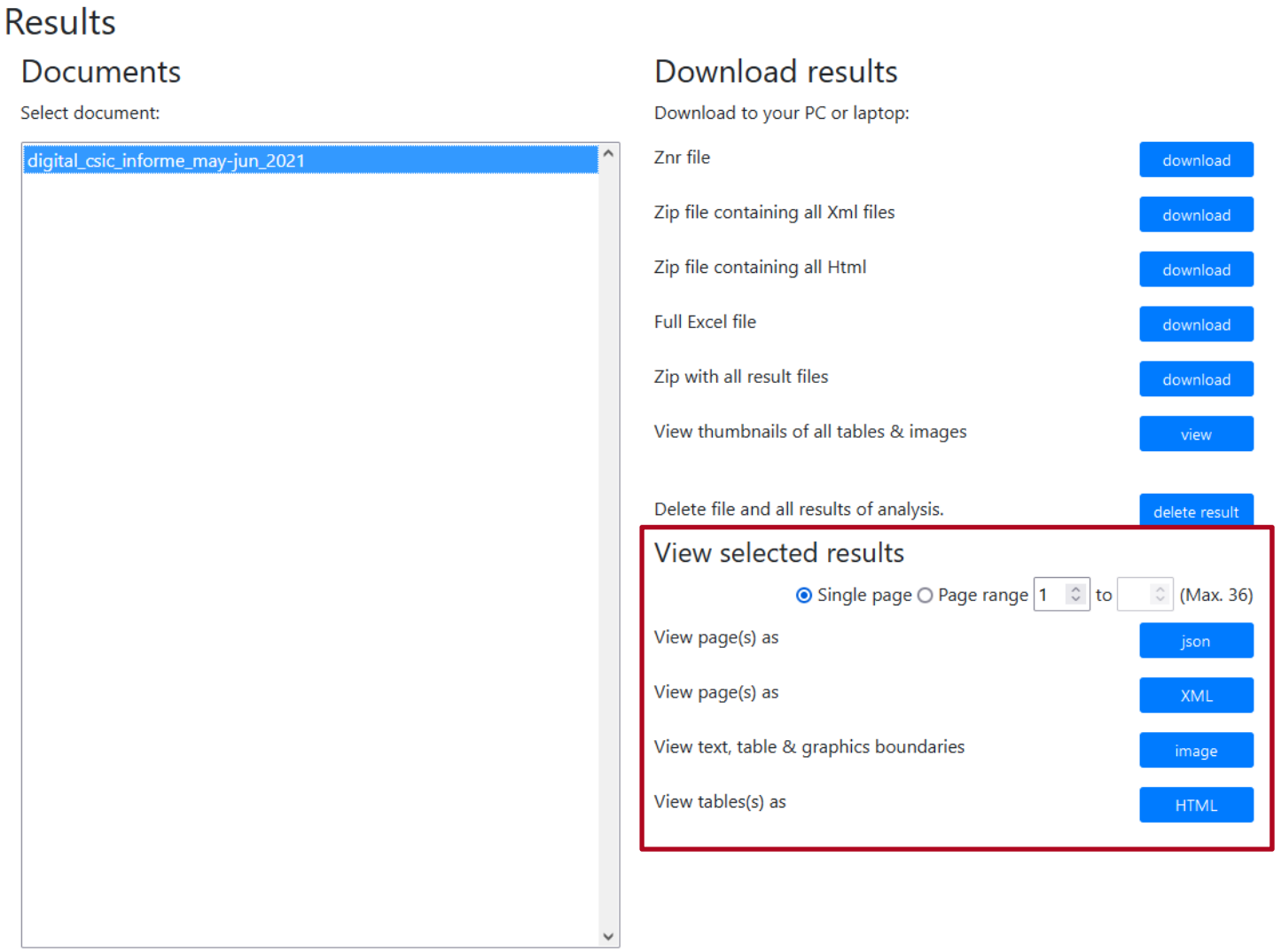
Este fichero XML no parece tener ninguna información de estilo asociada. Se muestra debajo el árbol del documento.

```
−<page number="6" width="595" height="842" rot="0" ctm1="1.00" ctm2="0.00" ctm3="0.00" ctm4="-1.00" ctm5="0.00" ctm6="841.92">
  −<table last_header_row="1" last_header_col="-1" last_header_row_pos="620" last_header_col_pos="-1">
    −<tr bgcolor="#F5F5F5">
        <td/>
        <td align="left">Nº de registros disponibles</td>
        <td align="left">Visualizaciones de registros</td>
        <td align="left">Descargas de registros</td>
      </tr>
    −<tr>
        <td align="left">Mayo</td>
        <td align="left">229.657</td>
        <td align="left">125.019.707</td>
        <td align="left">117.423.904</td>
      </tr>
    −<tr>
        <td align="left">Junio</td>
        <td align="left">232.159</td>
        <td align="left">126.302.291</td>
        <td align="left">118.627.942</td>
      </tr>
  </table>
</page>
```

DIGITAL.CSIC

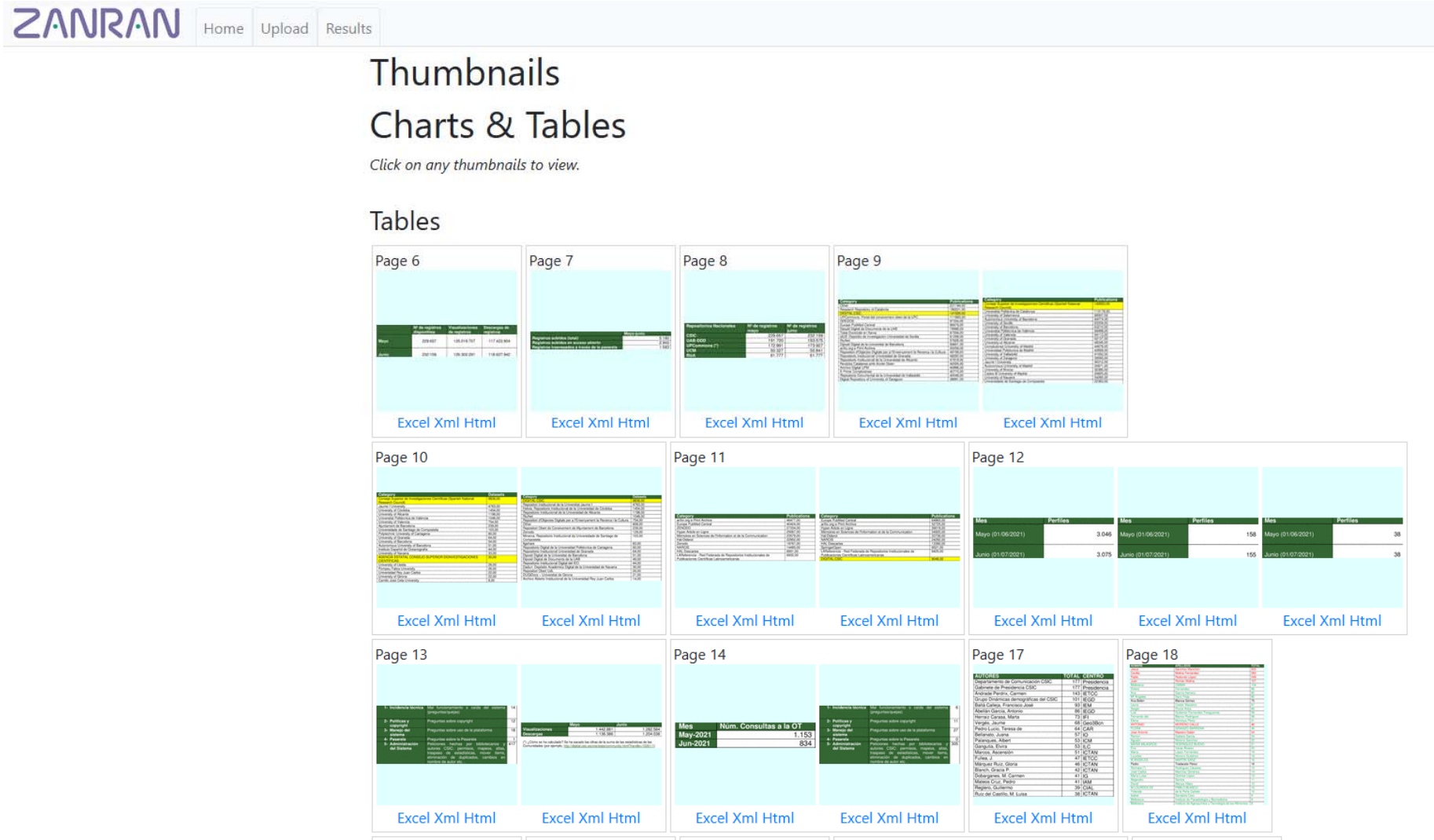
Zanran

5. Ver resultados



Zanran

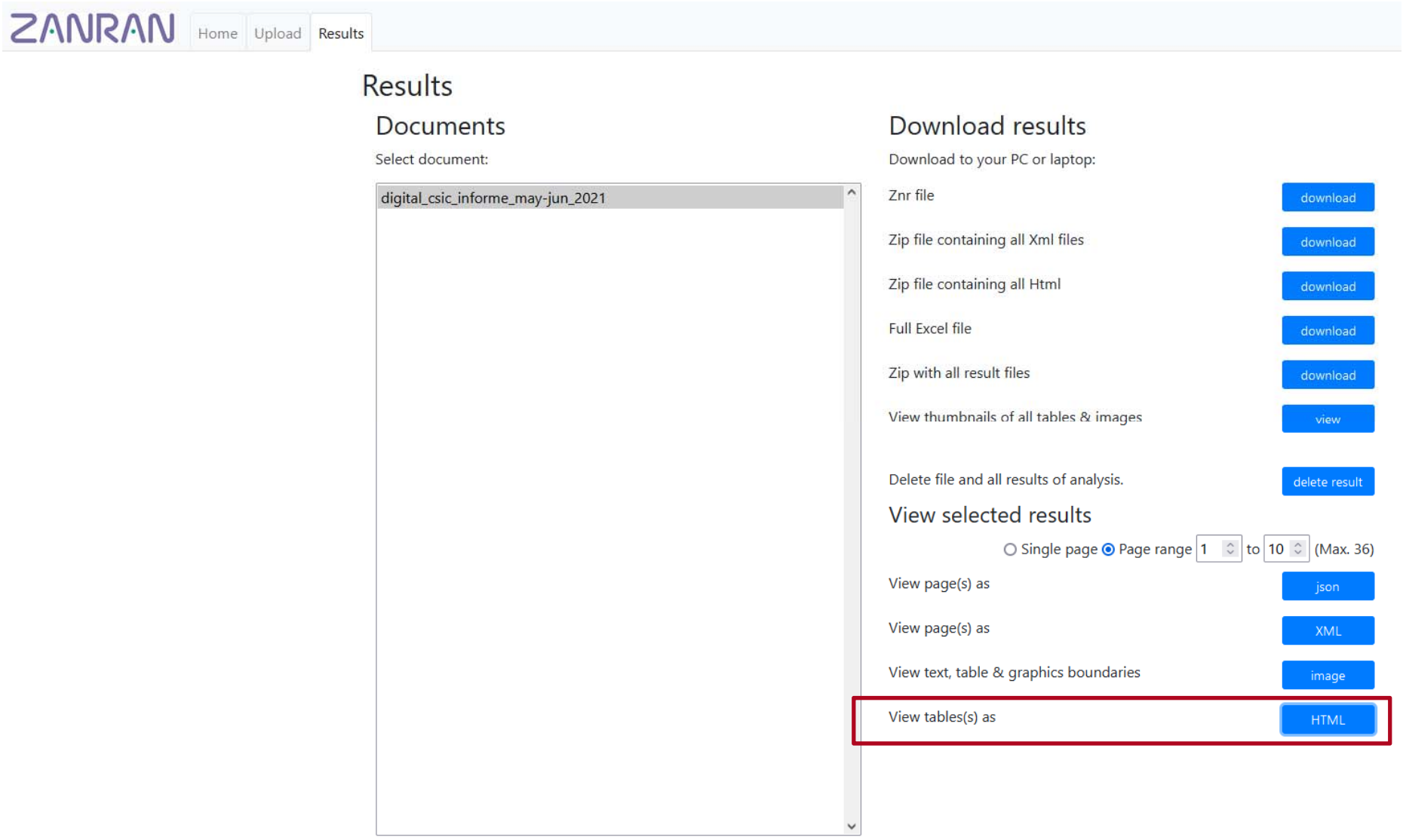
5. Ver resultados



DIGITAL.CSIC

Zanran

5. Ver resultados



DIGITAL.CSIC

Zanran

5. Ver resultados

digital_csic_informe_may-jun_2021 - page 8

Repositorios Nacionales	Nº de registros mayo	Nº de registros junio
CSIC	229.657	232.159
UAB-DDD	191.700	193.575
UPCommons (*)	172.991	173.927
UCM	50.327	50.841
RUA	61.777	61.777

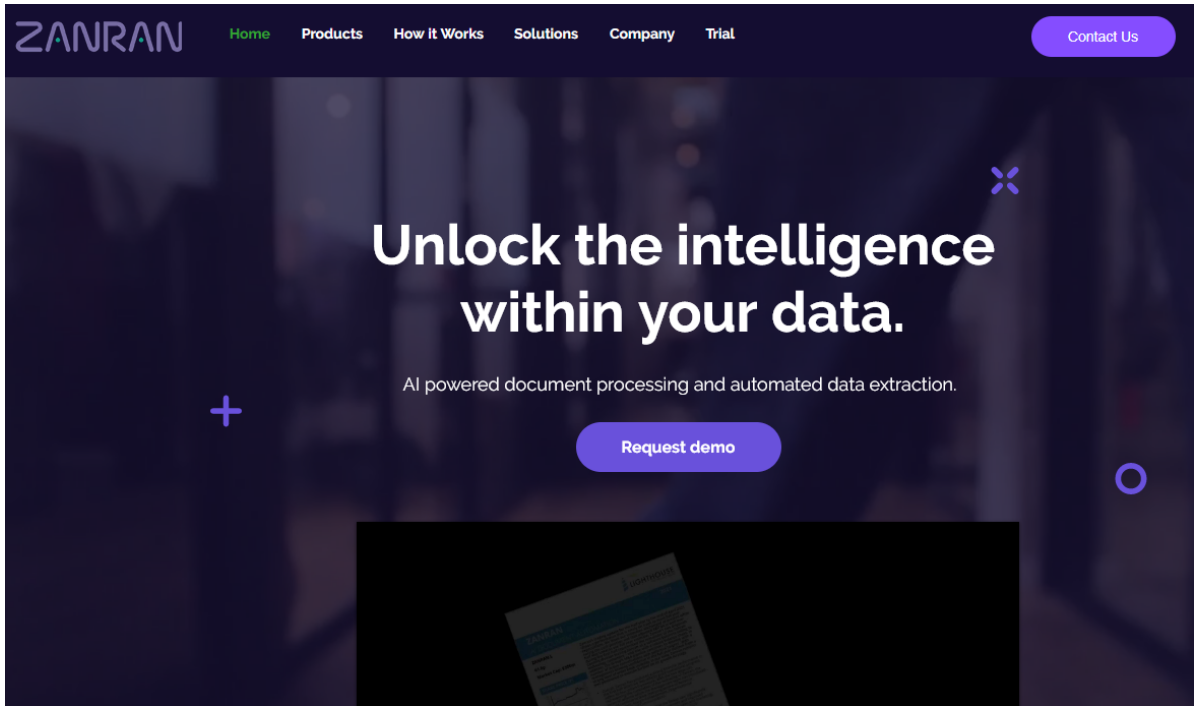
Caution: The extraction process can occasionally result in data corruption. Please compare table data against the original image below to ensure accuracy.

Repositorios Nacionales	Nº de registros mayo	Nº de registros junio
CSIC	229.657	232.159
UAB-DDD	191.700	193.575
UPCommons (*)	172.991	173.927
UCM	50.327	50.841
RUA	61.777	61.777

DIGITAL.CSIC

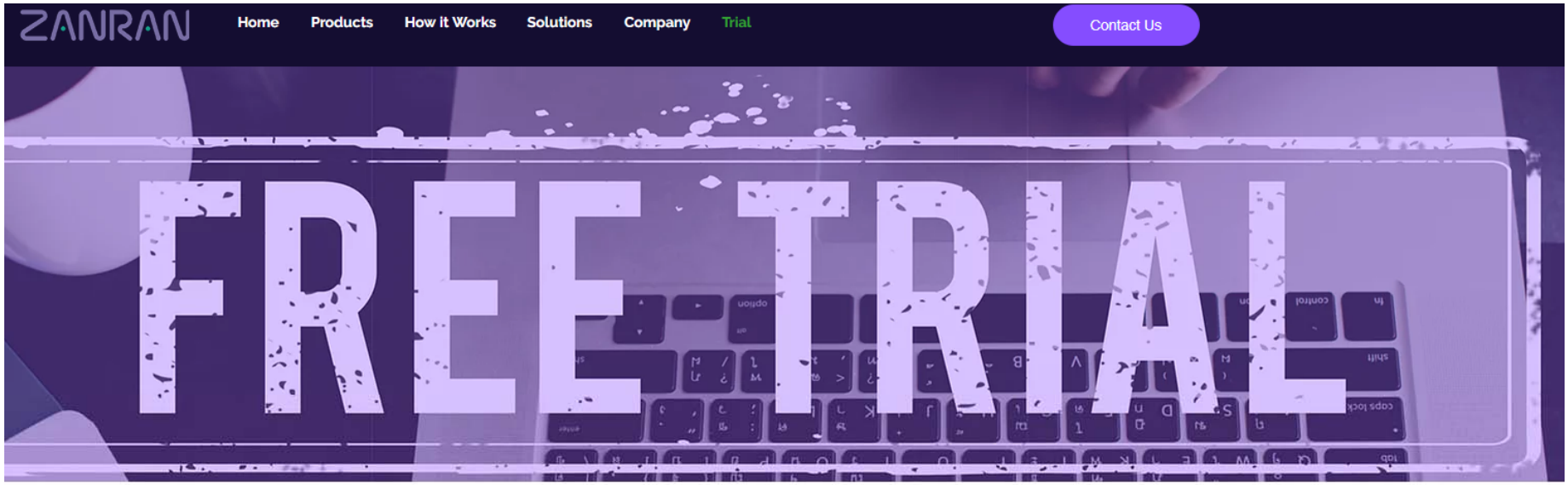
Cómo gestionar los datos de investigación: buenas prácticas, estándares, herramientas

PROGRAMA



https://www.zanran.com/

Cómo gestionar los datos de investigación: buenas prácticas, estándares, herramientas



Free trial - Zanran Scaffolder

This free-to-use server will enable you to convert your PDFs. For the trial, you **must** use digital PDFs - direct from Word or InDesign for example - **not** scanned documents.

Free trial - Zanran Scaffolder

This free-to-use server will enable you to convert your PDFs. For the trial, you **must** use digital PDFs - direct from Word or InDesign for example - **not** scanned documents.

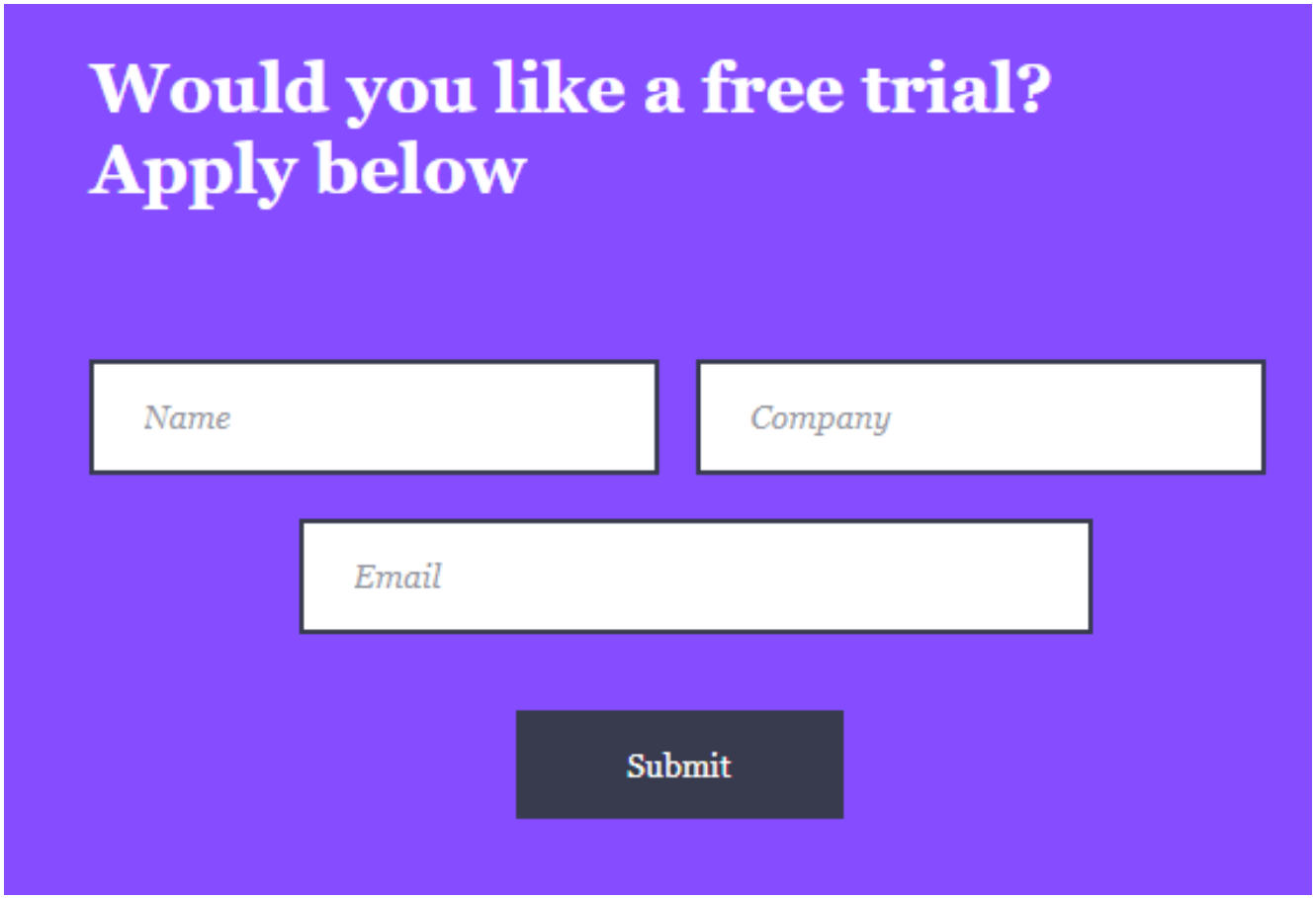
The Scaffolder algorithms in this server have been trained on:

- brokers reports
- financial statements (annual reports, 10-K, 10-Q, etc).

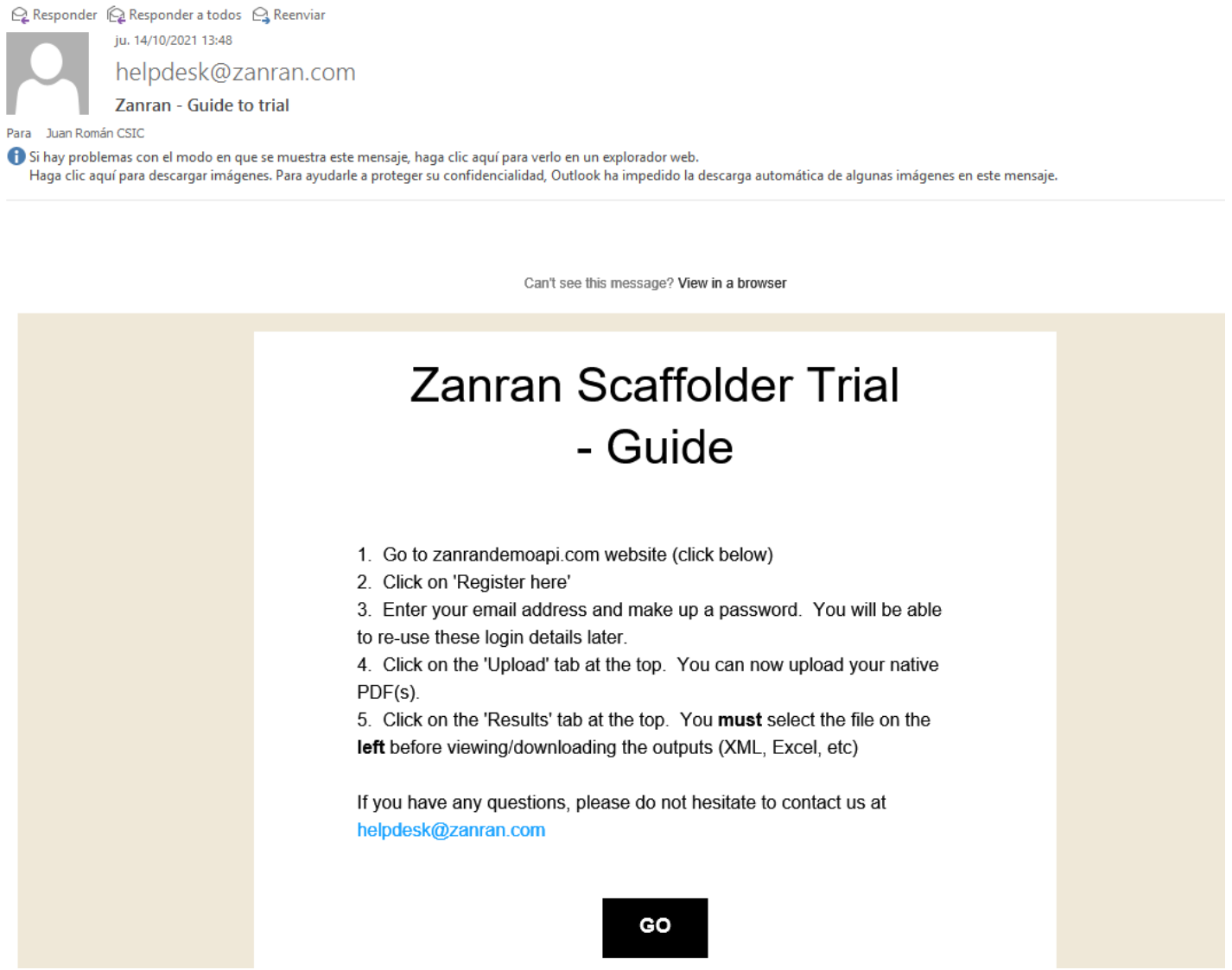
Other documents and reports will work, but with less accuracy.

You get back:

- Excel files of extracted tables
- HTML of extracted tables
- XML files giving a complete description of each page (text, tables, images)
- ZNR files for viewing on the Zanran Workbench


Would you like a free trial?
Apply below
Name
Company
Email
Submit

Cómo gestionar los datos de investigación: buenas prácticas, estándares, herramientas



Responder Responder a todos Reenviar

ju. 14/10/2021 13:48

helpdesk@zanran.com

Zanran - Guide to trial

Para Juan Román CSIC

Si hay problemas con el modo en que se muestra este mensaje, haga clic aquí para verlo en un explorador web.
Haga clic aquí para descargar imágenes. Para ayudarle a proteger su confidencialidad, Outlook ha impedido la descarga automática de algunas imágenes en este mensaje.

Can't see this message? **View in a browser**

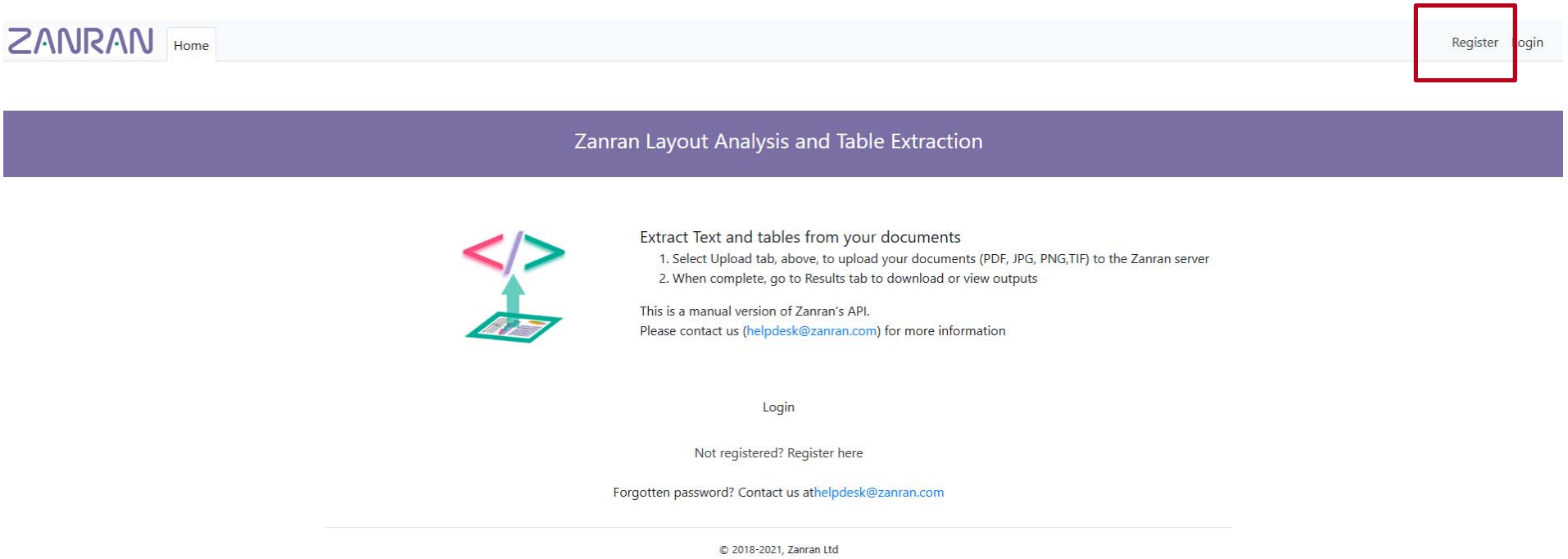
Zanran Scaffolder Trial - Guide

1. Go to zanrandemoapi.com website (click below)
2. Click on 'Register here'
3. Enter your email address and make up a password. You will be able to re-use these login details later.
4. Click on the 'Upload' tab at the top. You can now upload your native PDF(s).
5. Click on the 'Results' tab at the top. You **must** select the file on the **left** before viewing/downloading the outputs (XML, Excel, etc)

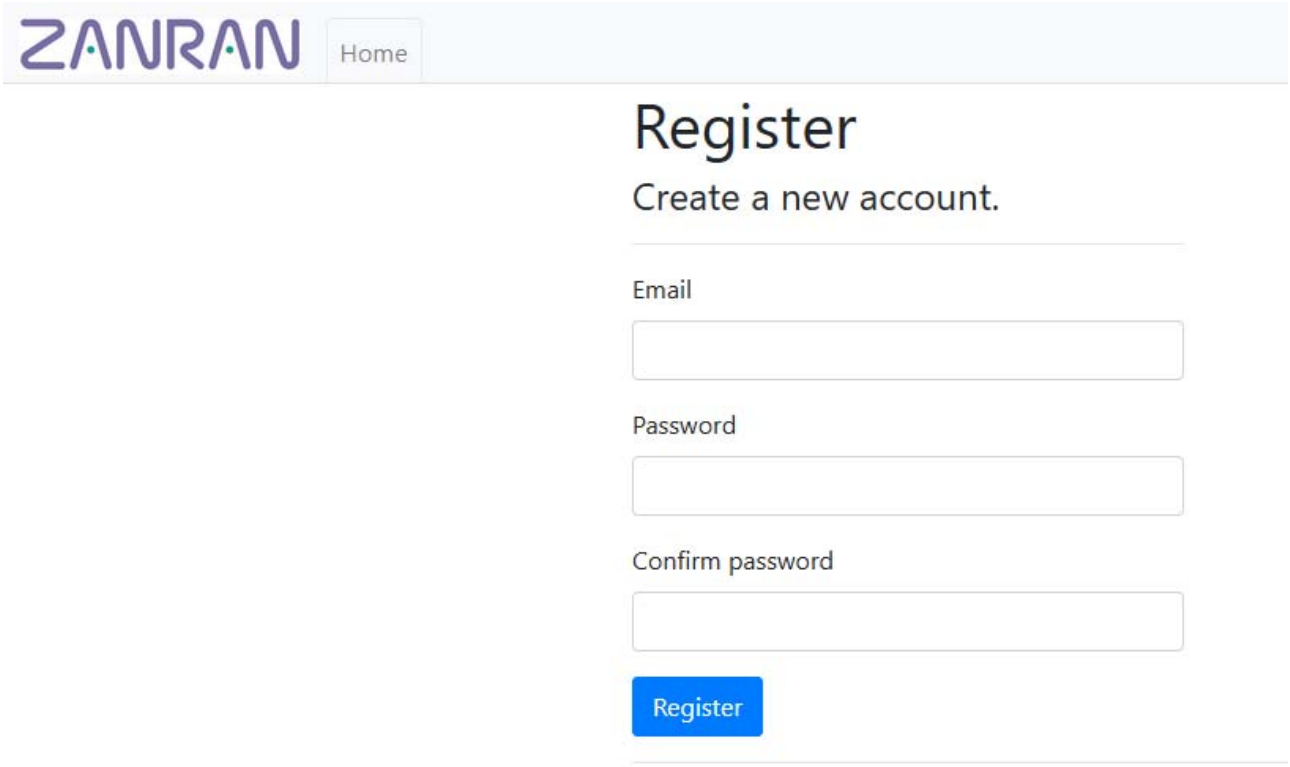
If you have any questions, please do not hesitate to contact us at helpdesk@zanran.com

GO

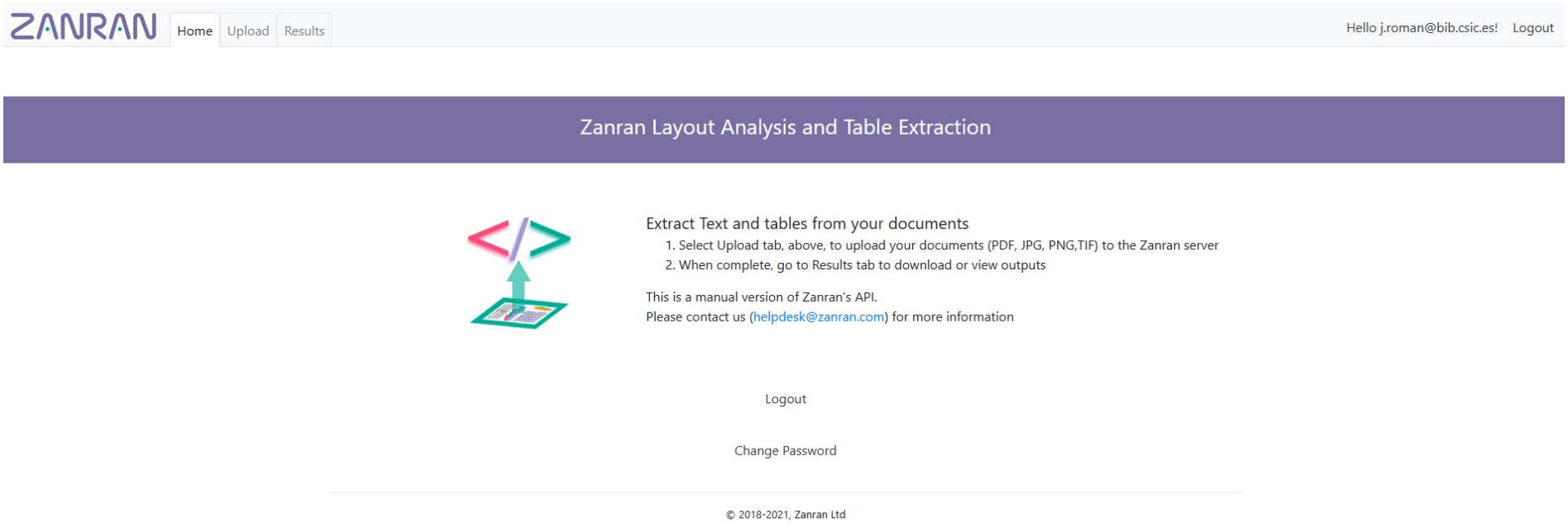
Cómo gestionar los datos de investigación: buenas prácticas, estándares, herramientas



Cómo gestionar los datos de investigación: buenas prácticas, estándares, herramientas

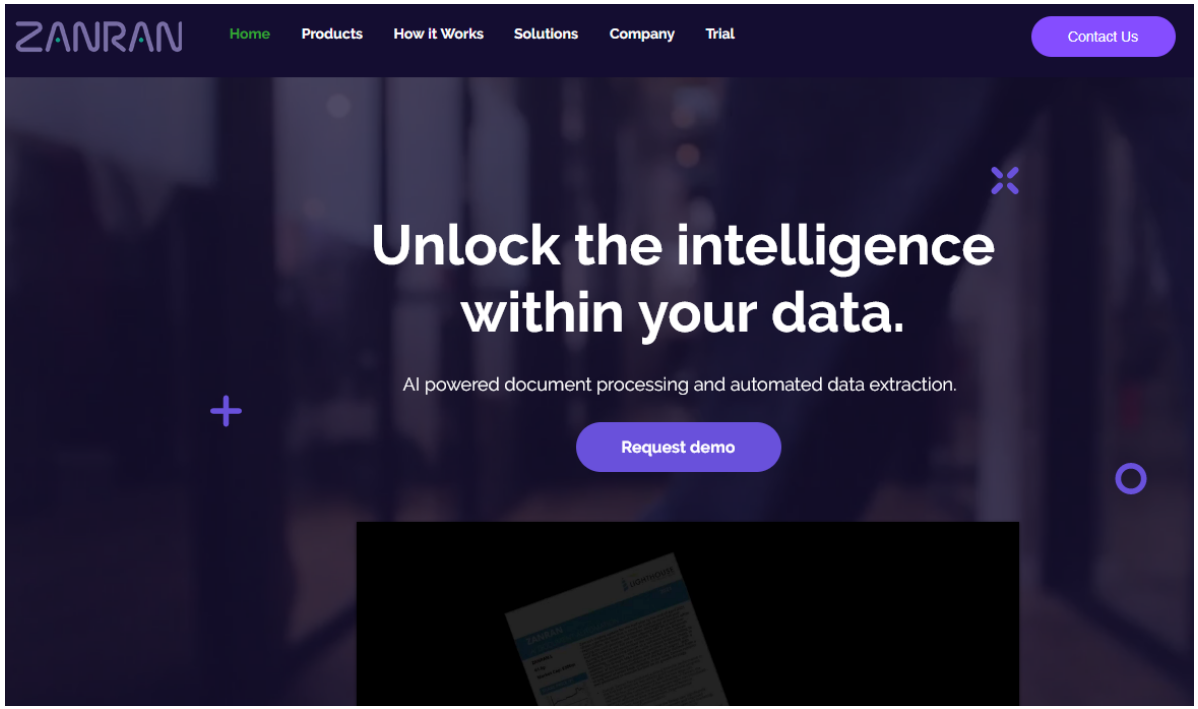


Cómo gestionar los datos de investigación: buenas prácticas, estándares, herramientas



DIGITAL.CSIC

PRÁCTICAS



https://www.zanran.com/

Cómo gestionar los datos de investigación: buenas prácticas, estándares, herramientas

ZANRAN

1. Solicita acceso a la herramienta (*free trial*)

   https://www.zanran.com/trial

2. Primera toma de contacto

   https://www.zanran.com/how-it-works

ZANRAN

grasslands_RothC.pdf
seoke_Botswana_pXRF.pdf
water_quality .pdf

1. Sube uno de los artículos propuestos
2. Descarga los resultados
3. Mira los resultados

DIGITAL.CSIC

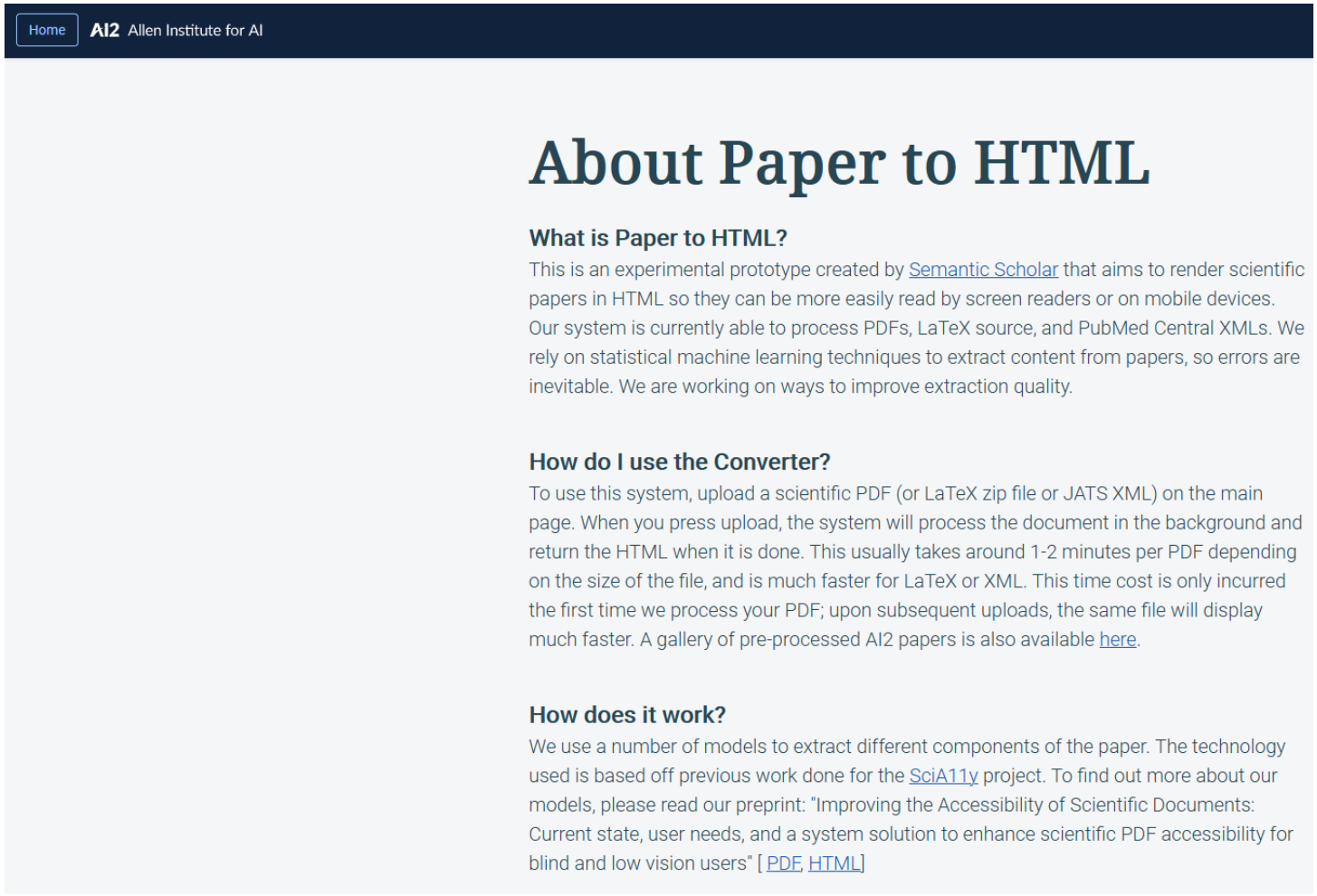
Paper to HTML Converter

1. Sobre la herramienta
2. Ejemplos
3. Carga de un trabajo
4. Ayuda y configuración
5. Advertencias

https://papertohtml.org/

Paper to HTML Converter

1. Sobre la herramienta



https://papertohtml.org/about

DIGITAL.CSIC

Paper to HTML Converter

Paper to HTML Converter

2. Ejemplos

AI2 Paper Gallery

This is a gallery of papers published by researchers at AI2. Clicking on one will render a reader-friendly HTML version.

RealToxicityPrompts: Evaluating Neural Toxic Degeneration in Language Models

2020 FINDINGS

Samuel Gehman Suchin Gururangan Maarten Sap Yejin Choi Noah A. Smith

Domain-Targeted, High Precision Knowledge Extraction

2017 TACL

Bhavana Dalvi Niket Tandon P. Clark

Expanding Holographic Embeddings for Knowledge Completion

2018 NeurIPS

Yexiang Xue Yang Yuan Zhitian Xu Ashish Sabharwal

What's in an Explanation? Characterizing Knowledge and Inference Requirements for Elementary Science Exams

2016 COLING

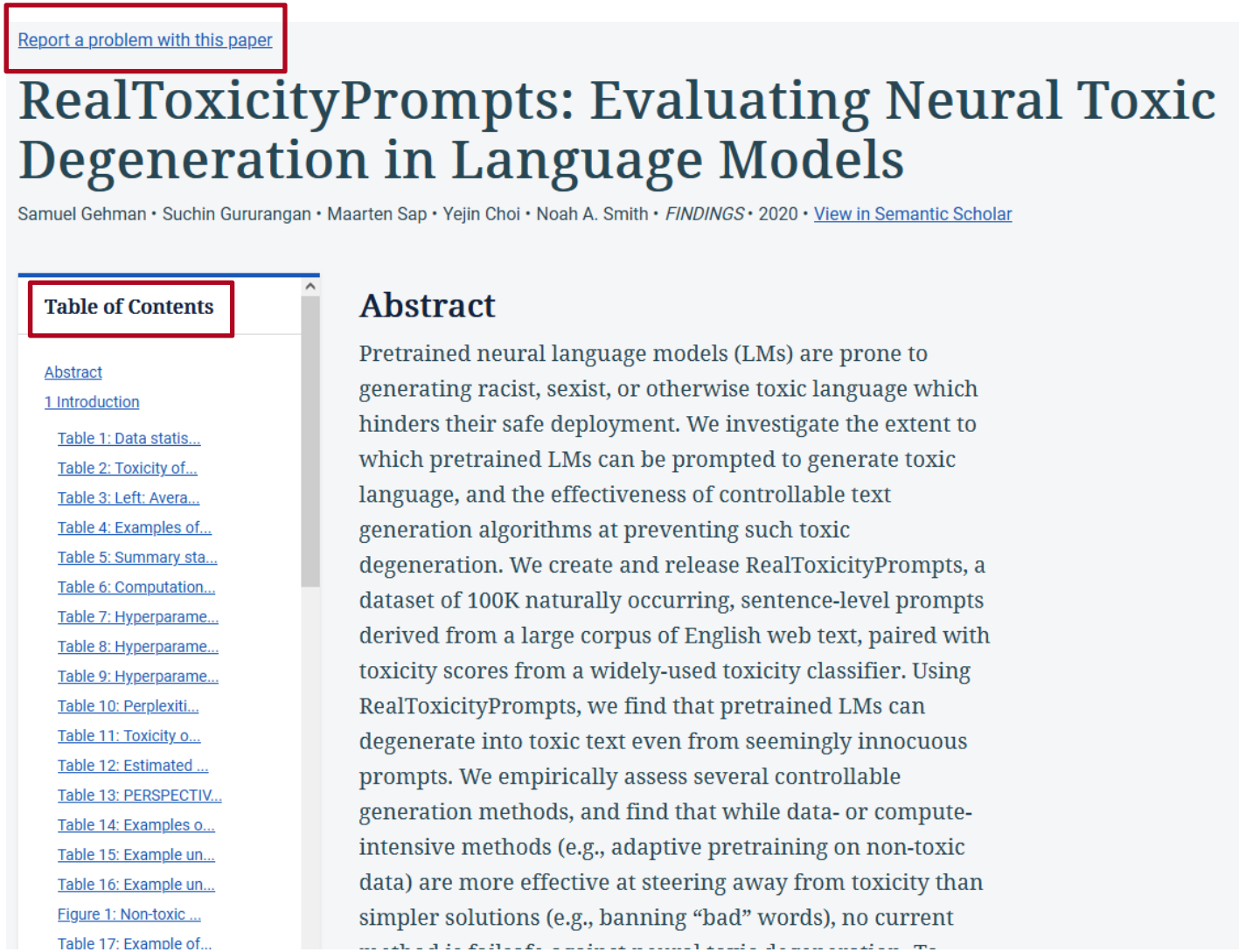
Peter Jansen Niranjan Balasubramanian M. Surdeanu P. Clark

DIGITAL.CSIC

Paper to HTML Converter

Paper to HTML Converter

2. Ejemplos



DIGITAL.CSIC

Paper to HTML Converter

Paper to HTML Converter

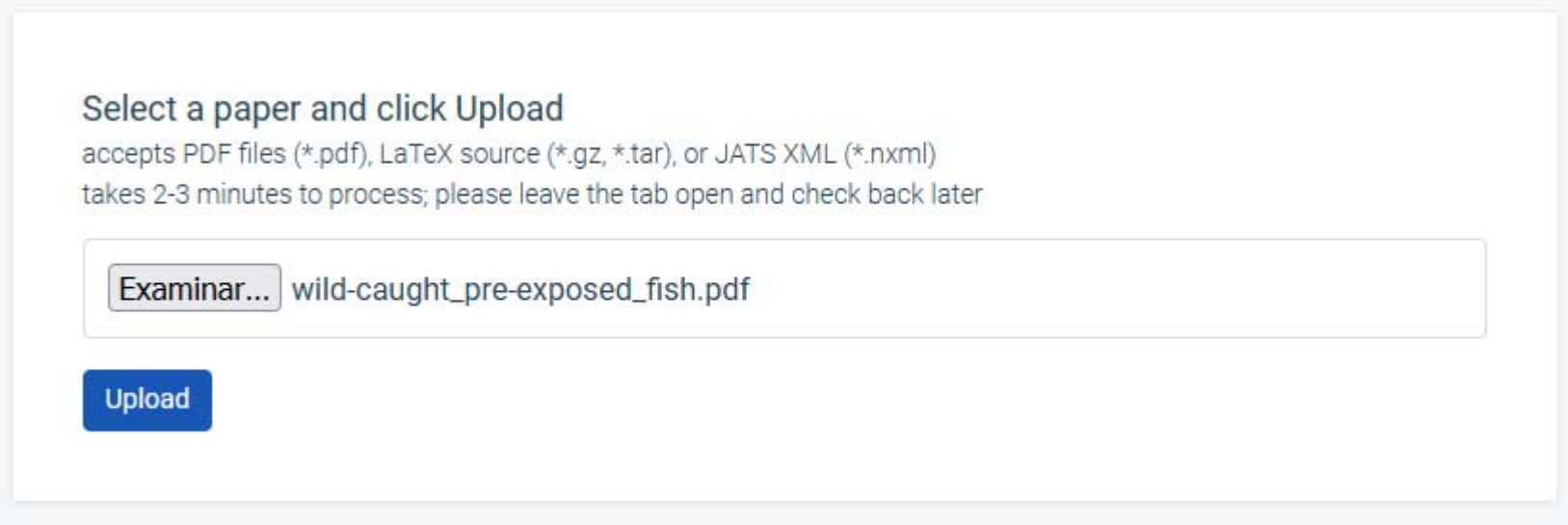
2. Ejemplos

```
<p tabindex="0">Attribute Conditioning (ATCON) Inspired by <a href="#BIBREF27"  id="BIBREF27-Exp-ref"> Ficler and Goldberg (2017)</a> and Keskar et al. 2019, we prepend a corresponding toxicity attribute token (<

<h2 id="section-body-13" tabindex="0">5.2 Decoding-Based Detoxification </h2>

<p tabindex="0">Noting the additional cost of training language models further, we explore three detoxifying strategies that only rely on altering the decoding algorithm and are therefore more readily usable by m

<p tabindex="0">Vocabulary Shifting (VOCAB-SHIFT) Inspired by <a href="#BIBREF25"  id="BIBREF25-5.2 Decoding-Based Detoxification-ref"> Eisenstein et al. (2011)</a> and Ghosh et al. 2017, we learn a 2-dimensional

<h2 id="section-body-14" tabindex="0">Word Filtering (Word Filter) </h2>

<p tabindex="0">We also implement a language model blocklist, disallowing a set of words from being generated by GPT-2. We set the probability of generating any word from a list 13 of profanity, slurs, and swearw

<p tabindex="0">PPLM We use the recently released PPLM <a href="#BIBREF16"  id="BIBREF16-Word Filtering (Word Filter)-ref"> (Dathathri et al., 2020)</a> . This decoding method operates on GPT-2 by altering the pa

<p tabindex="0">To prevent a list of banned words from being generated, we first encode each word as a sequence of BPE tokens. During generation, we set any vocabulary logits that would complete the token sequenc

<p tabindex="0">PPLM We replicate the experimental setup for language detoxification described by <a href="#BIBREF16"  id="BIBREF16-Word Filtering (Word Filter)-ref"> Dathathri et al. (2020)</a> using the release

<h2 id="section-body-15" tabindex="0">5.3 Effect Of Controllable Solutions On Generation Toxicity </h2>

<p tabindex="0">We investigate the effectiveness of our detoxification methods under REALTOXICITYPROMPTS, following the same generation procedures and experimental setups outlined in §4. Listed in <a href="#TABRE

<p tabindex="0">Prompts That Challenge All Models We find that certain prompts consistently cause all models to generate toxicity (e.g., the four prompts in Figure  1 ). Specifically, there are 327 prompts that y

<h2 id="section-body-16" tabindex="0">6 Analyzing Toxicity In Web Text </h2>

<p tabindex="0">To further investigate the phenomenon of neural toxic degeneration, and partially motivated by the surprising effectiveness of domain-adaptive pretraining on non-toxic data, we turn our focus to t
```

DIGITAL.CSIC

Paper to HTML Converter

3. Carga de un trabajo



DIGITAL.CSIC

Paper to HTML Converter

Paper to HTML Converter

3. Carga de un trabajo

Report a problem with this paper

Improving water quality does not guarantee fish health: Effects of ammonia pollution on the behaviour of wild-caught pre-exposed fish

Patricia Soler • Melissa Faria • Carlos Barata • Eduardo García-Galea • Beatriz Lorente • Dolors Vinyoles

Table of Contents

Abstract
Introduction
Study Area And Fish Sampling
Figure 1...
Table 1...
Plos One
Experimental Design
Figure 2...
Plos One
Biochemical Determination
Statistical Analyses
Plos One
Behavioural Variables
Figure 3...
Biochemical Determination
Figure 4...

Abstract

Ammonia is a pollutant frequently found in aquatic ecosystems. In fish, ammonia can cause physical damage, alter its behaviour, and even cause death. Exposure to ammonia also increases fish physiological stress, which can be measured through biomarkers. In this study, we analysed the effect of sublethal ammonia concentrations on the behaviour and the oxidative stress of Barbus meridionalis that had been pre-exposed to this compound in the wild. Wild-caught fish from a polluted site (pre-exposed fish) and from an unpolluted site (non-pre-exposed fish) were exposed, under experimental conditions, to total ammonia concentrations (TAN) of 0, 1, 5, and 8 mg/L. Swimming activity, feeding behaviour, and oxidative stress response based on biomarkers were analysed. Pre-exposed fish showed both an altered behaviour and an altered oxidative stress response in the control treatment (0 mg/L).

DIGITAL.CSIC

Paper to HTML Converter

Paper to HTML Converter

3. Carga de un trabajo

Table of Contents

Abstract

Introduction

Study Site And Sampling

Table 1...

Sediment Dating

Edna Analyses

Organic Matter, Corg And 13 Corg Analyses

Data Analysis

Dating And Sediment Accumulation Rates

Reconstruction Of H. Stipulacea Invasion Based On Edna

Figure 1...

Figure 2...

Organic Carbon Stocks, Burial Rates And Sources

Figure 3...

Table 2...

Figure 4...

after the arrival of H. stipulacea as well as between neighbour exotic (H. stipulacea) and native (C. nodosa and Posidonia oceanica) meadows since the invasion.

Study Site And Sampling

Location	Sampling date	Replicate ID	Depth (m)	DBD	OM	Dating	eDNA	C_{org}, and $\delta^{13}C_{org}$
Kalami (Greece)	18/07/17	*H. stipulacea* (1)*	7	✓	✓	✓	8	✓
		H. stipulacea (2)		✓	✓	✓	8	✓
		H. stipulacea (3)*		✓	✓			
		H. stipulacea (4)*		✓	✓			
		C. nodosa (1)	7	✓	✓	✓	6	✓
		C. nodosa (2)		✓	✓			
		C. nodosa (3)		✓	✓			
Maridati (Greece)	21/07/17	*H. stipulacea* (1)	10.5	✓	✓	✓	7	✓
		H. stipulacea (2)		✓	✓	✓	6	
		H. stipulacea (3)		✓	✓			
		H. stipulacea (4)	5.5	✓	✓		4	
		C. nodosa (1)		✓	✓	✓	8	✓
		C. nodosa (2)		✓	✓			
		C. nodosa (3)		✓	✓			
Limassol (Cyprus)	04/09/17	*H. stipulacea* (1)	4.3	✓	✓	✓	8	✓
		H. stipulacea (2)		✓	✓	✓	8	
		H. stipulacea (3)		✓	✓			
		C. nodosa (1)	3.5	✓	✓	✓	8	✓
		C. nodosa (2)		✓	✓			
		C. nodosa (3)		✓	✓			
Port (Cyprus)	04/09/17	*H. stipulacea* (1)	8	✓	✓	✓	8	✓
		H. stipulacea (2)		✓	✓	✓	8	
		H. stipulacea (3)		✓	✓			
	06/09/17	*P. oceanica* (1)	8	✓	✓	✓	8	✓
		P. oceanica (2)		✓	✓			
		P. oceanica (3)		✓	✓			

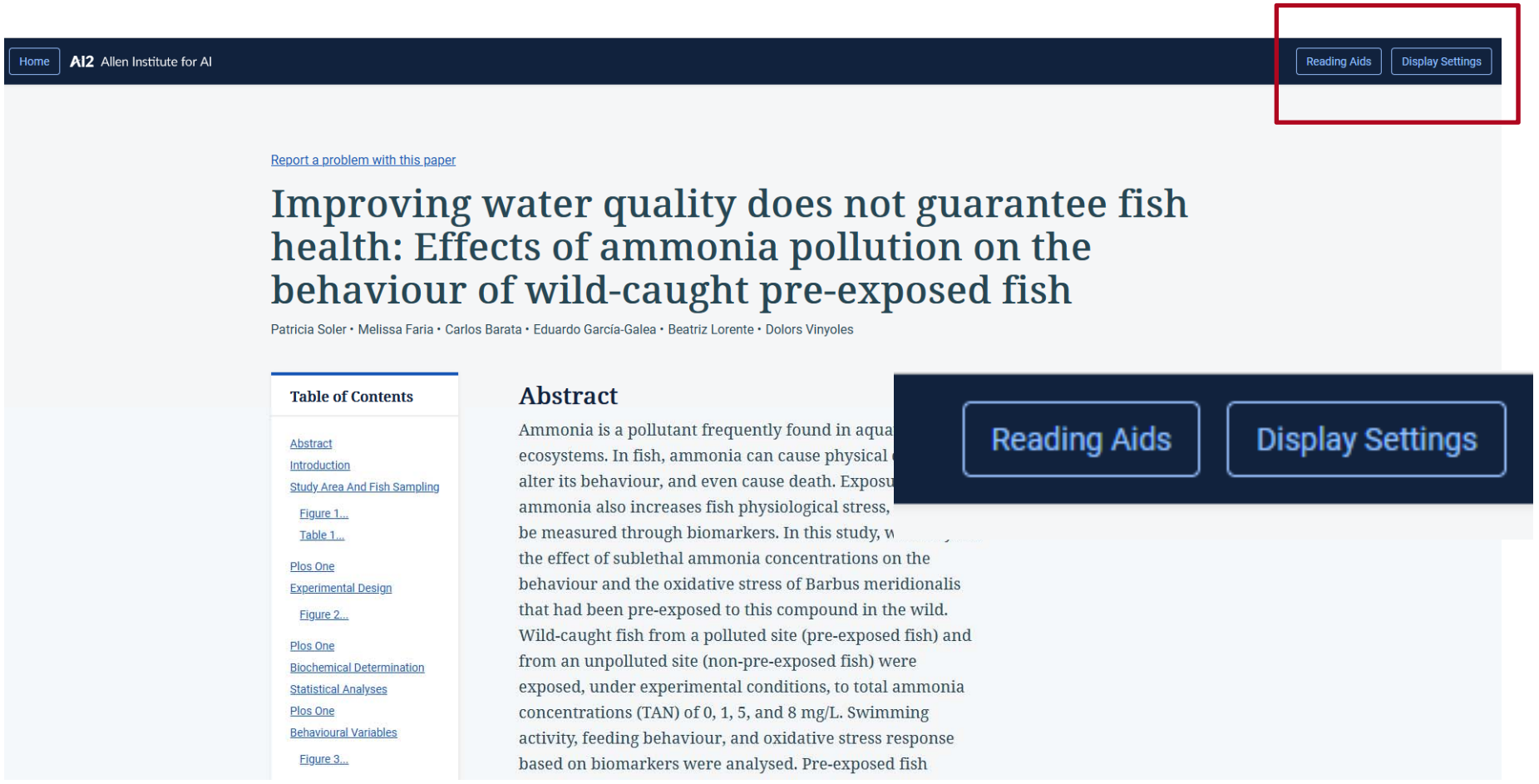
79

DIGITAL.CSIC

Paper to HTML Converter

Paper to HTML Converter

4. Ayuda y configuración del HTML

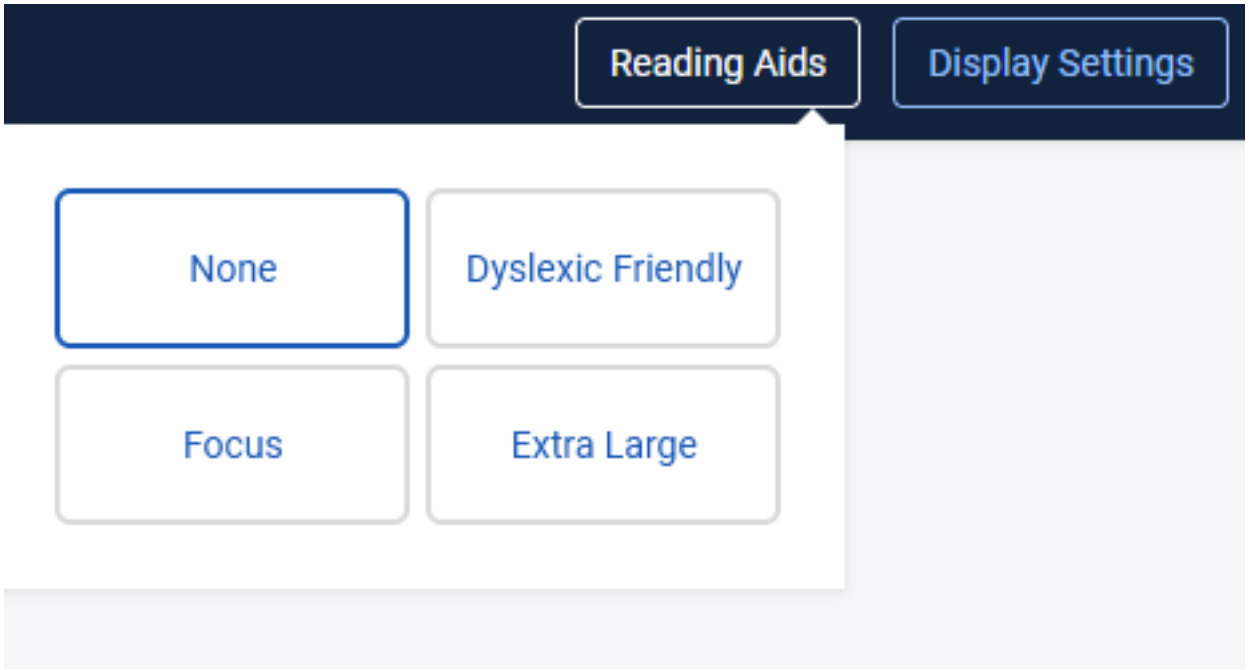


DIGITAL.CSIC

Paper to HTML Converter

Paper to HTML Converter

4. Ayuda y configuración del HTML



DIGITAL.CSIC

Paper to HTML Converter

Paper to HTML Converter

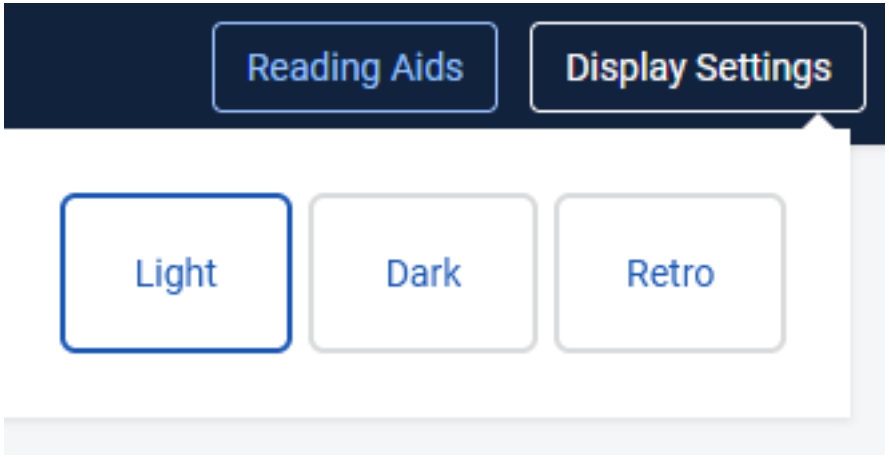
4. Ayuda y configuración del HTML

DIGITAL.CSIC

Paper to HTML Converter

Paper to HTML Converter

4. Ayuda y configuración del HTML

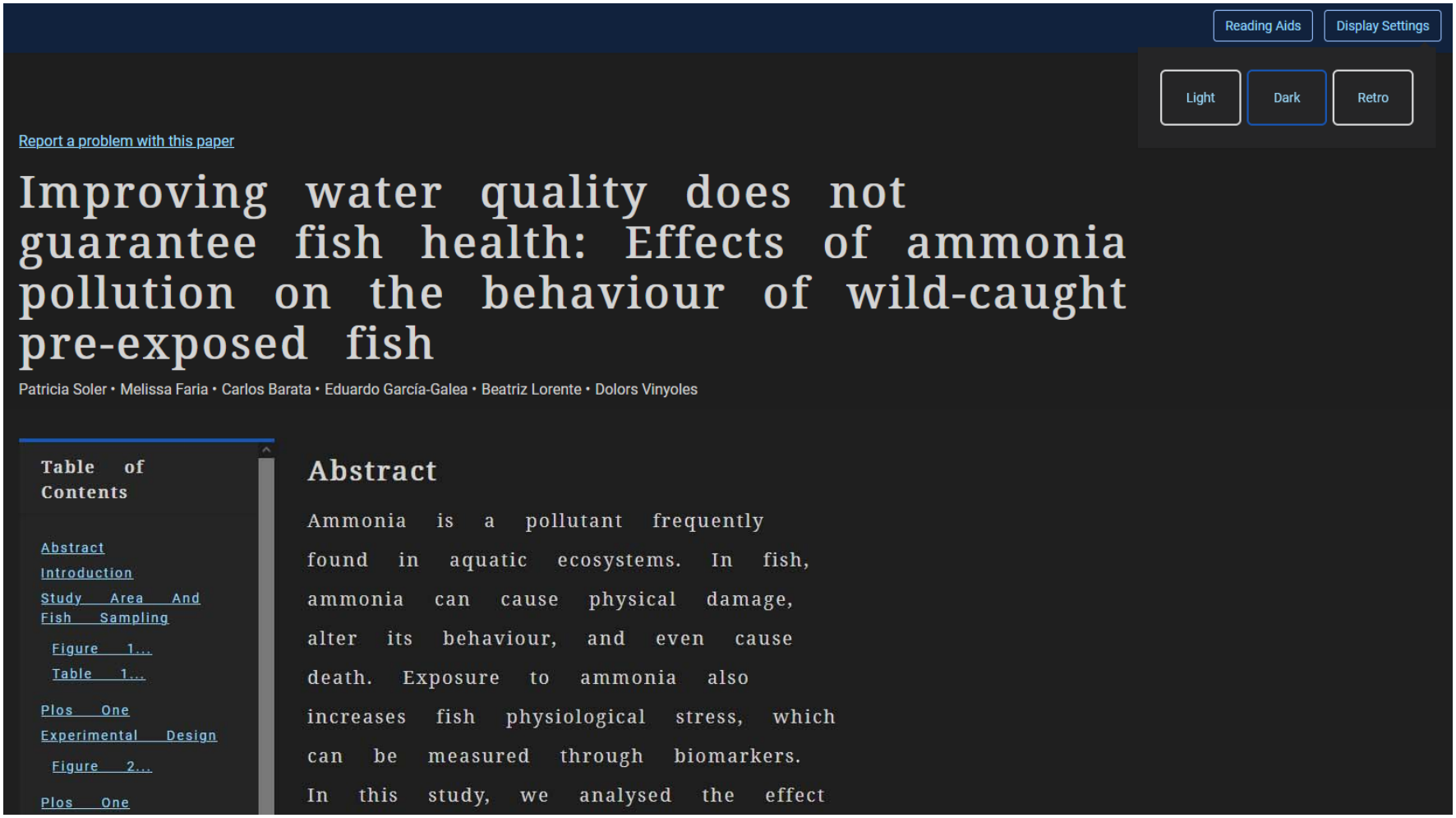


DIGITAL.CSIC

Paper to HTML Converter

Paper to HTML Converter

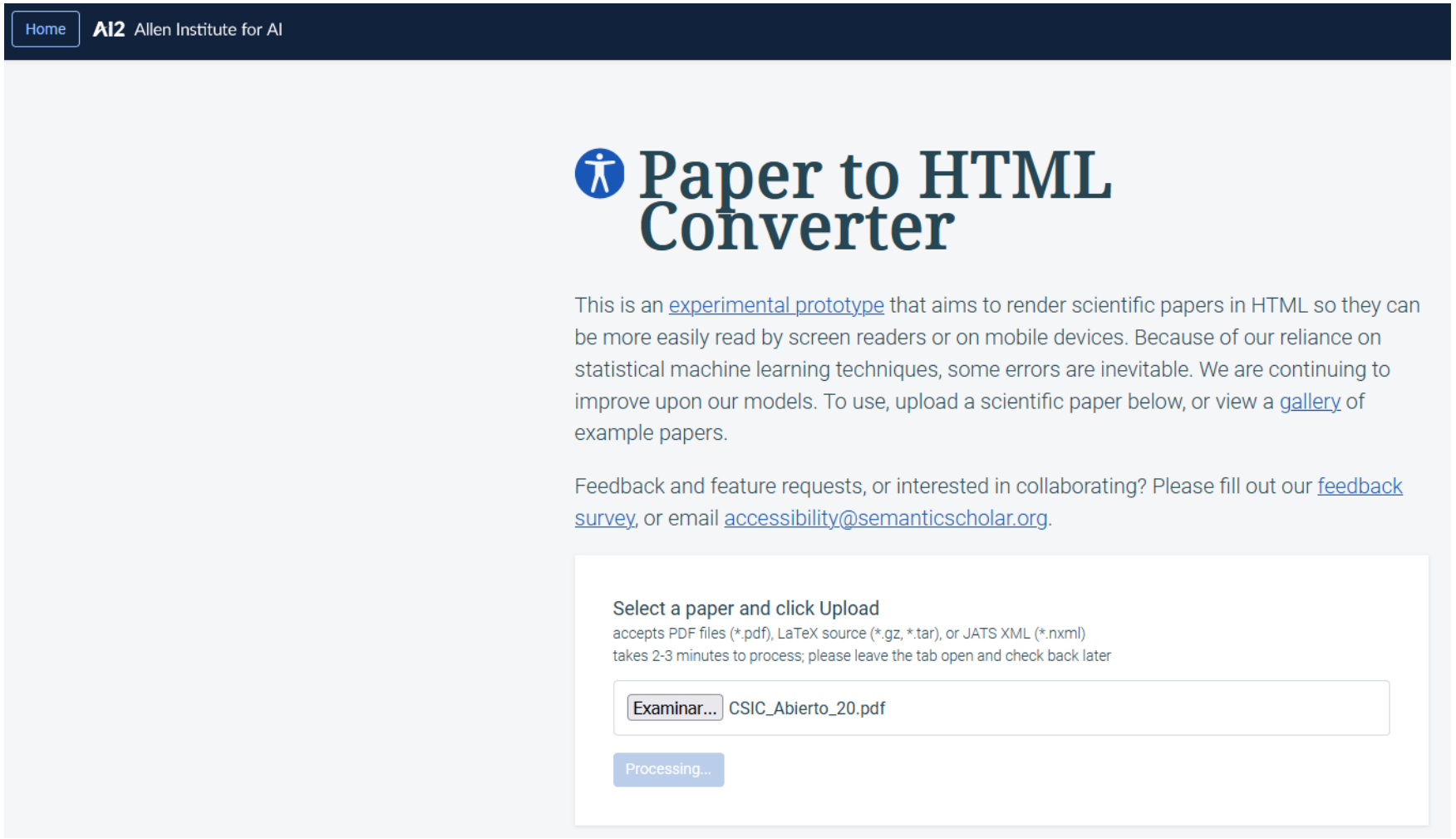
4. Ayuda y configuración del HTML



DIGITAL.CSIC

Paper to HTML Converter

5. Advertencias



Paper to HTML Converter

5. Advertencias

Paper to HTML Converter

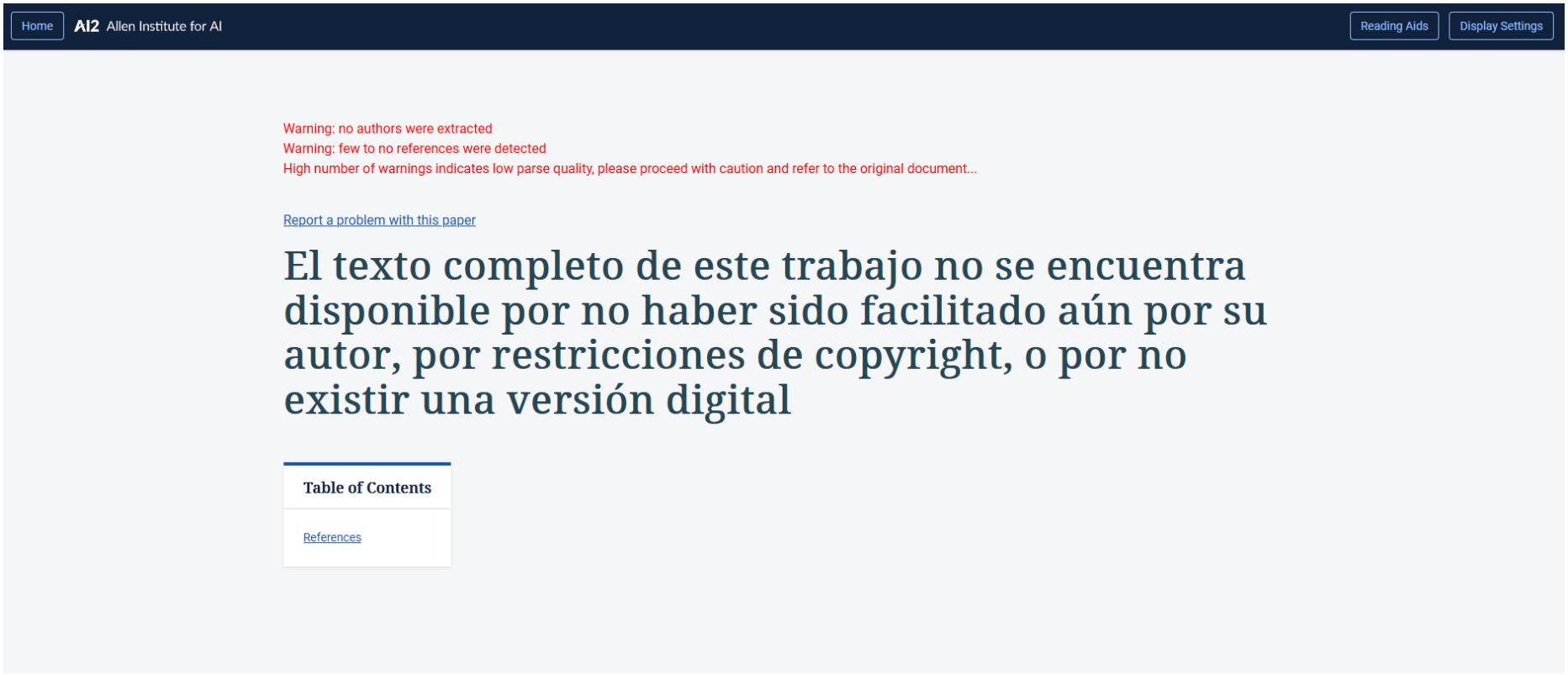
Paper to HTML Converter

5. Advertencias

Warning: no title was extracted
Warning: no authors were extracted
Warning: few to no references were detected
High number of warnings indicates low parse quality, please proceed with caution and refer to the original document...

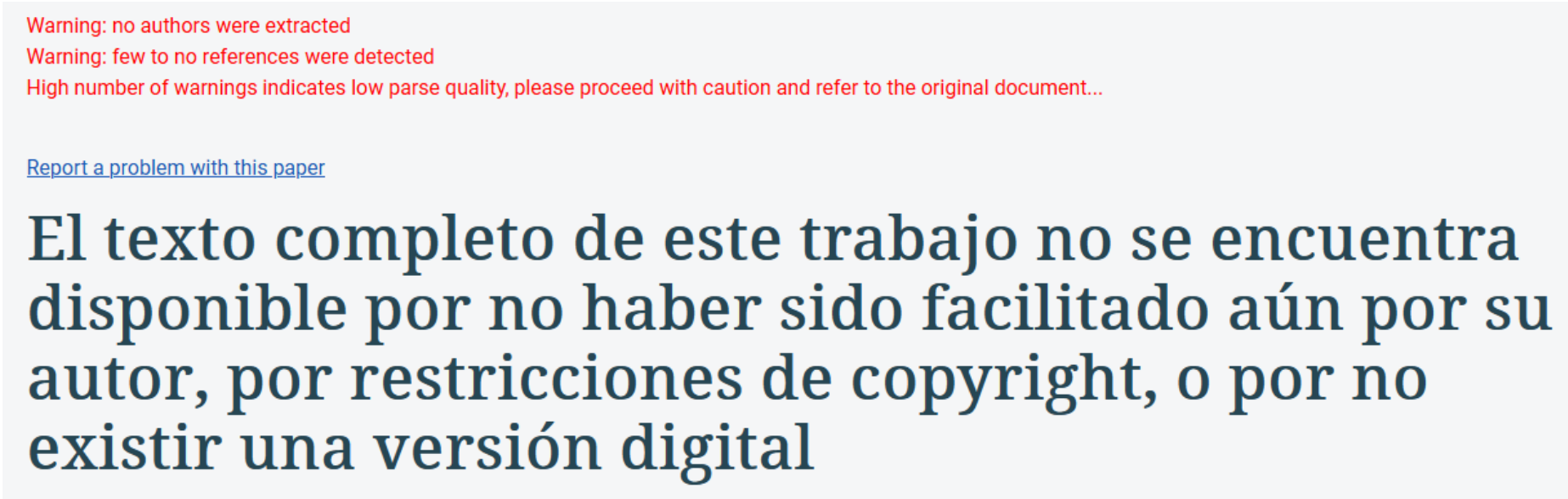
Paper to HTML Converter

5. Advertencias



Paper to HTML Converter

5. Advertencias



DIGITAL.CSIC

Cómo gestionar los datos de investigación: buenas prácticas, estándares, herramientas

PRÁCTICAS

https://papertohtml.org/

Paper to HTML Converter

Genomic_Prediction_revised_LaText.pdf
Halophila_stipulacea_Postprint.pdf

1. Busca en DIGITAL.CSIC algunos trabajos en acceso abierto o trabaja con los aportados
2. Conviértelos en HTML
3. Prueba las opciones de lectura y configuración

DIGITAL.CSIC

GRACIAS!!

dudas/preguntas

j.roman@bib.csic.es