

---

## Plan Overview

*A Data Management Plan created using DMPonline*

**Title:** REFOMAD: Análisis de la relación entre la Renta y la Formación en los distritos de Madrid.

**Creator:**EquipoMorado DLC

**Affiliation:** University of Oviedo

**Funder:** European Commission

**Template:** Horizon Europe Template

### **Project abstract:**

Este proyecto tiene como objetivo investigar la relación entre la formación y la renta en los distritos de Madrid. Para ello, se llevará a cabo un análisis detallado del nivel de formación de los habitantes mayores de 25 años en los diferentes distritos de Madrid y se comparará con la renta media de las personas de cada distrito. A través de este estudio, esperamos comprender mejor cómo la formación y la renta interactúan en la ciudad y cómo pueden afectar a la creación de oportunidades y la reducción de desigualdades.

**ID:** 114381

**Start date:** 01-02-2023

**End date:** 24-05-2023

**Last modified:** 22-01-2023

# REFOMAD: Análisis de la relación entre la Renta y la Formación en los distritos de Madrid.

---

## Data Summary

### Will you re-use any existing data and what will you re-use it for?

Sí. El proyecto reutilizará datos existentes procedentes de dos fuentes gubernamentales. Los administradores de datos del proyecto evaluarán minuciosamente la calidad y relevancia de los datos existentes que se van a reutilizar, y se asegurarán de que se atribuyan y citen adecuadamente en cualquier análisis o informe.

Por un lado, los datos sobre la renta media de las personas en los veintidós distritos de Madrid se toman del Atlas de Distribución de Renta de los Hogares (ADRH), una operación estadística elaborada por el INE que se basa en la explotación de registros administrativos con el objetivo de obtener información sobre el nivel y la distribución de renta desglosada según variables demográficas básicas de la población a nivel territorial muy detallado (todos los municipios, distritos y secciones censales en que se organiza territorialmente el Estado).

Por otro lado, los niveles de formación de los adultos mayores de 25 años que residen en los distintos distritos del Ayuntamiento de Madrid y la información acerca de los centros educativos se toman del "Panel de indicadores de distritos y barrios de Madrid", un estudio sociodemográfico elaborado por el Ayuntamiento de Madrid que ofrece una visión territorial de las variables socioeconómicas, de salud, demográficas, educativas, calidad de vida, vivienda, medio ambiente, equipamientos municipales, participación ciudadana y presupuesto de los distritos de Madrid.

### What types and formats of data will the project generate or re-use?

Se utilizarán únicamente datos de tipo numérico, incluyendo enteros y floats, con excepción de las etiquetas que identifican a las variables de interés y a los distritos del Ayuntamiento de Madrid, que serán de tipo carácter. Los datos, tanto los originales como los generados para el análisis, tendrán un formato estructurado. En particular, los datos recopilados del Ayuntamiento de Madrid y los generados para el proyecto serán archivos Excel XSLX, mientras que los recopilados por el INE tienen formato CSV (Comma-Separated Values, del inglés).

### What is the purpose of the data generation or re-use and its relation to the objectives of the project?

La finalidad de la reutilización de los datos en este proyecto es analizar en profundidad el nivel de formación alcanzado por los habitantes mayores de 25 años de los distintos distritos de Madrid en función de la renta media de las personas en estos distritos. Se trata de un aspecto crucial de los objetivos del proyecto, ya que comprender cómo se correlacionan estos factores puede proporcionar información muy valiosa sobre la dinámica educativa y económica del ayuntamiento. Por ejemplo, si los datos revelan que los distritos con mayores niveles de renta tienden a tener mayores niveles de formación, esto podría indicar que la educación es un factor clave para impulsar el éxito económico en estas zonas. Por otro lado, si los datos muestran que los distritos con rentas medias más bajas tienden a tener niveles educativos más bajos, esto podría sugerir que la falta de educación está contribuyendo a los problemas económicos de estos distritos. En cualquier caso, los datos recogidos y analizados a través de este proyecto serán de gran interés para fundamentar decisiones de índole política, así como las inversiones destinadas a mejorar la educación y la economía del ayuntamiento de Madrid.

### What is the expected size of the data that you intend to generate or re-use?

Los datos que se reutilizarán en el proyecto no serán de gran tamaño. Por ejemplo, el archivo del INE que contiene la información sobre la renta neta media en los 21 distritos del Ayuntamiento de Madrid durante los años 2016-2019 tiene un tamaño de 68 kB, mientras que los archivos obtenidos del portal de datos abiertos del Ayuntamiento de Madrid con información sobre formación y centros educativos durante los años 2016-2019 tienen un tamaño total de 21 MB (incluyen variables adicionales a las de interés). En cuanto a los datos generados para el análisis, se espera que su tamaño sea reducido. En particular, se medirán 15 variables numéricas en 22 localizaciones (los 21 distritos y la ciudad en su conjunto). Suponiendo que cada valor ocupa 8 bytes, esta información ocuparía aproximadamente  $22 \times 15 \times 8 \text{ bytes} = 2.64 \text{ kB}$ . Si también se consideran las variables de tipo carácter que identifican las localizaciones (22) y las variables de estudio (15), y se asume que como máximo tendrán 50 caracteres ASCII (cada uno ocupa 1 byte), entonces a la cantidad anterior deberíamos sumar aproximadamente  $37 \times 50 \text{ bytes} = 1.85 \text{ kB}$ . En total, los datos de cada año ocuparían 4.50 kB, lo que multiplicado por los 4 años que se analizan da un total de 18 kB.

## **What is the origin/provenance of the data, either generated or re-used?**

## **To whom might your data be useful ('data utility'), outside your project?**

En general, la utilidad de los datos de este proyecto es alta, ya que la información obtenida del análisis de los datos podría servir de base para una serie de intervenciones e iniciativas destinadas a mejorar la educación y las condiciones económicas en los diferentes distritos de Madrid. Por ejemplo, los datos generados en este proyecto podrían ser de utilidad para muchos colectivos, incluidos investigadores, instituciones gubernamentales, responsables políticos y/o pedagogos que trabajan para mejorar la educación y la economía del Ayuntamiento de Madrid. Asimismo, los datos también podrían ser de interés para el público en general que desee comprender la dinámica de la educación y los ingresos de los distintos distritos de la ciudad.

## **FAIR data**

### **2.1. Making data findable, including provisions for metadata: Will data be identified by a persistent identifier?**

Sí, se utilizará un identificador persistente para los datos del proyecto, asegurando así que puedan ser encontrados, identificados y referenciados fácilmente por investigadores y/o otras partes interesadas (stakeholders). Para ello, se asignará un DOI (Digital Object Identifier) al conjunto de datos, que permitirá su localización y acceso a largo plazo.

### **2.1. Making data findable, including provisions for metadata: Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.**

Sí, se proporcionarán metadatos para facilitar el descubrimiento y/o la localización del conjunto de datos utilizado en el proyecto. Los metadatos incluirán, como mínimo, la siguiente información sobre los datos: título, creadores, palabras clave, descripción, contribuidores, fecha, tipo, formato, identificador, idioma, cobertura espacio-temporal y derechos o licencias. Se utilizarán, al menos, los esquemas de metadatos Dublin Core y DataCite para garantizar que los metadatos sean fácilmente localizables, comprensibles y accesibles para un público amplio.

### **2.1. Making data findable, including provisions for metadata: Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?**

Sí, se proporcionarán palabras clave que caractericen adecuadamente el contenido y el contexto de los datos para maximizar el potencial de reutilización. Estas palabras clave se incluirán en los metadatos que acompañan a los datos y se seleccionarán para aumentar la facilidad con la que otros investigadores y/o stakeholders pueden encontrar los datos.

### **2.1. Making data findable, including provisions for metadata: Will metadata be offered in such a way that it can be harvested and indexed?**

Sí. El conjunto de datos utilizado en el proyecto se subirá al repositorio abierto Zenodo, que se adhiere al protocolo OAI-PMH (Open Archive Initiative-Protocol for Metadata Harvesting). Se trata de un protocolo de interoperabilidad para el intercambio y la difusión de metadatos procedentes de diversos repositorios. En concreto, este protocolo permite la recolección de metadatos utilizando peticiones HTTP.

### **2.2. Making data accessible - Repository: Will the data be deposited in a trusted repository?**

Sí, como se ha indicado anteriormente, los datos se depositarán en Zenodo, que es un repositorio de datos abiertos que cumple las normas de conservación y acceso a los recursos digitales establecidas por la iniciativa OpenAIRE de la Unión Europea, una red de repositorios cuyo objetivo es proporcionar acceso abierto a la investigación financiada por la Comisión Europea.

**2.2. Making data accessible - Repository: Have you explored appropriate arrangements with the identified repository where your data will be deposited?**

Question not answered.

**2.2. Making data accessible - Repository: Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?**

Sí, Zenodo garantiza que los conjuntos de datos que se depositan en su plataforma sean asignados un identificador persistente. En concreto, Zenodo asigna un DOI (Digital Object Identifier) a cada recurso digital que se deposita en el repositorio, lo que permite acceder de manera sencilla y rápida al recurso a través de un enlace. Cuando un usuario hace clic en el DOI, se le redirige a la página de destino del recurso, donde se proporciona información sobre él y se permite al usuario acceder al contenido del mismo.

**2.2. Making data accessible - Data:**

**Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.**

Sí, el conjunto de datos estará disponible en abierto en el repositorio de Zenodo, lo que facilitará la accesibilidad tanto para la comunidad científica como para el público en general.

**2.2. Making data accessible - Data:**

**If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.**

Question not answered.

**2.2. Making data accessible - Data:**

**Will the data be accessible through a free and standardized access protocol?**

Sí, los datos utilizados en el proyecto serán accesibles a través de protocolos de acceso gratuitos y estandarizados, como formatos universales y estructurados como Excel XLSX. También se podrá acceder a los datos a través de APIs (Application Programming Interfaces, del inglés). En el caso del repositorio de Zenodo, existe la Zenodo REST API que, entre otras cosas, permite la publicación y descarga de recursos digitales.

**2.2. Making data accessible - Data:**

**If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?**

No aplica. Los datos serán accesibles para el público general sin ningún tipo de restricción tanto durante el proyecto como una vez finalizado.

**2.2. Making data accessible - Data:**

**How will the identity of the person accessing the data be ascertained?**

No aplica. No será necesario que el usuario se identifique para acceder a los datos.

**2.2. Making data accessible - Data:**

**Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?**

Question not answered.

## 2.2. Making data accessible - Metadata:

**Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?**

Sí, los metadatos se pondrán a disposición del público general bajo la licencia Creative Commons Zero (CC0), lo que significa que cualquier usuario podrá utilizarlos para cualquier fin, incluido el uso comercial, sin necesidad de obtener permiso ni pagar ningún royalty. Estos metadatos se ceden al dominio público con el objetivo de hacerlos lo más accesibles posible, reconociendo al mismo tiempo la contribución de los creadores. Además, estos metadatos incluirán el identificador persistente de tipo DOI asociado a los datos, así como un enlace que permite descargar directamente el conjunto de datos al hacer clic en él.

## 2.2. Making data accessible - Metadata:

**How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?**

Los datos estarán disponibles y/o localizables durante el máximo tiempo posible (idealmente, de forma permanente). En particular, el equipo del proyecto intentará que los datos estén disponibles para su reutilización, al menos, mientras se consideren útiles y relevantes para la comunidad científica. Si por alguna razón los datos dejan de ser accesibles, los metadatos seguirán estando disponibles y proporcionarán información sobre los datos, permitiendo así a los investigadores y/o stakeholders comprender el alcance y el contexto del conjunto de datos. Esto último estará garantizado gracias al "Plan de Preservación de Datos" elaborado para el proyecto.

## 2.2. Making data accessible - Metadata:

**Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?**

El campo de descripción de los metadatos especificará que el conjunto de datos utilizado en el proyecto se publica en formato Excel XLSX y mencionará las herramientas utilizadas para leer, curar y analizar los datos.

## 2.3. Making data interoperable:

**What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?**

El conjunto de datos utilizado en el proyecto se procesará y publicará de manera que sea lo más interoperable posible para que el resto de la comunidad pueda reutilizarlo fácilmente. Se utilizarán herramientas de software libre, como Python, que promuevan la interoperabilidad de los datos y se seguirán buenas prácticas referentes a formatos de datos y metadatos. Por ejemplo, los datos se publicarán en el formato estructurado XLSX y se garantizará que los metadatos estén, al menos, en el formato Dublin Core y DataCite. Estas prácticas buscan garantizar que los datos estén organizados y formateados de manera que sean fáciles de integrar en otras fuentes de datos.

## 2.3. Making data interoperable:

**In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?**

Question not answered.

## 2.3. Making data interoperable:

**Will your data include qualified references<sup>[1]</sup> to other data (e.g. other data from your project, or datasets from previous research)?**

**[1] A qualified reference is a cross-reference that explains its intent. For example, *X is regulator of Y* is a much more qualified reference than *X is associated with Y*, or *X see also Y*. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source:**

<https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/>

Sí. Como se ha comentado anteriormente, el conjunto de datos empleado en el proyecto reutiliza datos de dos repositorios gubernamentales, a saber: el INE y el portal de datos en abierto del Ayuntamiento de Madrid. Estas fuentes de datos originales se referenciarán de la manera oportuna en el campo de metadatos de descripción y también se incluirán en el campo de metadatos de contribuidores.

#### **2.4. Increase data re-use:**

**How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?**

Con el objetivo de garantizar la transparencia y reproducibilidad del análisis, adicionalmente al conjunto de datos se subirá al repositorio de Zenodo el código de Python utilizado para leer, limpiar y analizar los datos. La memoria del proyecto también será una fuente valiosa de información sobre los datos y su estructura, en especial porque detalla las variables estudiadas y su significado.

#### **2.4. Increase data re-use:**

**Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?**

Los datos curados serán publicados en abierto bajo la Atribución 4.0 Internacional de Creative Commons (CC BY 4.0), lo que permite a cualquier usuario distribuir, adaptar y utilizar los datos libremente siempre y cuando se conceda el debido crédito al autor original y se indique cualquier modificación realizada sobre los mismos.

#### **2.4. Increase data re-use:**

**Will the data produced in the project be useable by third parties, in particular after the end of the project?**

Los datos utilizados en el proyecto podrán ser utilizados por terceros, tanto durante el proyecto como una vez finalizado, sin restricción alguna.

#### **2.4. Increase data re-use:**

**Will the provenance of the data be thoroughly documented using the appropriate standards?**

Question not answered.

#### **2.4. Increase data re-use:**

**Describe all relevant data quality assurance processes.**

Los datos en crudo se obtienen directamente de los repositorios del INE y del portal de datos abiertos del Ayuntamiento de Madrid, por lo que este proyecto no tiene control sobre los criterios de calidad utilizados por estas instituciones para recopilar los datos. Sin embargo, tanto el INE como el Ayuntamiento de Madrid están sujetos al Código de Buenas Prácticas de las Estadísticas Europeas, lo que garantiza que las estadísticas que publican sean independientes, fiables y de alta calidad.

Por otro lado, en la fase de preprocesado y curación de los datos, los administradores de datos del proyecto llevarán a cabo una fase exhaustiva de depuración de los datos para garantizar su consistencia. Esto incluirá la detección y corrección de valores anómalos, la imputación de valores faltantes, la eliminación de valores duplicados, la corrección de errores tipográficos y cualquier otra tarea necesaria para asegurar la calidad de los datos.

#### **2.4. Increase data re-use:**

**Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.**

Question not answered.

## Other research outputs

**In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).**

Question not answered.

**Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.**

Question not answered.

## Allocation of resources

**What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.) ?**

Los costes para cumplir con los principios FAIR en el proyecto son principalmente los relacionados con los recursos humanos, ya que los datos reutilizados son gratuitos y se analizarán con herramientas de software libre y se publicarán en un repositorio también gratuito. Los costes incluyen las remuneraciones de los miembros del equipo que se encargarán de limpiar, curar y publicar los datos y metadatos para que sean fácilmente localizables, accesibles, interoperables y reutilizables.

**How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)**

Los costes destinados a garantizar que los datos empleados en el proyecto sean FAIR quedarán cubiertos por el presupuesto del proyecto, las subvenciones y el dinero procedente de fuentes externas. En particular, los gastos relacionados con el acceso abierto a los datos podrán ser subvencionables por la Comisión Europea, en el marco de los proyectos Horizon 2020, si se ajustan a las directrices establecidas en el acuerdo de subvención.

**Who will be responsible for data management in your project?**

El proyecto REFOMAD tiene dos administradores de datos principales, que son Sergio Bolívar Gómez y Juan José Velasco Horcajada. No obstante, las responsabilidades individuales de cada uno de los miembros del equipo en materia de gestión de datos pueden consultarse en la división de paquetes de trabajo.

**How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?**

Question not answered.

## Data security

**What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?**

Además del almacenamiento de los datos en un repositorio seguro, se tomarán otras medidas para garantizar la seguridad de los mismos. Entre ellas, destacan: crear copias de seguridad tanto en la nube como en un disco duro externo de forma frecuente, utilizar un sistema de control de versiones y guardar todo el material generado en el proyecto en un disco duro externo una vez finalizado el mismo. Para una información más detallada, se puede consultar el plan de preservación de datos elaborado para el proyecto

## **Ethics**

**Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).**

Teniendo en cuenta que los datos son de dominio público y que no se manejan datos sensibles, no hay ninguna cuestión ética y/o legal que pueda tener un impacto negativo a la hora de compartir los datos.

**Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?**

Question not answered.

## **Other issues**

**Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?**

Question not answered.

7