A close up of a piece of paper

Description automatically generated

|  |
| --- |
|  |

|  |  |
| --- | --- |
| Deliverable Number | D1.1 |
| Lead Beneficiary | UNIPR |
| Authors | Mario Veneziani, Federico Antonioli, Giorgia Eranio, Carlos Leyva, Pablo Báez, Álvaro Fernández |
| Work package | WP1 |
| Delivery Date | M09 → M11 |
| Dissemination Level | Public |

D1.1 – Metodología y Ontología utilizada para la caracterización de las fuentes de datos.



Document Information

|  |  |
| --- | --- |
| Project title | Agent-based support tool for the development of agriculture policies |
| Project acronym | AGRICORE |
| Project call | H2020-RUR-04-2018-2019 |
| Grant number | 816078 |
| Project duration | 1.09.2019-31.8.2023 (48 months) |
| Deliverable Authors | Mario Veneziani, Federico Antonioli, Giorgia Eranio, Carlos Leyva, Pablo Báez, Álvaro Fernández |
| Deliverable Reviewers | Jaromir Krzyszczak, Ahmet Ali KOC |

Version History

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Description | Organisation | Date |
| 1.0 | Deliverable complete | UNIPR, IDENER | 31 July 2020 |
| 1.0 [SP | [Spanish]  summary translation | CAAND | 03 May 2022 |

Resumen ejecutivo

Este entregable presenta cuál ha sido la metodología definida dentro del proyecto AGRICORE para caracterizar las fuentes de datos utilizadas para realizar el análisis de investigación en el sector agrícola. Esta metodología se ha desarrollado como parte del primer paquete de trabajo definido en el proyecto AGRICORE. AGRICORE es un proyecto de investigación que propone una forma innovadora de aplicar la modelización basada en agentes para mejorar la capacidad de los responsables políticos de evaluar el impacto de las medidas relacionadas con la agricultura dentro y fuera del marco de la Política Agrícola Común. Este proyecto fue financiado por la Comisión Europea como resultado de la convocatoria RUR-04-2018, que forma parte del programa H2020.

El presente documento comienza proporcionando una introducción al lector sobre el marco de este desarrollo, abarcando el proyecto AGRICORE y centrándose en las partes relevantes del mismo, relacionadas principalmente con el uso de datos para realizar análisis de evaluación de impacto. A continuación, se explica el uso de ontologías como parte de la metodología de caracterización y se ofrece un análisis de los trabajos de investigación ya existentes en este ámbito. A continuación, se detalla la metodología propuesta, incluyendo también el proceso seguido para diseñarla. Como parte de esta metodología, los socios de AGRICORE han desarrollado una extensión del estándar del Catálogo de Datos de Aplicación (DCAT-AP) que servirá de base para recopilar la información necesaria durante el proceso de caracterización. Esta sección está disponible como parte de este entregable, pero también se ha publicado como documento independiente. Por último, se ofrecen algunas conclusiones sobre la caracterización y el mapeo (y las necesidades de éste) de las fuentes de datos.

Es importante señalar que, aunque este documento se ha desarrollado en el marco del proyecto AGRICORE, los socios participantes han pretendido ampliar el uso de la metodología propuesta. Como objetivo final de este paquete de trabajo, la herramienta de indexación de la UE propuesta (ahora rebautizada como herramienta de indexación de datos de investigación agrícola (ARDIT)) pretende servir como punto de entrada central para localizar conjuntos de datos útiles para la investigación agrícola. En consecuencia, la metodología aquí presentada se utilizará para caracterizar el conjunto de conjuntos de datos identificados por el consorcio

# Conclusiones

La entrega 1.1 ha aportado pruebas sobre el uso de ontologías para capturar y sistematizar ricos dominios de conocimiento, como la agricultura. El uso de una ontología para la gestión eficiente de mucha información puede ser más importante cuando los modelos de gran escala/complejos requieren hacer uso de más de una fuente de datos. Esto es especialmente cierto cuando es necesario garantizar que una o unas pocas variables estén a disposición del investigador para hacer funcionar el modelo o los modelos.

Debido a la falta de ontologías existentes capaces de identificar la información relevante sobre las variables clave contenidas en (un) conjunto(s) de datos, y las relaciones entre ellas, este entregable ha documentado el desarrollo de la ontología AGRICORE DCAT-AP 2.0. Se trata de una ampliación del modelo de datos DCAT-AP 2.0 que se ha llevado a cabo principalmente añadiendo nuevas clases -y relaciones- capaces de captar las necesidades de conocer las características (de las variables contenidas en) los conjuntos de datos que podrían emplearse para el análisis de los impactos de la(s) política(s) en la agricultura. Los modelizadores que participan en el proyecto AGRICORE expresaron estas necesidades de conocimiento durante la ejecución de la Tarea 1.1 ayudando a preparar una plantilla para la caracterización de los conjuntos de datos, que podría captar las características de los conjuntos de datos estadísticos y georreferenciados por igual. La plantilla permitirá recoger la información pertinente sobre los conjuntos de datos de interés para un investigador, sin tener acceso a los datos. Dado que la información recogida mediante la plantilla de caracterización será manipulada y gestionada por la ontología AGRICORE DCAT-AP 2.0, ambas se han desarrollado de forma sinérgica. La plantilla de caracterización constituirá la herramienta para llevar a cabo la caracterización de los conjuntos de datos que se realizarán en las tareas 1.2 a 1.6. No obstante, al comprobar que la plantilla puede quedarse corta a la hora de captar características importantes de los conjuntos de datos que se van a caracterizar, los socios del proyecto AGRICORE actualizarán tanto la plantilla como la ontología que organiza y gestiona la información.

La concienciación y el conocimiento por parte de los investigadores de qué variables, y sus características, están disponibles en qué conjuntos de datos son cruciales para que los esfuerzos de investigación sean eficaces. Por lo tanto, la plantilla de caracterización desarrollada simultáneamente y la ontología AGRICORE DCAT-AP 2.0 permitirán capturar muchos detalles sobre un gran número de conjuntos de datos de potencial interés para la comunidad investigadora. Esta información se almacenará en el ARDIT (que se entregará en la tarea 1.8) y se podrá buscar -también mediante servicios semánticos (que se proporcionarán en la tarea 4.4)- en la Internet pública, gracias a la ontología AGRICORE DCAT-AP 2.0. Las consultas de búsqueda devolverán información sobre los conjuntos de datos y, lo que es más importante, sobre las variables que contienen, que pueden emplearse para ejecutar modelos de análisis de los efectos de las políticas (reformas) en la agricultura. Se espera que el ARDIT se convierta en una herramienta de referencia para identificar los conjuntos de datos relevantes para los esfuerzos de modelización de la comunidad investigadora.