

|  |
| --- |
|  |

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

D1.1 – Una Metodologia Standardizzata ed un Gruppo di Ontologie per la Caratterizzazione delle Fonti di Dati



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 Bàez, Alvaro Fernandez |
| Deliverable Reviewers | Jaromir Krzyszczak, Ahmet Ali Koc |

Version History

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Description | Organisation | Date |
| 0.1 | First template version | UNIPR | 31 March 2020 |
| 0.2 | Revision of document structure | IDENER | 7 April 2020 |
| 0.3 | Background information added | UNIPR | 16 April 2020 |
| 0.4 | Description of Ontologies and previous efforts | UNIPR, STAM | 5 May 2020 |
| 0.5 | Restructuring of the document. Inclusion of DCAT- AP extension | STAM | 8 May 2020 |
| 0.6 | Content completion | UNIPR | 28 May 2020 |
| 0.7 | DCAT-AP extension details | STAM | 15 June 2020 |
| 0.8 | Methodology description | UNIPR, STAM | 22 June 2020 |
| 0.9 | Coordinator review, suggestions for updates | IDENER | 30 June 2020 |
| 0.10 | Further explanation of the methodology | IDENER, UNIPR | 10 July 2020 |
| 0.11 | Extension of DCAT- AP AGRICORE extension | STAM | 15 July 2020 |
| 0.12 | Intensive review and extension of the content and format | IDENER, UNIPR, STAM | 10 July 2020 – 30 July 2020 |
| 1.0 | First release version - Delivered to the EC | UNIPR, IDENER | 31 July 2020 |

RIASSUNTO

Questo documento propone la metodologia definita all'interno del progetto AGRICORE per caratterizzare le fonti di dati utili a completare analisi in ambito agricolo. Questa metodologia è stata sviluppata quale attività iniziale del progetto AGRICORE. AGRICORE è un progetto di ricerca che propone un modo innovativo di applicare i modelli *“ad agenti”* per migliorare le capacità dei decisori pubblici di valutare l'impatto delle misure di politica agricola, siano esse incluse nella o separate dalla Politica Agricola Comune (PAC). Questo progetto è stato finanziato dalla Commissione Europea come risultato del bando RUR-04-2018, parte del programma H2020.

Il presente documento inizia fornendo al lettore un'introduzione su tutti gli aspetti di questa attività di ricerca e sviluppo, presentando il progetto AGRICORE e concentrandosi sulle parti rilevanti di esso, principalmente legate all'uso dei dati per eseguire analisi di valutazione d'impatto. In seguito, il documento spiega l'uso delle ontologie nell’ambito della metodologia di caratterizzazione delle fonti di dati e presenta un'analisi critica degli studi già completati in questo ambito. Successivamente, si espone in maniera dettagliata la metodologia proposta, compreso l’iter per la sua progettazione. I membri del Consorzio di Ricerca che sta implementando il progetto AGRICORE hanno sviluppato un'estensione dello standard Data Catalogue Application Profile (DCAT-AP) che servirà come base per compilare le informazioni richieste durante il processo di caratterizzazione. Questa sezione è disponibile come parte di questo documento ma è anche rilasciata come documento separato. Infine, le conclusioni riguardano l’utilità e la necessità della caratterizzazione e della mappatura delle fonti di dati.

È importante sottolineare che, anche se questo documento è stato sviluppato per la realizzazione della attività previste nell’ambito del progetto AGRICORE, membri del Consorzio di Ricerca che sta implementando il progetto hanno ricercato un uso più esteso della metodologia proposta. Obiettivo finale di queste attività preliminari è la realizzazione dell'EU Index Tool proposto (ora rinominato Agricultural Research Data Index Tool (ARDIT)) che si propone di essere lo strumento preferito per conoscere, e potenzialmente ottenere, fonti di dati utili alla ricerca agricola. Pertanto, la metodologia qui presentata sarà utilizzata per caratterizzare l'insieme dei dataset identificati dal consorzio AGRICORE, non limitando l'analisi a quelli utilizzati nel progetto.

CONCLUSIONI

Il documento 1.1 ha fornito prove sull'uso delle ontologie per catturare e sistematizzare ricchi domini di conoscenza, come l'agricoltura. L'uso di un'ontologia per la gestione efficiente di molte informazioni può essere più importante quando modelli su larga scala e complessi richiedono l’uso di una molteplicità di fonti di dati. Questo è particolarmente rilevante quando è necessario assicurare che una o alcune variabili siano disponibili al ricercatore per far funzionare il modello (o i modelli).

A causa della mancanza di ontologie esistenti in grado di identificare le informazioni rilevanti sulle variabili chiave contenute in una o più fonti di dati, e le relazioni tra di esse, questo documento ha riportato lo sviluppo dell'ontologia AGRICORE DCAT-AP 2.0. Si tratta di un'estensione del modello di dati DCAT-AP 2.0 che è stata intrapresa principalmente aggiungendo nuove classi e relazioni in grado di soddisfare la necessità di conoscere le caratteristiche delle fonti di dati (e delle variabili in esse contenute) che potrebbero essere impiegate per l'analisi degli impatti delle politiche (o delle loro riforme) sull'agricoltura. I costruttori di modelli economici coinvolti nel progetto AGRICORE hanno espresso queste esigenze di conoscenza durante l'esecuzione dell'attività 1.1, aiutando a preparare un modello per la caratterizzazione delle fonti di dati che può essere applicato sia alle fonti di dati statistici sia geo-referenziate. Il modello di caratterizzazione proposto permette di sistematizzare informazioni rilevanti sui dataset di interesse per un ricercatore, senza avere accesso ai dati. Poiché le informazioni raccolte per mezzo del modello di caratterizzazione saranno trattate e gestite dall'ontologia AGRICORE DCAT-AP 2.0, le due sono state sviluppate in modo sinergico. Il modello di caratterizzazione costituirà lo strumento per attuare la caratterizzazione delle fonti di dati da intraprendere nelle attività da 1.2 a 1.6. Tuttavia, qualora si verificasse che il modello di caratterizzazione potrebbe essere incapace di rilevare caratteristiche importanti delle fonti di dati da caratterizzare, i membri del Consorzio di Ricerca che sta implementando il progetto AGRICORE aggiorneranno sia il modello di caratterizzazione sia l'ontologia che organizza e gestisce le informazioni.

La consapevolezza e la conoscenza da parte dei ricercatori di quali variabili - e le loro caratteristiche - sono disponibili in quali fonti di dati sono determinanti per rendere efficaci gli sforzi di ricerca. Pertanto, il modello di caratterizzazione sviluppato simultaneamente all'ontologia AGRICORE DCAT-AP 2.0 permetteranno di raccogliere molti dettagli su un gran numero di fonti di dati di potenziale interesse per la comunità di ricerca. Queste informazioni saranno memorizzate nell'ARDIT (da consegnare nell'attività 1.8) e saranno fruibili - anche per mezzo di servizi semantici (da fornire nell’attività 4.4) sull'internet pubblico - grazie all'ontologia AGRICORE DCAT-AP 2.0. Le stringhe di ricerca online restituiranno informazioni riguardanti le fonti di dati e - soprattutto - le variabili in esse contenute, che possono essere impiegate per l'esecuzione di modelli per l'analisi degli impatti delle politiche (e/o loro riforme) sull'agricoltura. Si spera che ARDIT diventi uno strumento di riferimento per identificare le fonti di dati rilevanti per gli sforzi di creazione di modelli da parte della comunità di ricerca.