



**AGENT-BASED
SUPPORT TOOL FOR
THE DEVELOPMENT
OF AGRICULTURE POLICIES**

D2.2. Modul zur Extraktion von Big Data



Deliverable Number	D2.2
Lead Beneficiary	AUTH
Authors	AUTH, IDE
Work package	WP2
Delivery Date	31/08/2022
Dissemination Level	Public

www.agricore-project.eu



The Agricore project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No. 816078





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	Michail Tsagris (AUTH), Vangelis Tzouvelekas (AUTH), Dimitrios Natos (AUTH), IDENER Team
Deliverable Reviewers	IDENER Team

Version History

Version	Description	Organisation	Date
1.0	Deliverable complete	AUTH	31 Aug 2022
1.0 GER	German summary translation	AXIA	05 Sep 2022

Executive Summary

AGRICORE ist ein von der Europäischen Kommission im Rahmen der Aufforderung RUR-04-2018 finanziertes Forschungsprojekt, das Teil des H2020-Programms ist. Es schlägt einen innovativen Weg vor, agentenbasierte Modellierung anzuwenden, um die Fähigkeit der politischen Entscheidungsträger zu verbessern, die Auswirkungen von agrarbezogenen Maßnahmen innerhalb und außerhalb des Rahmens der Gemeinsamen Agrarpolitik (GAP) zu bewerten.

In diesem Beitrag wird das AGRICORE-Datenextraktionsmodul (DEM) vorgestellt, das die Lokalisierung, Extraktion und Speicherung aller Arten von Daten ermöglicht, die für die Nutzung der verschiedenen Instrumente und Module der AGRICORE-Plattform erforderlich sind. Außerdem werden die wichtigsten Arten von Informationen vorgestellt, die für die Initialisierung, Kalibrierung und Nutzung der verschiedenen AGRICORE-Module erforderlich sind, sowie die wichtigsten Datenquellen, aus denen diese Informationen extrahiert werden können.

Zu den Funktionalitäten des Datenextraktionsmoduls gehört auch die statistische Analyse der in den importierten Datensätzen enthaltenen Variablen sowie die Ermittlung von Korrelationsbeziehungen zwischen ihnen. In diesem Beitrag wird die Verarbeitungssequenz vorgestellt, die für jede einzelne Variable sowie für alle Variablen durchgeführt wird, um die gegenseitigen Abhängigkeiten zu ermitteln.

Die Ergebnisse dieser Operationen erzeugen statistische Metadaten und verbotene Richtungen zwischen den Variablen. Beide Ergebnisse sind Argumente, die auch als Eingabe für das Datenfusionsmodul dienen, das in Beitrag D2.3 vorgestellt wird.

Conclusions

In diesem Beitrag werden die Funktionen des Datenextraktionsmoduls von AGRICORE vorgestellt, einschließlich der Beschaffung von Daten von lokalen oder Cloud-Speicherorten, der grundlegenden Umwandlung und Erkundung der erhaltenen Datensätze und des Ladens dieser in das Data Warehouse. Die Verbindungen des DEM mit ARDIT und mit dem DWH selbst, die in Abschnitt 4 erläutert werden, sind von grundlegender Bedeutung für das korrekte Funktionieren des Prozesses der Datensatzsuche und -erfassung.

Der generische Fluss der Datensätze von ihrer Lokalisierung durch ARDIT bis zu ihrem letztendlichen Zusammenführungsprozess, um angereicherte abgeleitete Datensätze zu erzeugen, wurde ebenfalls vorgestellt. Zu diesem Zweck wurde ein klimatologischer Indikator (SPEI12), dessen räumliche Auflösung geändert werden muss ($\{Lat, Lon\}$ zu NUTS3), um sich an das vom biophysikalischen Modul von AGRICORE geforderte Format anzupassen, als Beispiel verwendet.

In diesem Dokument werden auch die Arten von Informationen vorgestellt, die für die Initialisierung und Konfiguration der Module, aus denen das AGRICORE-Tool besteht, erforderlich sind. Es werden auch kurz die wichtigsten Datenquellen beschrieben, aus denen die erforderlichen Daten normalerweise extrahiert werden.

Die DEM ist nicht nur ein notwendiges Element für den Betrieb der übrigen Module, sondern sie ist auch von grundlegender Bedeutung für den Prozess der Erzeugung synthetischer Populationen, da sie (zusammen mit ARDIT) das Werkzeug sind, das es ermöglicht, die Rohdaten zu erhalten, aus denen das DFM das mathematische Objekt (Bayes'sches Netz) erzeugt, das es ermöglicht, den Attributen jedes Agenten numerische Werte zuzuweisen.

Die nächsten Schritte in diesem Task T2.2 sind die Integration des DEM-Codes in den DWH-Einsatz und die Funktionstests des Prozesses zur Erzeugung synthetischer Populationen, sobald das SPG-Modul fertiggestellt ist (zukünftiges Ergebnis D2.4).