



D4Dairy

**Digitalisation, Data integration,
Detection and Decision support
in Dairying**

Programme: COMET – Competence
Centers for Excellent Technologies

Programme line: COMET-Project

Type of project: D4Dairy,
01.10.2018 – 30.09.2022, strategic,
multi-firm

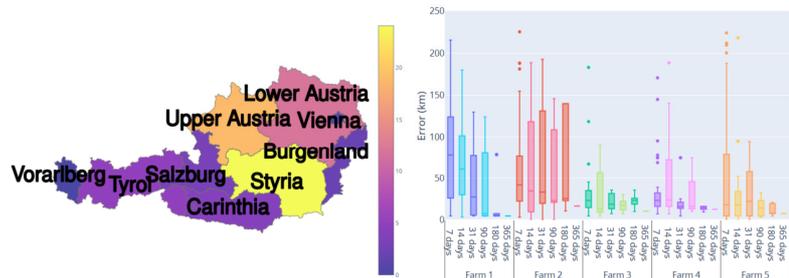


Abbildung 1: (a) Dichte der landwirtschaftlichen Betriebe Bundesland in Österreich im Datensatz. (b) Lokalisierungsfehler von ausgewählten Tieren aus landwirtschaftlichen Betrieben in mehreren österreichischen Bundesländern. F. Papst, N. Stricker, R. Entezari, O. Saukh, [To Share or Not to Share: On Location Privacy in IoT Sensor Data](#), IEEE/ACM IoTDI 2022.

DATENINTGRATION MIT DATENSCHUTZ, KONSISTENZ UND QUALITÄTSKONTROLLEN

D4DAIRY INTEGRIERT BETRIEBSBEZOGENE DATEN, AUS MEHREREN QUELLEN UNTER BEACHTUNG VON DATENSCHUTZ UND DATENQUALITÄT

Detaillierte und wertvolle Daten zum landwirtschaftlichen Tagesgeschäft, zur Effizienz und zur Tiergesundheit befinden sich in getrennten Datensilos, was es schwierig macht, genaue Vorhersagemodelle zu entwerfen und diese in hochwertige Entscheidungswerkzeuge zu integrieren. Die innerhalb von D4Dairy entwickelte Lösung verbindet und integriert alle projektbezogenen Subsysteme und Datenquellen in einer einzigen Datenbank namens D4Dairy Data. Es umfasst jetzt Datenströme von RDV, hochauflösende Sensordaten von Industriepartnern, automatischen Melksystemen und Stallklima. D4Dairy Data läuft in einem Rechenzentrum zur weiteren Verwendung über die D4Dairy-Projektdauer hinaus. Das System implementiert Datenanonymisierung, Datenfusion und statistische Datenqualitätsprüfungen. Obwohl das Kernsystem aus Standardkomponenten aufgebaut ist, integriert es neuartige Algorithmen, die das System speziell auf die Milchwirtschaft zugeschnitten machen.

Datenschutz für standortbezogene Daten. Das Teilen von Daten ist entscheidend für die Zusammenstellung hochwertiger Datensätze zur Verbesserung von Vorhersagemodellen. Es gibt jedoch auch berechtigte Bedenken bezüglich des Datenschutzes. Öffentliche und geteilte Daten werden in der Regel pseudonymisiert, also alle eindeutigen Identifikatoren, wie z. B. Identifikationsnummern von Betrieben, deren Standorten und Namen werden entfernt. Die Forschung zeigte jedoch, dass das bloße Entfernen dieser Identifikatoren nicht ausreicht: Die Daten selbst enthalten Informationen über den Datenlieferanten. Ein Linkage-Angriff kombiniert zwei Datensätze und verknüpft ähnliche oder identische Datensätze miteinander. Standortdatenschutz ist besonders wichtig, wenn Betriebe Sensordaten teilen. Auch wenn Ortsinformationen nicht explizit geteilt werden, ist es dennoch möglich, mit hoher Sicherheit aus den Daten auf den Standort zu schließen. Dies kann durch

SUCCESS STORY

die Kombination Sensordaten mit öffentlich verfügbaren Wetterdaten erreicht werden. Wir haben gezeigt, dass die Daten von Aktivitätssensoren verwendet werden können, um ein Tier innerhalb eines Landes mithilfe eines Linkage Angriffs zu lokalisieren (siehe Abbildung 1). Wir haben auch einen Datenschutzmechanismus implementiert, um solche Linkage Angriffe auf geteilten Sensordaten zu verhindern, indem wir die Wetterabhängigkeit in den Daten durch maschinelles Lernen aufweichen.

Datenqualität. Zur Datenvalidierung und Datenqualitätssicherung haben wir ein statistisches Sensordatenverarbeitungs-Framework entworfen, das die Co-Abhängigkeit zwischen Datenqualität und Modellrobustheit nutzt, um Leistungseinbußen in Vorhersagemodellen zu erkennen. Wir haben gezeigt, dass Verschiebung der Verteilung der Eingabedaten die Qualität des Modells beeinflussen, und einen Indikator vorgestellt, der solche Verschiebungen im laufenden Betrieb erkennen kann. Das Framework ermöglicht es, die Vorhersagequalität für Lahmheits-Prognosen um 62 % zu verbessern.

Datenkonsistenz und Harmonisierung. D4Dairy Data prüft auf Datenkonsistenz und harmonisiert Daten von Industriepartnern (siehe Abbildung 2). Innerhalb des Projekts wurde

das System verwendet, um Forschern eine konsistente Datensicht der von verschiedenen Quellen gesammelten Daten in der gewünschten zeitlichen Auflösung bereitzustellen.

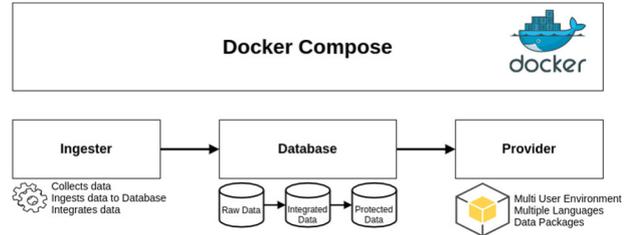


Abbildung 2: Die Architektur unseres Systems besteht aus drei Komponenten: (1) Modul zum Sammeln und Einspeisen der Daten (2) Datenbank zur Datenspeicherung (3) Data-Science-Plattform. F. Papst, K. Schodl, O. Saukh, [Exploring Co-dependency of IoT Data Quality and Model Robustness in Precision Cattle Farming](#), ACM SenSys 2021.

Konklusion. Obwohl D4Dairy aus Standardkomponenten besteht, enthält es mehrere Funktionen, z.B. Funktionen für Datenschutz oder Qualitätssicherung, die es ermöglichen wertvolle Erkenntnisse für Landwirte und die Milchwirtschaft zu liefern. Dies macht D4Dairy Data besonders, da es, verglichen zu anderen Systemen für den Datenaustausch eine Qualitätskontrolle der Daten vornimmt.

Project coordination (Story)

Prof. Olga Saukh
 Institut für Technische Informatik
 TU Graz
 saukh@tu-graz.at

D4Dairy / COMET-Project

ZuchtData EDV-Dienstleistungen GmbH
 Dresdner Straße 89/B1/18, 1200 Wien
 forschung@zuchtdata.at
 www.d4dairy.com

Project partners



This success story was provided by the consortium leader and by the mentioned project partners for the purpose of being published on the FFG website. D4Dairy is a COMET-Project within the COMET – Competence Centers for Excellent Technologies Program and funded by BMK, BMDW, Vienna and Lower Austria. The COMET Program is managed by FFG. Further information on COMET: www.ffg.at/comet

 Federal Ministry
 Republic of Austria
 Climate Action, Environment,
 Energy, Mobility,
 Innovation and Technology

 Federal Ministry
 Republic of Austria
 Digital and
 Economic Affairs

Austrian Research Promotion Agency
 Sensengasse 1, A-1090 Vienna
 P +43 (0) 5 77 55 - 0
 office@ffg.at
 www.ffg.at