

Informationen effektiv verknüpfen: Datenintegration in der Landwirtschaft

Dipl.-Ing. Martin Wischenbart, M.Sc.
Josephinum Research, Wieselburg

08.02.2018

Moderne Technik und Sensoren im Stall wie auf dem Feld liefern heutzutage große Mengen an Daten. Deren automatisierter Austausch zwischen verschiedenen Systemen ist allerdings noch immer mangelhaft, sodass Auswertungen erschwert werden und Mehrfacheingaben notwendig sind. Im Folgenden wollen wir anhand von Beispielen die Grundlagen und Prinzipien der Datenintegration diskutieren und aktuelle Entwicklungen zur Vereinfachung des Datenaustauschs vorstellen.

Zur Einleitung zwei Beispiele: Ein *Rinderbetrieb* führt seinen Tierbestand in mehreren Systemen: in Melksystem, Transponderfütterung und Herdenbuch des Kontrollverbandes für Milchkühe, sowie für alle Rinder das behördliche eAMA RinderNet und Tier-Gruppen für die Lockfütterung am Futtertisch. Wird beispielsweise eine Milchkuh verkauft oder umgestallt, so müssen Änderungen mehreren Systemen eintragen werden, anstatt in nur das zentrale integrierte Herdenmanagementsystem. Ein *Ackerbaubetrieb* besitzt einen modernen Traktor mit Satellitennavigation und Terminal. Für Fruchtfolge- und Düngeplanung erwägt er die Anschaffung einer modernen Ackerschlagkartei-Software. Dabei möchte er sicherstellen, dass am Terminal aufgezeichnete Bewegungsdaten auch zur Protokollierung von Tätigkeiten und für Auswertungen (z. B. Berechnung der Flächenleistung) genutzt werden können. Außerdem möchte er für die erste Erfassung aller Feldstücke seine Schlaggrenzen und vergangene Nutzungen aus dem eAMA INVEKOS-GIS für Flächenerfassung automatisch importieren.

Grundlagen: Datenbanken und Heterogenitäten

Solche zusammengehörigen Daten, beispielsweise über Tiere oder Flächen in Tabellenform, nennt man eine *Datenbank*. Auf dem Beispiel-Rinderbetrieb werden mehrere Datenbanken in verschiedenen Systemen getrennt verwaltet. Die vorhandenen Tabellen und der Spaltenüberschriften beschreiben die Struktur der Daten, das sogenannte *Datenmodell*. Datenbank und Datenmodell, sowie die Art des Datenzugriffs unterscheiden sich aber zwischen den Systemen oft erheblich, obwohl sie eigentlich dieselben Kühe oder Flächen beschreiben. In der folgenden Abbildung werden einige verschiedene Arten von solchen Unterschieden, in der Fachliteratur genannt *Heterogenitäten*, beispielhaft veranschaulicht: „Kuhnr“ und „OM-Nr“ in System Nr. (1) heißen „Nr“ und „Lebensnummer“ in (2), während „cow ID“ in (3) vorkommt, aber die behördliche Ohrmarkennummer dort gar fehlt. Überdies gibt es auf Ebene der Daten Unterschiede in Datumsformat und Schreibweisen der Ohrmarkennummer. Schließlich unterscheiden sich die Methoden des Zugriffs bzw. Datenformate: manueller Export einer CSV-Datei (1), Nutzung eines Web-Service (2), oder relationale Datenbank (3).

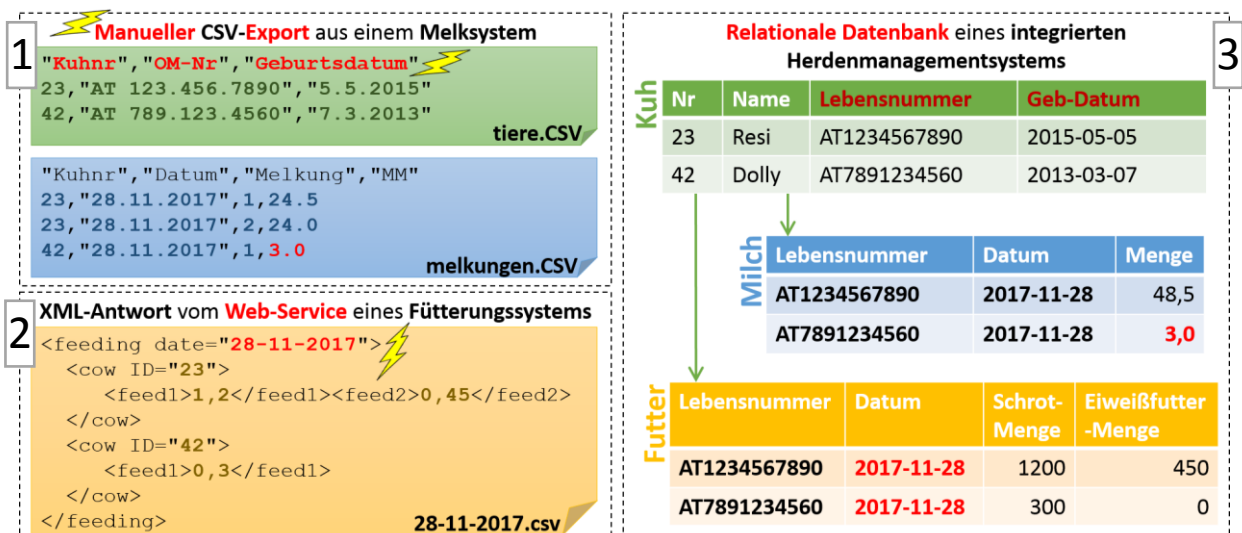


Abbildung 1: verschiedene Datenbanken und Datenmodelle sowie auftretende Heterogenitäten in drei unterschiedlichen Systemen auf einem Rinderbetrieb

Die Ursache für diese Heterogenitäten liegt meist in der *Verteilung* und in der *Unabhängigkeit* der verschiedenen Systeme bzw. Hersteller. Während ein Fütterungssystem *direkt auf einem Computer am Betrieb* läuft (Datenbank befindet sich auf dem Betrieb) wird ein Herdenmanagementsystem möglicherweise *in der Cloud* gehostet (Datenbank liegt auf fernem Rechenzentrum). Die verschiedenen Systeme werden von *Drittherstellern* geliefert, die *unabhängig* von den Betrieben und *autonom* in ihren Entscheidungen sind. Sie haben bei der *Entwicklung* der Systeme *freie Wahl* bezüglich Datenmodellen und Formaten. Im Gegensatz zur „Industrie 4.0“ sind in kleinstrukturierten österreichischen landwirtschaftlichen Betrieben daher oftmals unterschiedlichste Systeme verschiedenster unabhängiger internationaler Hersteller im Einsatz. Damit gibt es auch praktisch immer eine *Vielzahl* solcher *Heterogenitäten*.

Integration: Methoden, Schnittstellen und Architekturen

Die Überwindung von Heterogenitäten ist die Hauptaufgabe der Datenintegration. Dazu werden zuerst sogenannte *Korrespondenzen* festgelegt. Beispielsweise könnte man definieren, dass „OM-Nr“ und „Lebensnummer“ Synonyme sind, oder dass Punkte und Leerzeichen aus „AT 123.456.7890“ entfernt werden müssen um das korrekte Format zu erreichen (Übersetzung vom Quell- zum Ziel-System). In der Praxis kommt es auch vor, dass mehrere Werte zu einem einzelnen verrechnet werden müssen. Dabei liegt eine besondere Schwierigkeit im Finden einer zuverlässigen Übersetzungsfunktion, die alle möglichen Werte korrekt abdeckt, beispielsweise zur Übersetzung zwischen Geo-Koordinaten und Post-Adressen (48°07'38.4"N 15°08'39.9"E (WGS 84) vs. X: 534828.39, Y: 471352.27 (MGI / Austria Lambert) vs. „Rottenhauser Straße 1“). Während ein Landwirt üblicherweise leicht die Korrespondenz zwischen „OM-Nr“ und „Lebensnummer“ erkennen kann, müssen für die Entwicklung von sogenannten *Adaptoren* zur tatsächlichen Übersetzung Programmierer mit Experten aus dem jeweiligen Fachgebiet zusammenarbeiten. Der *Aufwand* für solche Entwicklungen wird *häufig unterschätzt* und die Personalkosten machen sie häufig besonders kostspielig. Nicht zuletzt müssen bei der Übersetzung möglicherweise noch *Datenfehler erkannt und behoben* werden (z. B. unterschiedliche Schreibweisen oder Tippfehler; Dolly vs. Doly), *fehlende Daten ergänzt* werden (z. B. fehlende Tiere; oder Lebensnummern werden nicht oder nur teilweise eingegeben), oder Werte mit unterschiedlicher Genauigkeit oder Einheiten erkannt werden (z. B. 8,5 l vs. 8550 ml).

Voraussetzung ist, dass der *Zugriff* auf die Daten aus den verschiedenen Quellen überhaupt *gewährt* wird, sowie, dass entsprechende Nutzungsrechte vorliegen. Insbesondere größere Hersteller bieten

aber teilweise keine oder nur eingeschränkte *Schnittstellen (APIs)* für den *automatisierten Datenaustausch* mit Systemen anderer Hersteller (Kundenbindung: gesamte Produktpalette von einem einzelnen Anbieter). Für die Implementierung eines Adapters ist überdies eine gute *Beschreibung der Schnittstellen* erforderlich, sodass ein firmenfremder Entwickler diese bedienen kann.

Mit der Zurverfügungstellung von *Schnittstellen* und deren *Beschreibung*, sowie mit der *Implementierung von Adaptern* ist also häufig ein großer Aufwand verbunden. Möchte man obendrein Daten mit mehreren verschiedenen Systemen austauschen, so steigt die *Anzahl* der notwendigen *Adapter* überproportional an. Während man bei 2 Systemen 2 Adapter benötigt, sind es bei 3 Systemen schon 6 Adapter und bei 8 Systemen sogar 56 Adapter (bei Austausch eines jeden Systems mit jedem anderen und in beide Richtungen), wie veranschaulicht in der linken Grafik der nächsten Abbildung weiter unten. Damit wird verständlich, dass beispielsweise ein kleiner Hersteller von Fütterungstechnik keine Anbindung zum Import von täglichen Milchmengen zu allen im Markt agierenden Herstellern von Melktechnik umsetzen kann, während andererseits einzelne Landwirte für solche vermeintlich einfachen Funktionen aber nicht die entsprechenden Honorare bezahlen können.

Standards und aktuelle Entwicklungen

Um die Datenintegration zu erleichtern kann durch *Standardisierung* von Datenmodellen und Schnittstellen die *Autonomie* der Systeme *eingeschränkt* und damit Homogenität erzwungen werden. Beispielsweise bietet ISO 17532 („ISOagriNET“) [3] einen Standard für Vernetzung und Datenaustausch zwischen Geräten in der Innenwirtschaft. Die Norm ISO 11783 („ISOBUS“) [2] definiert Stecker, Datenformate und Schnittstellen für den Datenaustausch zwischen Traktoren und Anbaugeräten. Mit einem universellen Standard kann folglich auch die Anzahl der notwendigen Adapter für den Datenaustausch verringert werden, wie in der rechten Grafik der folgenden Abbildung veranschaulicht.

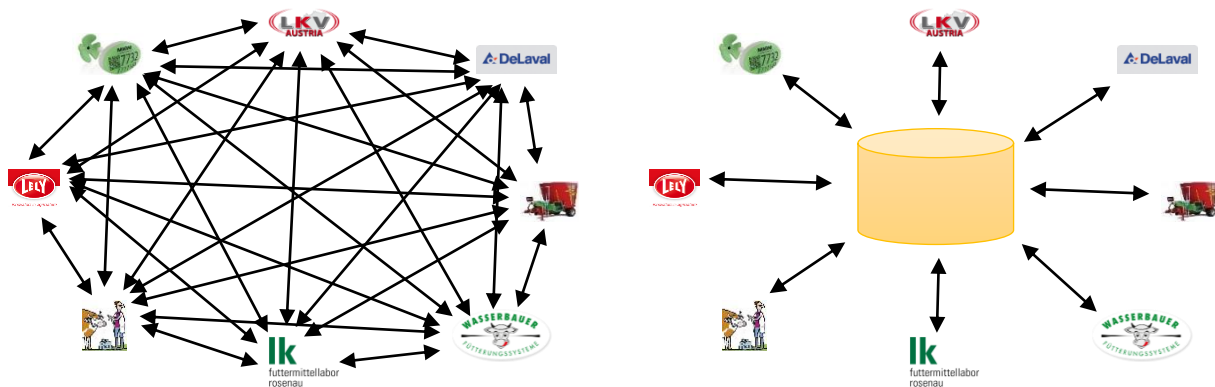


Abbildung 2: Alternative Topologien für Datenaustausch (links: direkter Austausch zwischen allen Systemen ohne gemeinsamen Standard; rechts: Austausch über ein zentrales Integrationssystem mit universell standardisiertem Datenmodell und Schnittstellen)

Während die genannten Standards auf Austauschformate fokussieren, propagiert aktuell *DKE Data* (ein europäisches Konsortium von mehreren größeren Landtechnikherstellern) eine umfangreiche kommerzielle *Datenaustauschplattform* zur Verknüpfung von landwirtschaftlicher Software auf Betrieben und Landmaschinen. Dieser „AgriRouter“ soll also die *Datenübertragung* zwischen den Systemen *bewerkstelligen*. Eine aktuelle Publikation dazu [4] erläutert den Nachrichtenaustausch über Postkästen und die verschiedenen Sicherheitsmechanismen des Systems (Authentifizierung, Zertifikate, Tokens und TANs). Im Gegensatz dazu arbeitet AgGateway (ein gemeinnütziges

Konsortium aus Nordamerika) an dem *Softwareentwicklungs-Framework „ADAPT“* [5], einer Sammlung von Software-Werkzeugen zur tatsächlichen Konvertierung von Daten zwischen verschiedenen Formaten in der Landwirtschaft.

Ob es sich um offizielle Normierungen durch internationale Konsortien handelt oder ob ein De-facto-Standard durch Institutionen oder Firmen definiert wird, neben wirtschaftlichen Interessen ist für den Endbenutzer oder Landwirt vor allem die *Akzeptanz und Verbreitung* entscheidend. Erst wenn sich die Hersteller der eingesetzten Systeme auf ein Datenaustauschformat einigen und die entsprechenden Schnittstellen anbieten, werden Datenmehrfacheingaben endgültig der Vergangenheit angehören. Auf österreichischer Ebene könnten hier die für das Jahr 2018 erwarteten API-Schnittstellen zum eAMA RinderNet und zum eAMA INVEKOS-GIS für Flächenerfassung ein richtungsweisendes Beispiel und eine Basis für standardisierte Austauschformate in der Landwirtschaft österreichweit liefern.

Über den Autor:

Dipl.-Ing. Martin Wischenbart ist Mitarbeiter bei Josephinum Research in Wieselburg. Seine Expertise liegt vor allem im Bereich "Data and Information Sciences". Im Rahmen der LFI-Bildungskampagne "Digitalisierung in der Land- und Forstwirtschaft" schreibt er über Integration und Verknüpfung von Daten bei landwirtschaftlichen Prozessen.

Mehr über die Bildungskampagne erfahren Sie unter <http://www.lkdigital.at/>

Literaturverweise

- [1] Ulf Leser, Felix Naumann, *Informationsintegration: Architekturen und Methoden zur Integration verteilter und heterogener Datenquellen*, ISBN 978-3-89864-400-6, dpunkt.verlag GmbH, 2007.
- [2] AEF, *ISOBUS*; <http://www.aef-online.org/the-aef/isobus.html>, letzter Zugriff am 29.1.2017
- [3] Bauförderung Landwirtschaft e.V., *ISOagriNET*, <http://www.isoagrinet.org/de/>, letzter Zugriff am 29.1.2017
- [4] Johannes Sonnen, Jens Möller, *The Technical Concept of a Manufacturer-Independent Web-Based Data Exchange Platform for the Agricultural Sector*, LAND.TECHNIK AgEng 2017, ISBN 978-3-18-092300-0, VDI, 2017
- [5] ADAPT Framework; <http://adaptframework.org/>, letzter Zugriff am 29.1.2017