

scch { }

**Software Competence
Center Hagenberg GmbH**

Programm: COMET -
Competence Centers for
Excellent Technologies

Förderlinie: COMET-
Zentrum (K1)

Projekttyp: 2023-2026



DER SEMANTISCHE DATENQUALITÄTS-WORKFLOW

Von starren Kennzahlen zu intelligenten KPIs

Gemeinsam mit der Robert Bosch AG haben wir einen semantischen Datenqualitäts-Workflow entwickelt, der die Messung der Datenqualität auf Basis von Domänenwissen ermöglicht, das in Form eines Wissensgraphen dargestellt wird.

Moderne Produktions- und Lieferketten sind in intelligente Infrastrukturen eingebettet und bilden ein riesiges, miteinander verbundenes Daten- und Informationsnetzwerk. Wissensgraphen (KGs) haben sich als leistungsstarke Technologie etabliert, um diese Komplexität zu bewältigen. Sie nutzen Ontologien, um Rohdaten mit einem reichhaltigen, domänen-spezifischen Kontext anzureichern.

Dieser datenzentrierte Ansatz hat jedoch eine kritische Schlüsselkomponente: die Datenqualität. Das Prinzip „Garbage in, Garbage out“ war noch nie so relevant wie heute. Sind die Daten unvollständig, inkonsistent oder ungenau, führen auch die darauf basierenden Entscheidungen zu Fehlern – mit Folgen

wie ineffizienter Produktion, Compliance-Verstößen und hohen Kosten.

Die Herausforderung liegt darin, dass sich Datenqualität nicht mit einem universellen Ansatz lösen lässt, da sie kontextabhängig ist. Sensorwerte, die alle zehn Sekunden aktualisiert werden, können für bestimmte Anwendungsfälle als „vollständig“ gelten (z.B. nicht kritische Überwachung), für andere jedoch als „unvollständig“ (z.B. Echtzeit-Sicherheitskontrolle).

Bisherige Ansätze zur Lösung dieses Kontextproblems nutzten Ontologien, um Datenqualitäts-Metriken zu definieren [1], wie beispielsweise die Data Quality Definition Ontology (DQD). Dieser Ansatz war ein wichtiger erster Schritt, hatte jedoch eine grundlegende Schwäche: Entwickler mussten den Kontext (das „Wie“ der Überprüfung) direkt in die Metrik fest codieren.

Die Definition einer Vollständigkeitsprüfung für zwei ähnliche Sensorströme, beispielsweise für Temperatur- und Drucksensoren, würde zu zwei nahezu identischen Metriken führen. Dieser Ansatz ist starr, zeitaufwändig und schränkt die Wiederverwendbarkeit von Metriken stark ein, was einen Engpass bei der Skalierung der Datenqualitätssicherung verursacht.

Anstelle solcher starren Definitionen führt DQD 2.0 eine neuartige Methode für *dynamische, wissensgraphbasierte Parameter* ein. Eine Datenqualitätsmetrik wird nur einmal definiert, der Kontext (das „Wie“ der Überprüfung) wird als flexibler Parameter zur Laufzeit aus dem Wissensgraphen abgerufen und an die Metrik übergeben [2, 3]. Unser patentierter Datenqualitäts-Workflow mit DQD 2.0 führt drei Aufgaben durch:

1. **Definition:** Durch eine Ontologie werden Datenqualitäts-Metriken flexibel und wiederverwendbar definiert.
2. **Ausführung:** Datenqualitäts-Metriken werden auf Rohdaten evaluiert, die entweder als tabellarische Daten oder als Wissensgraphen vorliegen.
3. **Annotation:** Die Ergebnisse der Datenqualitätsprüfung werden in einem Wissensgraphen gespeichert und direkt mit den überprüften Daten verknüpft.

Auswirkungen und Effekte

Massive Wiederverwendbarkeit und Skalierbarkeit. Ein Qualitätssicherungsteam kann eine einmalig definierte Datenqualitäts-Metrik dynamisch auf eine beliebige Anzahl unterschiedlicher Kontexte anwenden. Dadurch entfallen die manuelle Erstellung und Pflege zahlreicher redundanter Definitionen, was Zeit spart und das Risiko für menschliche Fehler reduziert.

Die zentrale Innovation unseres Workflows liegt in der klaren Trennung zwischen Datenqualitäts-Metriken und dem erforderlichen Domänenwissen.

Datenqualität wird sichtbar. Ein wichtiger Aspekt ist die Speicherung der Ergebnisse der Datenqualitätsprüfung im Wissensgraphen (siehe Abb. 1). Bisher waren die Ergebnisse oft in separaten Berichten oder Datenbanken isoliert und nicht direkt mit den überprüften Daten verknüpft. Die Ergebnisse der Datenqualitätsprüfung sind nun direkt mit den Daten verbunden, wodurch sie leicht auffindbar, analysierbar und nutzbar sind. Dies ermöglicht die Erstellung von Echtzeitdashboards, automatisierten Warnmeldungen oder einen kontinuierlichen Verbesserungszyklus im Daten-Governance.

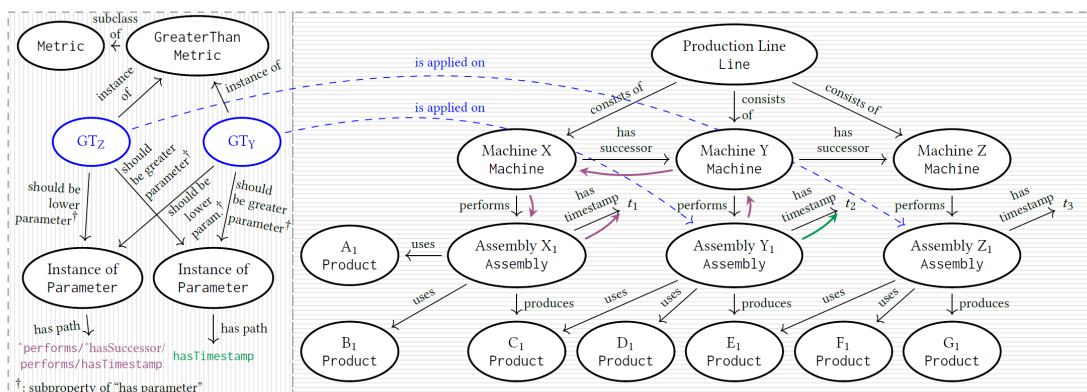


Abbildung 1: Ergebnis der semantischen Datenqualitätsprüfung als Wissensgraph dargestellt

Related Publications

[1] J. Schrott, R. Meindl, C. Lettner, W. Wöß, and L. Ehrlinger, DQD: The Data Quality Definition Ontology, in *Metadata and Semantic Research*, vol. 2048, E. Garoufallou and F. Sartori, Eds., in Communications in Computer and Information Science, vol. 2048. , Cham: Springer Nature Switzerland, 2024, pp. 291–297. doi: 10.1007/978-3-031-65990-4_27.

[2] J. Schrott, R. Meindl, C. Lettner, S. Hammer, and M. Leitner, Dynamic Knowledge Graph-based Measurement of Data Quality, in *Proceedings of the VLDB Endowment*. ISSN, [Online]. Available: https://www.vldb.org/2025/Workshops/VLDB-Workshops-2025/QDB/QDB25_3.pdf

[3] Patent Application DE102024129324A1: *Method and device for diagnosing or monitoring the (data) quality of a production plant, its production units and products*;
<https://patents.google.com/patent/DE102024129324A1/en>

Project Partner

- Software Competence Center Hagenberg
- Robert Bosch AG

Contact

Mag. Christian Lettner
Senior Research Project Manager Data Science
+43 50 343 837
christian.lettner@scch.at

Diese Success Story wurde von der Software Competence Center Hagenberg GmbH und den genannten Projektpartnern zur Veröffentlichung freigegeben. Das Software Competence Center Hagenberg wird im Rahmen von COMET – Competence Centers for Excellent Technologies durch BMK, BMDW und Land Oberösterreich gefördert. Das Programm COMET wird durch die FFG abgewickelt. Weitere Informationen zu COMET: www.ffg.at/comet