

## WORKSHOP

# ANFORDERUNGEN AN FAIRE DATENQUALITÄTSINDIKATOREN

Draeger, Christian, Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE)

Tute, Erik, Peter L. Reichertz Institut für Medizinische Informatik der TU Braunschweig und der Medizinischen Hochschule Hannover

Wodke, Judith, Universitätsmedizin Greifswald, Institut für Community Medicine, Abt. Medizininformatik Schmidt,

Carsten Oliver, Universitätsmedizin Greifswald, Institut für Community Medicine, Abt. Medizininformatik

Löbe, Matthias, Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE)

Ziel der Medizininformatik-Initiative (MII) ist es, dass Daten der Gesundheitsversorgung der Forschung zugänglich gemacht werden. Dazu werden Daten an den Datenintegrationszentren (DIZen) gesammelt und bereitgestellt. Die FHIR-Profile des sog. Kerndatensatzes bilden dabei das Common Data Model der MII. In regelmäßigen Projectathons wird der Fortschritt dieser Bemühungen evaluiert.

Da Datenqualitätssicherung bei der Sekundärnutzung der Versorgungsdaten in medizinischen Use Cases eine Grundvoraussetzung für verwertbare, wissenschaftlichen Standards genügende Forschungsergebnisse darstellt, wurde in den letzten Projectathons Datenqualität (DQ) explizit adressiert. Dabei wurde ein Framework für die Datenqualitätssicherung entworfen [1] sowie erste Datenqualitätsanalysen auf den Daten des MII Kerndatensatzes durchgeführt [2]. Grundlage aller DQ-Diskussionen in der MII bildet das Terminologie-Framework von Kahn et al. [3].

Die im Rahmen der Projectathons entstandenen Datenqualitätsindikatoren (DQIs) wurden mit dem R-Tool dataquieR [4] praktisch angewendet. Eine Harmonisierung der verwendeten Tools war hier möglich, ist jedoch im Regelbetrieb der DIZe unwahrscheinlich. Auch steht sie den diversen Interessen verschiedener Forscher und den unterschiedlichen Anforderungen ihrer Use Cases entgegen. Um diese zu berücksichtigen, wurden Tool-unabhängige DQIs in einem Metadaten Repository hinterlegt, um einen Austausch von DQIs über technische Restriktionen

hinweg zu ermöglichen. Sie referenzieren den MII-Kerndatensatz, ermöglichen darüber hinaus aber auch die Verwendung von standardisierten medizinischen Terminologien wie ICD-10. Diese Verschränkung ermöglicht eine Nachnutzung der erstellten DQIs in anderen MII Use Cases und darüber hinaus, ganz im Sinne der FAIR Kriterien [5]. Außerdem wird aus der Perspektive von Kohortenstudien vorgestellt, wie man DQIs FAIR machen kann und inwieweit diese Erfahrungen auf Anforderungen der MII übertragbar sind. Dies betrifft einerseits die maschinenlesbare Aufbereitung von Metadaten im Vorfeld der Analyse von DQIs sowie andererseits deren Ablage in Metadaten-Repositoryn.

Ziel des Workshops ist es, die Anforderungen an solche FAIR DQIs zu evaluieren und die verschiedenen Interessen strukturiert zu erfassen. Neben den Erfahrungen aus einem MII DIZ und den MII Projectathons stellen wir eine mögliche Repräsentation der FAIR DQIs vor. Anhand dieser werden vorhandene Schwächen diskutiert und anschließend gemeinsam weitere Anforderungen an FAIR DQIs erarbeitet, die eine Nutzung dieser in anderen, etablierten Tools zur Datenqualitätssicherung ermöglichen.

## **Ablauf**

### *Session 1*

- FAIRe Analysen der Datenqualität in Kohortenstudien - Translationspotenzial für die MII
- Analysen der Datenqualität in den ETL-Stecken eines MII DIZ
- Vorstellung des DQ-Frameworks aus den MII-Projectathons
- mögliche Repräsentation von FAIR DQIs

### *Session 2*

- Anforderungen an FAIRe DQIs aus Sicht der MII
- gemeinsame Diskussion und Erfahrungsaustausch

## Literatur

- [1] Draeger C, Löbe M, Meineke FA. Data Quality Assessment in the seventh MII-Projectathon. German Medical Science GMS Publishing House; 2023. DOI: 10.3205/23gmds139
- [2] Draeger C, Tute E, Schmidt CO, Waltemath D, Boeker M, Winter A, Löbe M. Identifying Relevant FHIR Elements for Data Quality Assessment in the German Core Data Set. Studies in Health Technology and Informatics. IOS Press; 2023. accepted for publication
- [3] Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. Vol. 4, eGEMs (Generating Evidence & Methods to improve patient outcomes). Ubiquity Press, Ltd.; 2016. p. 18. DOI: 10.13063/2327-9214.1244
- [4] Schmidt CO, Struckmann S, Enzenbach C, Reineke A, Stausberg J, Damerow S, et al. Facilitating harmonized data quality assessments. A data quality framework for observational health research data collections with software implementations in R. Vol. 21, BMC Medical Research Methodology. Springer Science and Business Media LLC; 2021. DOI: 10.1186/s12874-021-01252-7
- [5] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). DOI: 10.1038/sdata.2016.18