

Implementierung der FAIR-Prinzipien im Forschungsdatenmanagement: eine Terminologie-basierte Strategie für die inhaltliche Beschreibung numerischer Faktendatensätze

Giacomo Lanza, Joachim Meier / Physikalisch-Technische Bundesanstalt, Referat Q.11
„Wissenschaftliche Bibliotheken“, Braunschweig und Berlin
Ulrich Schwarldmann / Gesellschaft für Wissenschaftliche Datenverarbeitung Göttingen

In der Open Science-Ökonomie stellen numerische Faktendaten eine sehr große Herausforderung für die praktische Umsetzung der vier FAIR-Prinzipien dar. Dies resultiert aus der unüberschaubar großen Anzahl der Datensätze und aus der großen Vielfalt an verwendeten Datenstrukturen, die in den verschiedenen Disziplinen gängig sind. Diese Heterogenität erschwert den Vergleich von Forschungsdatenstrukturen unterschiedlichen Ursprungs, sowie die Festlegung eines einheitlichen Standards zu ihrer Metadatenbeschreibung: z.B. sieht das DataCite-Metadaten-Schema keine Felder für eine detaillierte Beschreibung jenseits der Angabe unkontrollierter Schlagwörter vor. Vor diesem Hintergrund ist bereits das F-Prinzip (Auffindbarkeit) nur eingeschränkt umsetzbar: zielgerichtetes, feingranulares Suchen und präzises Finden auf Datenrepositorien-übergreifender Ebene ist auf dem jetzigen Stand nicht möglich.

Wir setzen zur Lösung des Problems bei den typischen Eigenschaften numerischer Faktendaten an. Diese sind gekennzeichnet durch Messgrößen, Maßeinheiten, numerische Wertebereiche, Rollen (z.B. Messgröße, Messvariable, Messparameter) und, bei quantitativ bewerteter Zuverlässigkeit der Faktendaten, auch die Messunsicherheitsangabe und das Messunsicherheitsmodell. Als Module eines „Metrology Terminology Directory“ (MTD) werden in abgegrenzten Namensräumen kontrollierte Vokabulare für Messgrößen, Maßeinheiten, Messverfahren und verschiedene Charakteristiken von Messobjekten mehrsprachig entwickelt und jeweils über spezifische Persistent Identifiers in einer sogenannten Data Type Registry sprachübergreifend adressierbar gemacht. In von uns neu entwickelten Faktendaten-spezifisch strukturierten Metadatenmodulen dienen diese Vokabulare zur Beschreibung des Faktendatensatzes. Auf diese Weise werden die wesentlichen Eigenschaften numerischer Faktendatensätze für komplexes Suchen und Finden mit geeigneten Retrieval-Werkzeugen zugänglich gemacht.

Durch Implementierung der Metadaten-Schema-Module seitens der Hersteller von digitalen Messgeräten, digitalen Sensoren, Messverarbeitungs-Software, könnte zukünftig erreicht werden, dass die Metadatenbeschreibung schon bei der erstmaligen analog/digital-Wandlung von Messdaten beginnen und dann über die weitere Verarbeitungskette bis zur Archivierung und Publikation angereichert werden kann. Das würde die Dokumentationsarbeit erleichtern und zur vollständigen Umsetzung der FAIR-Prinzipien beitragen können.

Über eine Pilotrealisierung der MTD und ausgewählter Metadaten-Module werden wir im Vortrag berichten.