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

Giacomo Lanza¹, Joachim Erich Meier¹, Ulrich Schwardmann² und Thomas Wiedenhöfer¹

¹Physikalisch-Technische Bundesanstalt (PTB) ;

²Gesellschaft für die Wissenschaftliche Datenverarbeitung Göttingen (GWVG)

Abstract - Deutsch

In der Open Science-Ökonomie stellen numerische Faktendaten eine disziplinübergreifende Herausforderung für die praktische Umsetzung der vier FAIR-Prinzipien dar. Die zu erwartende unüberschaubar große Anzahl von Datensätzen und die heute schon in den verschiedenen Disziplinen gängige große Vielfalt an verwendeten Datenformaten und -strukturen sind wesentliche Ursache des Problems. Diese Heterogenität erschwert die Festlegung eines einheitlichen Standards zur Metadaten-Beschreibung und Archivierung von Forschungsdaten unterschiedlichen Ursprungs, und folglich deren Nachnutzung: z.B. sieht das *Data-Cite*-Metadaten-Schema keine Felder für eine detaillierte Beschreibung numerischer Faktendaten 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 derzeitigen Stand nicht möglich.

Wir setzen zur Lösung des Problems bei den typischen Eigenschaften numerischer Faktendaten an. Diese sind gekennzeichnet durch Messgrößen, Maßeinheiten, numerische Werteangaben, 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, Messdatenverarbeitungs-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, einen gewissen Qualitätsstandard einführen und somit zu einer effizienten Umsetzung der FAIR-Prinzipien beitragen können.

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

Abstract - Englisch

Within the Open Science economy, a major challenge for the practical realisation of the FAIR principles is represented by numerical factual data. This is a result of the overwhelming big quantity of data sets, as well as of the big variety of data structures currently used in the different disciplines. This heterogeneity makes it difficult to compare data structures of different origins, as well as the establishment of unique standards for their description through metadata: for instance, DataCite metadata scheme doesn't include any fields for the detailed description of factual data, other than the inclusion of free-text – thus not controlled – keywords. Under these conditions there are limited possibilities only to implement even the first of the FAIR principles, Findability: a guided advanced search of factual data over a plurality of repositories is currently not feasible.

We are approaching a solution to this problem via the typical characteristics of numerical factual data, which are commonly described by quantity names, units, numerical value ranges, roles (measured quantity, variable or parameter) and some estimate of the data reliability, such as the uncertainty budget and the uncertainty model. Within a newly defined "Metrology Terminology Directory"(MTD), we are developing a collection of multilingual controlled vocabularies for physical quantities, measuring units, experimental techniques and selected information about research objects; the correct designation of each item takes place in a language-independent manner via a persistent identifier. These vocabularies are then applied within an ad hoc defined metadata module for numerical factual data for a thorough description of a dataset. That way, the relevant features of numeric factual data are made accessible for complex searches and finding with suitable retrieval tools.

If the proposed metadata module is adopted and implemented by the producers of digital measuring instruments, digital sensors and software for data analysis, it will be possible in the future to implement a standardised metadata description already at the point of the first analog-to-digital conversion, and then propagate and enrich it somehow automatically along all subsequent data transformation steps. This would facilitate documentation work considerably, as well as would contribute to the realisation of all elements of the FAIR principles.

Within the talk a pilot realisation of the MTD will be presented, along with selected metadata modules.

1. Hintergrund und Problemstellung

Der freie Austausch von Informationen und Wissen ist eine der tragenden Säulen der Wissenschaft. Mit zunehmender Digitalisierung in der Wissenschaft und mit der Verbreitung der Open-Science-Philosophie werden in wachsendem Ausmaß auch Forschungsdaten der Öffentlichkeit verfügbar gemacht. Diese Daten fallen in einer Vielzahl offener und proprietärer Formate an. Für deren Beschreibung werden Metadaten verwendet, deren Umfang i. d. R. von den Möglichkeiten der Datenportale oder von Fachgemeinschaften bestimmt werden. Die Qualität der Metadatenerfassung ist dabei abhängig von der fachlichen Expertise und dem Vollständigkeitsanspruch der mit dem Datenkuratieren beauftragten Personen.

Eine Nachnutzung dieser Daten setzt die Fähigkeit voraus, sie mittels geeigneter Suchmaschinen selektiv recherchieren und filtern zu können. Eine feingliedrige, „erweiterte“ Suche bedarf einer gemeinsamen „Sprache“ zwischen den Suchmaschinen und den Datenrepositorien. Das wurde zum Teil erreicht mit der Festlegung standardisierter Metadatenschemata (DataCite [1]) und Protokolle für das *Metadata Harvesting* (OAI-PMH); darüber hinaus wenden einige Fachdisziplinen zusätzliche Metadatenmodule an, die eine feingranulare Beschreibung ermöglichen. In gewissen Fällen wird auch eine Suche im Datenbestand selbst angeboten.

Angesichts der wachsenden Datenmengen und der anfallenden multidisziplinären Fragestellungen wird der Bedarf immer offensichtlicher an einer interoperablen Struktur für die dokumentarische Beschreibung der Daten, die das zuverlässige Finden, Filtern und Vergleichen von Forschungsdaten unterschiedlichen Ursprungs ermöglicht. Die weitgehend KI-unterstützten Methoden der *Big Data*-Analyse und des *Machine Learning* erfordern, dass die Daten nicht nur für Menschen zugänglich, sondern auch maschinenlesbar und -verständlich sind. Dies erfordert eine Auswahl standardisierter, langlebig lesbarer Datenformate, sowie eine eindeutige Kodierung der Metadatenbeschreibungen mit Standardisierung sowohl der Attributfelder, als auch deren zulässiger Inhalte (Attributwerte). Die für Forschungsdatenmanagement eingeführten vier FAIR-Prinzipien [2], die Daten sollen auffindbar (*Findable*), zugänglich (*Accessible*), interoperabel (*Interoperable*) und wiederverwendbar (*Reusable*) sein, geben die Ziele vor, machen aber keine Angaben für Lösungen dieser nicht trivialen Aufgabenstellung.

Sehr stark automatisierungs- und messtechnisch geprägte Industrieunternehmen sind sich dieser Herausforderungen seit längerem bewusst. Firmeneigene Schemata und Protokolle für die digitale Übertragung von Informationen sind bereits entwickelt [3, 4, 5, 6].

2. Lösungsansatz

Forschungsdaten ohne eine standardisierte, feingranulare, den Suchmaschinen zugängliche Metadaten-Beschreibung sind schwierig zu finden bzw. mit zunehmender Komplexität schwer oder nicht immer eindeutig interpretierbar.

Um die Interoperabilität von Forschungsdaten zu gewährleisten, ist die Einführung einer gemeinsamen Grundlage für die Metadaten-Beschreibung unentbehrlich. Diese entsteht aus (1) einer definierten Palette von Attributen (Metadatenfelder) sowie (2) je einer kontrollierten Liste zulässiger Werte, die eine eindeutige, reproduzierbare und zweifelsfreie

Erfassung ermöglicht. Das Ganze sollte international und womöglich fachübergreifend abgestimmt werden. Ein solcher Standard besteht aus den folgenden Bausteinen:

- Ein Metadatenchema, also eine hierarchische Anordnung kontrollierter Metadatenfelder mit festgelegten Benennungen und vorgegebenen Regeln zur inhaltlichen Belegung (Variabeltyp, Freitext / kontrollierter Text). Erstrebenswert ist hier ein modularer Aufbau, bei dem neben einem gemeinsamen Kernschema (z.B. eine Erweiterung des aktuell verwendeten DataCite-Metadatenchema) unterschiedliche fachspezifische oder methodenspezifische Metadaten-Module optional verwendet werden können.
- Mehrere kontrollierte Vokabulare, die für die Inhalte ausgewählter Metadatenfelder eine Liste zulässiger Werte (Terme) zur Verfügung stellen. Je nach Komplexität können diese Vokabulare als Thesauri oder Ontologien realisiert werden. Um die Mehrdeutigkeiten der natürlichen Sprache zu überwinden, ist es ratsam, die einzelnen Begriffe mit langlebigen Kennzeichen (Persistent Identifiers) zu versehen. Hierdurch wird die Verfügbarkeit der Vokabulare in mehreren (menschlichen) Sprachen unter Beibehaltung der logischen Struktur möglich.

Eine derartige Vorgehensweise überwindet die Begrenztheit generischer Metadaten-schemata (wie z.B. das von DataCite) und ermöglicht gleichzeitig die eindeutige Interpretation (Disambiguierung) der Benennungen von Feldern und Inhalten, die in unterschiedlichen Fachbereichen unterschiedliche Bedeutung tragen könnten.

3. Beispiel: Größen und Einheiten

In den quantitativen Wissenschaften (z. B. Chemie, Physik, Materialwissenschaften und andere technische Wissenschaftsgebiete) spielen numerische Faktendaten eine zentrale Rolle. Die grundlegenden Informationen für deren eindeutige Identifizierung sollten Angaben beinhalten über

- Versuchsobjekt: Probenotyp, Proben-ID, chemische Identität, Hersteller, Auftraggeber;
- Datengenerierung: experimentelle Methode bzw. Simulationsprozedur, Identifikation des Messplatzes, Zeitpunkt, Mitarbeiter. . . ;
- Faktendaten-Merkmale: Größe, Maßeinheit, numerischer Wertebereich, numerische Auflösung, Messunsicherheit, Rolle (Messgröße, Variable oder Parameter).

Im vorgestellten Vorgehen wurde zunächst die Darstellung der Datenwerte behandelt. In der Folge wurde ein atomares Metadatenchema für numerische Faktendaten und Vokabulare für physikalische Größen und Einheiten entwickelt.

4. Ergebnisse: Metadatenchema

Im Rahmen des innerhalb Horizon 2020 europäisch geförderten Projektes *SmartCom* wurde für industrielle Anwendung ein atomares Metadatenchema (**D-SI** [7]), basierend auf internationalen Richtlinien der Messtechnik, für die Beschreibung numerischer Faktendaten definiert. Das Schema sieht Felder für die Angabe eines Messwerts mit Maßeinheit und Messunsicherheit vor. Die Unsicherheit kann entweder als absolute erweiterte Unsicherheit samt Erweiterungsfaktor, oder als Intervall samt Intervallgrenzen dargestellt werden; in beiden Fällen wird die Überdeckungswahrscheinlichkeit angegeben. Optionale Angaben dazu sind der Größenname, der Zeitstempel sowie die angenommene Wahrscheinlichkeitsdichteverteilung des Messunsicherheitswertes. Die Qualitätskontrolle legt den größten Wert auf die Angabe der Einheiten, wobei SI-Basiseinheiten ohne Präfixe stark empfohlen werden.

Tabelle 1.: Vorgeschlagene Metadatenfelder für die Beschreibung numerischer Faktendaten. Für jede Größe werden ein PID, die Rolle, die Einheit, die numerischen Extremwerte und die Unsicherheit angegeben. Die **fett** markierte Felder stellen eine Erweiterung im Vergleich zum heutigen **D-SI**-Schema dar.

Feld	Beschreibung	Beispiel	Obligation
block	Einzeldatei, oder zusammenhängender Tabellenbereich		M
block/npoints	Anzahl Messwerte	941	M
block/quantity	In einer Datei enthaltene Größe		M
block/quantity/ role	Funktion der aufgeführten Größe: "measurand", "parameter" oder "variable"	https://unserDatenraum.boh/ role:variable	M
block/quantity/ qty_id	Identifikator der Größe	https://unserDatenraum.boh/ qty:wavelength	M
block/quantity/description	Ausführliche Beschreibung	Wellenlänge des Lasers zur Bestrahlung der Probe	O
block/quantity/list/label	Kurzbezeichnung / Symbol	X	O
block/quantity/list/real[label="min"]/value	Minimalwert	230	M
block/quantity/list/real[label="max"]/value	Maximalwert	700	M
block/quantity/list/ unit_id	Identifikator der Einheit	https://unserDatenraum.boh/ unit:nm	M
block/quantity/list/uncertainty	Erweiterte Messunsicherheit (wenn role="measurand")	1	MA
block/quantity/list/coverageFactor	Erweiterungsfaktor (default=2)	2	MA

Für die Beschreibung von Forschungsdaten für Open Science-Anwendungen wird dieses Schema erweitert. Für jede im Datensatz auftauchende Größe werden sowohl der Wertebereich (Minimum und Maximum) als auch die Unsicherheit angegeben. Die Benennung

der Größe ist verpflichtend und soll möglichst aus einem normierten Vokabular stammen. Zusätzlich wird die Rolle der Größe (Messgröße, Messparameter oder Messvariable) sowie eine wörtliche Erläuterung angegeben.

Das vorgeschlagene Schema bildet die hierarchische Anordnung der relevanten Felder (Tabelle 1) ab. Die korrekte Eingabe der Metadatenwerte wird durch Anwendungsregeln unterstützt. Struktur und Regeln sind bewusst syntaxneutral angelegt und ermöglichen deshalb die äquivalente Auszeichnung bzw. Export der Metadaten in XML, JSON, YAML oder einem anderen beliebigen Auszeichnungsformat bzw. Notation.

5. Ergebnisse: kontrollierte Vokabulare

Die entwickelten Vokabulare werden in einem *Metrology Terminology Directory* gesammelt und sollen in maschinenlesbarem Format (JSON, RDF oder OWL) für die Wissenschaft, die Wirtschaft und die Öffentlichkeit zur Verfügung gestellt werden. Eine Testinstanz befindet sich auf dem von der GWDG betriebenen ePIC-Server [8].

The screenshot displays the ePIC Data Type Registry (testing) interface. The top navigation bar includes 'Introduction', 'All', 'Types', and a 'Sign In' button. The main header features the ePIC logo (Persistent Identifiers for eResearch) and the GWDG logo. A search bar is present below the header.

The main content area is titled 'qty:temperature' and includes a 'Type: PID-BasicInfoType-Metrology' label. It offers four view options: 'Digital Object View', 'JSON View', 'Versions View' (selected), and 'Show Relationships'.

The 'Identifier' section shows the value '21.T11148/29481f511f6208d8170a'. Below this, the 'Names in different languages' section is displayed as a table:

Language Code *	Descriptive Name *
de	thermodynamische Temperatur
fa	دماي ترموديناميكي
hi	ऊष्मतिकीय तापमान
zh	热力学温度

Each row includes a small text block explaining the ISO 639-1 (two-letter) code for the language in which the name is given.

The right-hand side of the interface shows 'Properties' and 'Restrictions on the type'. The 'Dimensions' section contains the symbol 'Θ¹' and a description: 'For measurement types (quantity / unit / constant), an expression relating the object to the SI fundamental quantities / units.' The 'Relations to other objects' section includes a table with columns: 'Nature of Relation *', 'Object Name *', 'Identifier', 'Issued By', and 'Details'. A row is shown with 'has_unit' as the relation, 'unit.K' as the object name, and 'BIPM' as the issuer.

Below the relations table, there are sections for 'symbols' and 'object', each with a dropdown menu for 'alphabet' (set to 'Latin') and 'symbol' (set to 'T, (Θ)').

Abbildung 1.: Ein Beispiel aus dem Vokabular für physikalische Größen ($qty=quantity$). Für jeden Eintrag (in diesem Fall, die Größe "Thermodynamische Temperatur") werden ein Identifikator und standardisierte mehrsprachige Bezeichnungen zugewiesen.

Das Vokabular für physikalische Größen besteht derzeit aus ca. 400 Einträgen, von denen ein Beispiel in Abbildung 1 wiedergegeben wird. Für jeden Eintrag werden die folgenden Attribute angeboten:

The screenshot displays the ePIC Data Type Registry (testing) interface. The main content area shows the entry for the unit 'Kelvin' (unit:K). The entry includes an identifier (21.T11148/7791c58cc21bc2ed0c8e) and names in various languages (de, fa, hi, zh, ja, ms). A table lists the names in different languages, including German (Kelvin), Arabic (كَلْوِن), Hindi (केल्विन), Chinese (开尔文), Japanese (ケルビン), and Malay (kelvin). The entry also shows the definition of the object in the chosen language, details, remarks, or special cases, and the symbol 'K' in different alphabets (Latin, Cyril, Arab, LaTeX). The 'Dimensions' section shows the unit is derived from the SI fundamental quantities / units. The 'Relations to other objects' section shows the unit is derived from unit A, is interchangeable with unit A, is the definition of unit A, is used by unit A, and is the unit for type A. A table lists the nature of relation, object name, identifier, issued by, and details for the relation 'is_unit_for' with the object 'qty:temperatu' and identifier 'BIPM'.

Abbildung 2.: Ein Beispiel aus dem Vokabular für Einheiten (unit). Zum Eintrag (in diesem Fall die Einheit "Kelvin") gehören ein Identifikator und die Benennungen in mehreren Sprachen. Das Einheitensymbol in verschiedenen Schriftarten (lateinisch, kyrillisch, arabisch, LaTeX-Kodierung) wird ebenso angezeigt.

- Benennung, derzeit in 18 Sprachen.
- Symbol und Definitionsformel.
- Dimensionen nach BIPM Broschüre [9]; Bezug zur vorgegebenen SI-Einheit.
- Kurzer Definitionstext und Anmerkungen.
- Notation aus einer mehrstufigen Klassifikation.

Das Vokabular für Naturkonstanten, in Anlehnung an die Liste von CODATA [10], enthält derzeit 84 Einträge. Für jede Naturkonstante werden die schon bei Größen gelisteten Attribute angegeben, und zusätzlich:

- Der numerische Wert der Naturkonstante mit absoluter und relativer Standardunsicherheit und die Gültigkeitszeitspanne dieses Wertes (basiert auf den CODATA-Ausgaben von 1969, 1973, 1986, 1998, 2002, 2006, 2010, 2014 und 2018).
- Die Korrelationskoeffizienten zu den anderen Naturkonstanten, mit Gültigkeitszeitspanne.

Das Vokabular für Einheiten listet derzeit ca. 100 Einträge auf; ein Beispiel wird in Abbildung 2 gezeigt. Dazu gehören, konform zur BIPM-Broschüre, die 22 kohärenten Einheiten und die derzeit noch akzeptierten Nicht-SI-Einheiten, sowie deren Kombinationen nach Summen und Produkten (z. B. J/K, N·m). Für die Einheiten werden folgende Attribute vergeben:

- Benennung, derzeit in 18 Sprachen.
- Symbol in lateinischem, kyrillischem und arabischem Alphabet, sowie LaTeX-Quellcode.
- Definitionsformel mit Umrechnungsfaktor, sofern notwendig.
- Dimensionen nach BIPM-Broschüre und Bezug zu den damit verbundenen Größen und Konstanten.
- Kurzer Definitionstext und Anmerkungen.

Geplant ist ein viertes Vokabular, das auf Basis des VIM (Internationales Wörterbuch der Metrologie) [11] und des GUM (Guide to the expression of uncertainty in measurement) [12] die grundlegenden metrologischen Begriffe mehrsprachig und maschinenlesbar darstellen wird. Unter anderen werden die Grundbegriffe in Bezug auf Messverfahren, Messunsicherheit sowie die Rolle einer Größe aufgeführt.

6. Beispielanwendung

Das hier eingeführte Schema samt Vokabularen stellt eine mögliche Grundlage dar für die interoperable Erfassung von Forschungsdaten in einem Datenportal oder Repositorium, das eine feingranulare Suche ermöglicht. Dazu soll eine spezialisierte Suchmaschine eingerichtet werden, die nach Größennamen und Wertebereichen indexieren und filtern kann.

Ein möglicher Anwendungsfall: Ein Materialwissenschaftler könnte nach aktuellen Erkenntnissen über die Temperaturabhängigkeit (*Variable*) verschiedener Materialeigenschaften (*Messgrößen*) einer Substanz unter gewissen Umweltbedingungen (*Parameter*) suchen. Eine solche Recherche sollte alle Datensätze finden und wiedergeben, in denen die gewünschte Substanz in der Probenbeschreibung, die Materialeigenschaft als Größe vorkommt und die Temperaturwerte zwischen vorgegebenen Grenzwerten (z. B. zwischen 200 K und 1000 K) liegen.

Eine solche Suchplattform, der *MessdatenMetaViewer* (Abbildung 3 (a)), wurde in unserer Arbeitsgruppe entwickelt. Die Anwendung basiert auf einer Datenbank, in der die

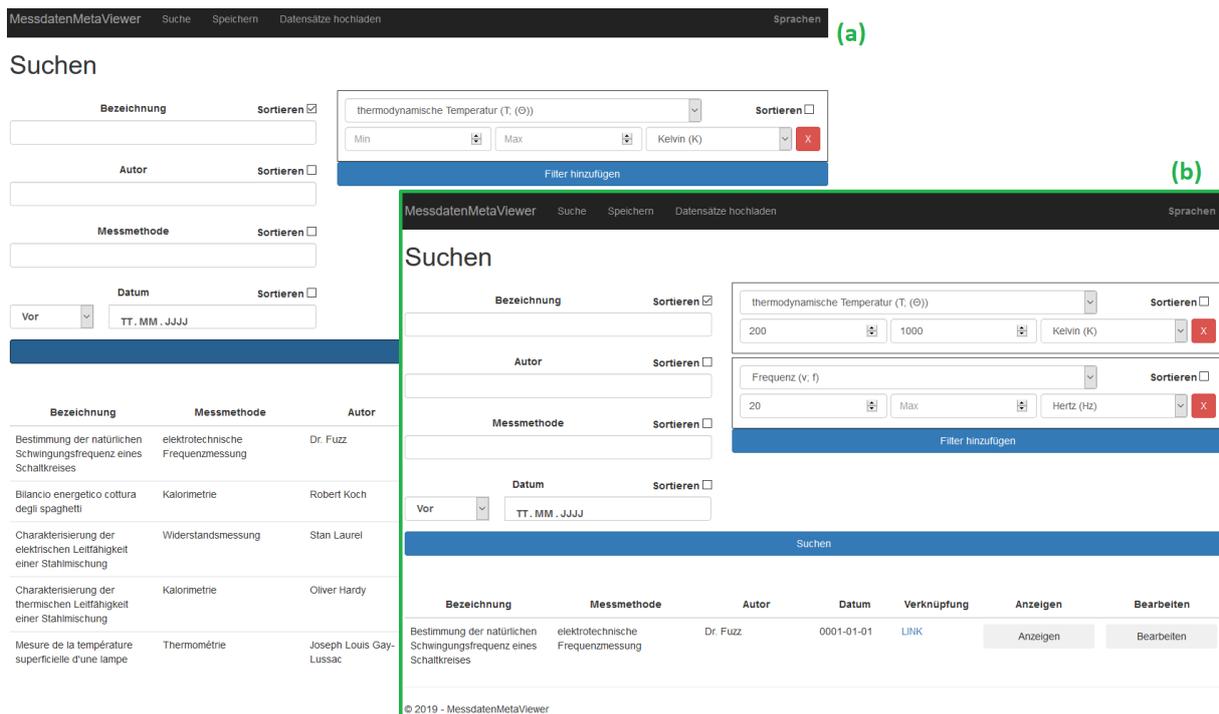


Abbildung 3.: Die Suchplattform *MessdatenMetaViewer*, die eine erweiterte Suche auf Basis der von uns definierten Felder und der festgelegten Vokabulare ausführt und nach Größen und Wertebereichen filtern kann. (a): Ein Schnappschuss aus einer kleinen Forschungsdatensammlung, mit der Suchmaske. (b): Die gleiche Datensammlung nach einer Suche nach den Größen *Thermodynamische Temperatur* und *Frequenz*.

Datensätze nach Autor, Datum, Messmethode und zugehörigen Größen erfasst sind; für jede Größe werden zusätzlich die dazugehörige Einheit und der Wertebereich angegeben. Die Suchfunktion filtert sowohl nach Größen, als auch nach Wertebereichen und macht es deswegen möglich, die passenden Datensätze aus der Menge auszuwählen. Ein Beispiel wird in Abbildung 3 (b) gezeigt, in der nur die Datensätze mit Frequenz- und Temperaturangabe in einem gewissen Bereich wiedergegeben werden.

7. Schlussfolgerung

Das vorgeschlagene Metadatenmodul ermöglicht eine feingranulare Erfassung von Datensätzen durch wenige bedeutsamen Metadatenfelder und lässt sich leicht als modulare Erweiterung in andere, auch fachspezifische Metadaten schemata integrieren. Die Speicherung der Metadatenwerte als Persistent Identifiers garantiert die eindeutige maschinelle Lesbarkeit der Inhalte und ermöglicht eine schnelle Übersetzung in beliebige menschliche Sprachen für diverse Anwendungen. Da jeder erfassten Größe eine eindeutige Rolle (Mess-Größe, -Parameter, -Variable) und das Intervall deren numerischen Werte zugewiesen wird, ermöglicht das die zielgerichtete Suche nach numerischen Faktendaten.

Die Verwendung persistenter Identifikatoren für die kontrollierten Vokabulare für die

Benennung der Metadatenfelder und -inhalte aus einem Thesaurus ermöglicht die Kommunikation mit Suchmaschinen, die über verschiedenen Datenbanken recherchieren. Somit ermöglicht der hier vorgestellte terminologische Ansatz einen vielversprechenden Weg, numerische Faktendaten nachhaltig und interoperabel zu beschreiben, zu speichern und gezielt wieder zu finden.

8. Ausblick

Schon heute werden bei einer softwaregesteuerten Messung wichtige Parameter über den Messablauf von den Messgeräten selbst in den Kopfzeilen der erzeugten Rohdaten oder in zusätzlichen begleitenden Dateien protokolliert. Auch erlaubt bestimmte Software für die wissenschaftliche Datenauswertung die Speicherung von Informationen über enthaltene Variablen und deren numerischen Werte. Die Lesbarkeit und Nachnutzbarkeit dieser Metadaten hängt vom Format und von der verwendeten Syntax ab. Wären sie in einem standardisierten Format abgespeichert, so könnten sie leicht von anderer Software gelesen, interpretiert und genutzt werden. Somit würde die Metadatenbeschreibung eines Messobjektes entlang des Datenverarbeitungsprozesses stetig angereichert werden und am Ende würde die gesamte Dokumentation über den Messprozess und seine Ergebnisse vorliegen. Metadatenerfassung durch einen Menschen würde so weitgehend vermieden und beschränkte sich daher auf diejenigen Felder, die nicht im Verlauf des Prozesses automatisch beschrieben werden können. Weitere wesentliche Vorteile dieser Vorgehensweise sind, dass fehleranfällige nachträgliche „händische“ Erfassung von Metadaten minimiert wird und die Kooperationsbereitschaft für Forschungsdatenmanagement mit Zielrichtung Open Science beim für die Datengewinnung zuständigen Personal wächst.

Um die Interoperabilität, eine breite Akzeptanz und die Praxistauglichkeit sicherzustellen, wird die Einbeziehung von Messgeräteherstellern und Herstellern wissenschaftlicher Software (ELN - Elektronische Laborbücher, LIMS - Laborinformationsmanagementsysteme) bei der Festlegung des Formats angestrebt.

Danksagung

Wir bedanken uns bei unseren PTB-Kollegen Robin Becker für die Entwicklung der Suchplattform *MessdatenMetaViewer* und Sascha Eichstädt für nützliche Diskussionen.

Literaturverzeichnis

- [1] The DataCite Schema. <https://schema.datacite.org>
- [2] FORCE 11: the FAIR data principles. <https://www.go-fair.org/fair-principles/>
- [3] ASAM, the ASAM Open Data Services format. <https://www.asam.net/standards/detail/ods/wiki/>
- [4] Woopsa Protocol Specifications. <http://www.woopsa.org/specifications/>

- [5] eCl@ss - Standard für Stammdaten und Semantik für die Digitalisierung. <https://www.eclasscontent.com/index.php>
- [6] Das Referenzarchitekturmodell Industrie 4.0 (RAMI 4.0). <https://www.plattform-i40.de/I40/Redaktion/DE/Downloads/Publikation/struktur-der-verwaltungsschale.pdf>
- [7] EMPIR 17IND02 (SmartCom). D-SI: XML implementation of the Digital SI (D-SI) meta data model for the exchange of metrological data - version 1.0.1-beta. https://www.ptb.de/si/smartcom/d-si/v1_0_1/SI_Format.xsd
- [8] Projekt ePIC: Persistent identifiers for research. <http://dtr-test.pidconsortium.eu/>
- [9] SI, Le Système International d'Unités—The International System of Units, 9th edn. BIPM 2019, ISBN 978-92-822-2272-0. <https://www.bipm.org/fr/publications/si-brochure/>
- [10] P. J. Mohr, D. B. Newell and B. N. Taylor. CODATA recommended values of the fundamental physical constants: 2014. https://ws680.nist.gov/publication/get_pdf.cfm?pub_id=920686
- [11] Bureau International des Poids et des Mesures, JCGM 200:2012. Vocabulaire international de métrologie - Concepts fondamentaux et généraux et termes associés (VIM). https://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf
- [12] Bureau International des Poids et des Mesures, JCGM 100:2008. Évaluation des données de mesure - Guide pour l'expression de l'incertitude de mesure (GUM). https://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_F.pdf