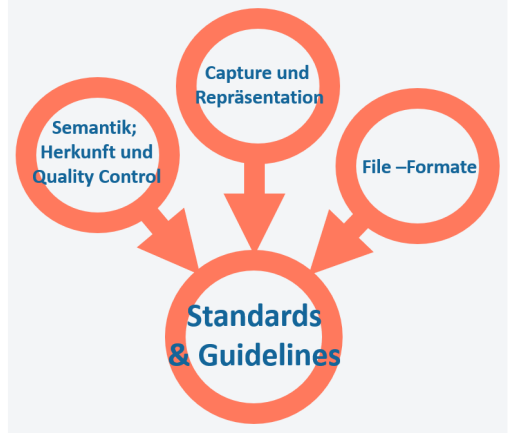




Minimal Information Standards (MIS)

Durch Open Science und andere Bestrebungen wird es zunehmend möglich Queranalysen mit Daten aus Repositorien und Archiven zu betreiben – eine große Chance für die Wissenschaft. Aber *aus großer Kraft folgt große Verantwortung*. Nämlich auch die eigenen Daten so aufzubereiten, dass sie vollständig dokumentiert, transparent, reproduzierbar und in strukturierte Datenbanken überführbar sind. Für diese Zwecke gibt es Metadaten Standards – und für Daten aus Hochdurchsatz-Verfahren eben die Minimal Information Standards (MIS).



Standards, Guidelines, MIS umfassen üblicherweise Metadaten-Angaben zu den drei genannten Bereichen.

Metadaten zu verschiedenen Zwecken

Es gibt zahlreiche Standards und Guidelines – ein guter Standard macht Angaben, die drei wichtige Bereiche der Metadaten abdecken (siehe Abbildung). Ein Bereich umspannt Vorgaben zu den File-Formaten. Ein anderer die benötigten Metadaten, um die Daten erfassen, darstellen und interpretieren zu können. Der dritte Bereich betrifft die Maschinenlesbarkeit: Es muss eine (weltweit) eindeutige Semantik für die Daten geben. Ihre Herkunft (Provenance) muss klar sein, dies entspräche z.B. den Hardware-Komponenten und -Settings, den Messparametern. Metadaten zur Quality Control (QC) der Daten und Messergebnisse zeigen an, ob diese genau, konsistent und vollständig sind.

Die gewählten Minimal Information – und Metadaten Standards, und was sie für die Forschungsgruppe konkret bedeuten, lassen sich auch super im DMP (Daten Management Plan) festhalten und dort nachlesen; so wird er zum lebendigen Dokument.

Standards & Frameworks

Minimal Information Standards wurden speziell für Daten aus Hochdurchsatz-Methoden entwickelt, derer man in den Naturwissenschaften einige findet (Genomics, Proteomics, Neurosciences, ...). Da jeder solche Forschungsbereich andere Anforderungen an eine ausreichende Dokumentation hat, gibt es zahlreiche MIS. Alle bedeutenden sind auf der Wikipedia-Seite zu dem Thema zu finden: https://en.wikipedia.org/wiki/Minimum_information_standard

Für Biowissenschaftler*innen mit Publikationsambitionen gibt es zusätzlich das Materials Design Analysis Reporting (MDAR) Framework, das auch Best Practice Empfehlungen enthält: <https://osf.io/2k3va/>

Metadaten an einer medizinischen Uni

Minimum Information Standards fragen auch nur ein *Minimum* ab, das nötig ist um Daten zu interpretieren, teilen und reproduzieren. *Über dieses Minimum hinauszustreben* und bei Best Practice anzukommen, liegt im Sinne der Guten Wissenschaftlichen Praxis.

- Es gibt um die tausend medizinische Ontologien - [BioPortal.bioontology.org: Ontologien und Klassen finden](https://bioportal.bioontology.org/Ontologien_und_Klassen_finden)
 - Ein schwieriges Unterfangen ist es einzelne Ontologien miteinander zu vereinen, weil Unterschiede in der Begriff-Definition schon damit anfangen, ob man sie prä- oder postkoordiniert zusammen setzt
 - Es gibt drei verschiedene Arten von Ontologie-Sprachen, deren Akronyme auch beschreiben welche logischen Verknüpfungen jeweils verwendet werden
 - Beispiel für die verwendete Logik:

$$\text{Mumps} \equiv \exists \text{Auslöser. MumpsVirus} \cap \exists \text{Pathologie. Infektion}$$
(Mumps ist gleichbedeutend mit der Existenz von Mumpsvirus als Auslöser und einer vorhandenen Infektion)
- Aufbau einer Ontologie – Quelle & mehr dazu:
<https://www.heise.de/hintergrund/Ontologien-in-der-Medizin-Struktur-und-Erstellung-9797489.html>

Ontologische Streifzüge

Medizinische Daten können nur sekundärgenutzt und verglichen werden, wenn Terminologien von Krankheitsbildern und klinischen Prozessen in einer standardisierten Semantik dargestellt werden. Noch besser ist es, wenn zusätzlich die logischen Beziehungen zwischen den einzelnen Elementen und deren Abfolge in Sequenzen beschrieben werden kann. Dann spricht man von einer Ontologie. Sie kann auch komplexe Datenmodelle darstellen. Weil Ontologien eine formale Logik-Sprache ermöglichen, sind sie und

Bereichsspezifische Standards & Richtlinien

Für medizinische bzw. Life Science Daten gibt es eine Vielzahl an Metadaten Standards. Die Befolgung dieser macht die Daten u.a. FAIR (Findable, Accessible, Interoperable, Re-usable). Einen Überblick der Standards & MIS gibt es im Anhang zu den Research Data Management Checklists:

<https://muniverse.medunigraz.at/forschung/forschungsdatenmanagement/fdm-checklisten-vorlagen>

Häufig sind medizinische Daten auch persönliche Daten, die der DSGVO unterliegen und daher strenger reguliert sind (mehr dazu: <https://forschungsdaten.info/themen/rechte-und-pflichten/datenschutzrecht/>).

An der Med Uni Graz gibt es Policies für Forschungsdatenmanagement, gute wissenschaftliche Praxis, Open Access, sowie ein Positionspapier zur Datenweitergabe, die ebenso zu beachten sind. Diese und weitere Richtlinien sind hier gesammelt aufgelistet: <https://muniverse.medunigraz.at/forschung/forschungsdatenmanagement/richtlinien>

die daraus hervorgehende deskriptive Logik fester Bestandteil von Wissensrepräsentation und künstlicher Intelligenz geworden.

Sie erfüllen ihren Zweck nur dann gut, wenn sie von einer breiten Masse angenommen und verwendet werden, weshalb es auch in der Forschung bedeutend ist bestehende Ontologien einzelner Fachgebiete miteinander zu verknüpfen. Gleichzeitig müssen Ontologien von der jeweiligen Institution unterstützt und den Forschenden zur Verfügung gestellt werden – Hilfestellung dazu gibt es an der Med Uni Graz am IMI: <https://imi.medunigraz.at/forschung>

RDM Team
Service Fflver:



Gudrun Kainz, RDM Team der Med Uni Graz

<https://www.medunigraz.at/services-fuer-forschende#c54299>

rdmsupport@medunigraz.at

2024

DOI: 10.25365/phaidra.605_01

<https://creativecommons.org/publicdomain/mark/1.0/deed>

Diese Lizenz schließt Projekt- und Institutslogos und individuell referenzierte Materialien nicht mit ein!

