MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

@BoenischCaro

FAIR+R: Making Clinical Data Reliable through Qualitative Metadata

Caroline Bönisch

Research Associate

Dep. Medical Informatics, University Medical Center Göttingen

AIDH IMIA

Introduction

- Application of FAIR Guiding Principles in medical research
- Inclusion of data + corresponding metadata
- Necessity of reliable data in medical research
  - Base for data analysis and potential to result validation
- FAIR Principle lack reliability
- Metadata inherit major aspects in providing reliability

Introduction

- Why metadata?
  - Organizational, structural or content-related metadata
  - Restriction in some repositories: data access only via metadata
  - First impression of data before filing for data use agreements

Information about data quality not fully available within metadata

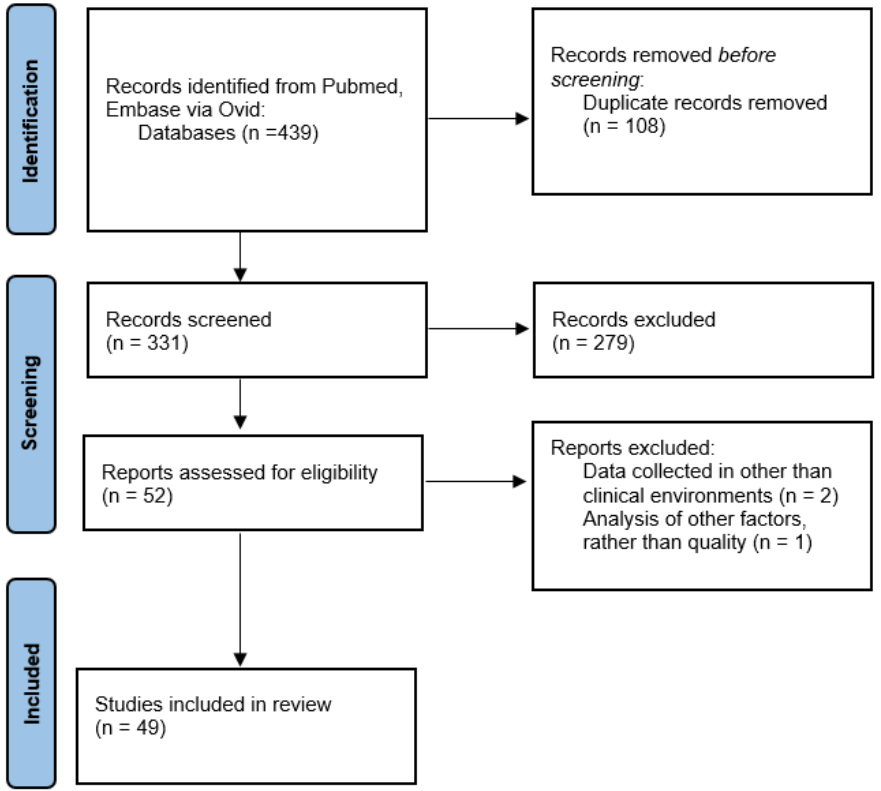
- Anticipation of automated metadata quality measurement

Methods

- Literature search via Pubmed & Embase
- Analyzation criteria of literature:
  - Data collected and stored within clinical environment
  - Methods or evaluation schemes for estimating (meta)data quality
  - Analysis of (meta)data quality factors
  - Earlier approaches to make data reliable in other disciplines
- PRISMA statement as conceptual guide

Methods

PRISMA Flow Diagram based on the literature search:



Results

Proposed metadata quality measures:

Measure	Description
Completeness	All mandatory data fields are filled with information
Consistency	Metadata should be conform to existing standards and formats
Correctness	The information describes the metadata in an accurate and distinct way
Correspondence	Metadata that is linked or inter-dependent represents the same information through every instance
Relevance	The metadata corresponds to the requirement/expectations of the user
Semantic Specificity	Average specificity of a semantic concept in metadata information
Timeliness	Currency of the metadata information describing a resource information
Accessibility	The information of the metadata must be physically available and understandable either by human or machine
Reproducibility	Metadata quality scores should be reproducible and not lack clarity in terminology

Results

Proposed extension to FAIR Guiding Principles:

Extension to the FAIR Guiding Principles
To be Reliable: RL1. (Meta)data should be complete and consistent RL2. (Meta)data should be correct and correspond coherently when inter-record dependent RL3. Metadata should be relevant to the data context of use

Discussion

- Missing: highly accepted definition of metadata quality
- Stated most important quality factor: metadata regarding completeness of the data
- But not the only one!
- Target components of qualitative metadata: relevance, consistency and timeliness

Metadata quality assurance is still seen as more of a casualty

Conclusion

- Quality issues in data repositories gain more visibility
- Sparse literature on metadata quality in medical research
- Inclusion of further scientific fields, to obtain more measures for metadata
- Addition of a fifth block "Reliability" to the FAIR Guiding Principles
- Outlook: Automatization of metadata quality assessment