

Richtlinien zur Erfassung von Daten

Empfehlungen zur Datenqualität bei der Erhebung

Ziel der Richtlinie zur Erfassung von Daten ist es, die Qualität der Daten bei der ersten Erfassung (oder einer Bearbeitung) qualitativ zu verbessern.

Anwendungsbereich

Diese Richtlinien beziehen sich auf die Ersterfassung und Bearbeitung von Daten. Das Dokument dient als Richtlinie, um die Qualität der Daten zum Erfassungszeitpunkt anzuheben.

Richtlinien

1. URIs sollten zur Identifizierung von Entitäten verwendet werden. Im Idealfall sollten dabei URIs aus Referenzservices ([GND](#), [VIAF](#), etc.) verwendet werden. Insbesondere Personen, Institutionen, Publikationen, sollten über Referenzressourcen eindeutig identifiziert werden. Im Idealfall ab der ersten Erfassung.
2. Für eigene Objekte – oder Objekte, welche nicht über einen Referenzservice gefunden werden können – sollten, um deren global eindeutige Identifikation/Referenzierbarkeit zu gewährleisten, URIs kreiert werden, welche mit einer Domäne beginnen, über die die ausliefernde Institution verfügen kann.
3. Begriffe, Konzepte, Schlagworte, usw. sollten nicht als Strings angegeben werden, sondern durch die Verwendung der entsprechenden URI, die auf einen Eintrag in einem kontrollierten Vokabular wie [Getty AAT](#), [Iconclass](#), [Wikidata](#), [GeoNames](#) oder [VIAF](#) verweisen.
4. URIs müssen generell [RFC 3986](#) folgen und dürfen demnach u. a. keine Leerzeichen enthalten. Sonderzeichen wie Leerzeichen müssen enkodiert werden. URIs dürfen höchstens einmal enkodiert sein.

5. Zu den Metadaten sollten Sprach-Tags hinzugefügt werden. Diese geben die Sprache des Textinhalts eines Feldes an. Bei mehreren Sprachen in einem Text kann maximal eine Sprache festgehalten werden, hierbei sollte die „hauptsächliche“ Sprache verwendet werden. Sprach-Tags müssen [BCP 47](#) konform sein.
6. Sprach-Tags sind getrennt vom Inhalt eines Textfelds zu führen. D. h. in einer konkreten Datenbank sollte der Sprach-Tag nicht als Teil des Textes innerhalb des Feldes erfasst, sondern in einem getrennten Feld festgehalten werden.
7. Schlagworte und Begriffe sollten nicht als Listen innerhalb eines Feldes geführt werden, sondern das Feld mit je nur einem Begriff so oft wie nötig wiederholt werden.
8. Textfelder für Schlagworte sollten keine Encodings, Sonderzeichen oder Markup enthalten.
9. Fließtextfelder sollten ebenfalls keine Encodings, Sonderzeichen oder Markup (wie z. B. HTML, TEI, Markdown) enthalten. Der einzige zulässige Spezialfall ist dabei der Newline-Character zur Kennzeichnung von Zeilenumbrüchen.
10. Weitere Standards, wie etwa die ISO-Standards, helfen bei der Vereinheitlichung und der damit einhergehenden besseren Auffindbarkeit der Objekte.
Ein Feld, in dem Daten einem ISO-Standard entsprechend festgehalten werden, sollte von einem zweiten Feld begleitet werden, in dem der entsprechende Standard festgehalten wird. Zumindest sollte der entsprechende Standard Teil der strukturierten Daten sein.
11. Beispielsweise normiert der Standard [ISO 8601-1](#) die Zeit- bzw. Datumsangabe. Die Erweiterung ISO 8601-2 erlaubt es auch, dass ungefähre oder unbekannte zeitliche Angaben strukturiert und sinnvoll erfasst und auffindbar werden.
12. Das letzte Bearbeitungsdatum eines Datensatzes sollte automatisch erfasst und mit dem Datensatz ausgeliefert werden können.

e

r

Version #10

Erstellt: 23 September 2024 06:12:16 von EW

Zuletzt aktualisiert: 16 Oktober 2024 09:06:06 von JP