

DIN ISO 26162-3

ICS 01.020; 35.240.30

Einsprüche bis 2023-06-14

Entwurf**Management von Terminologieressourcen –
Terminologiedatenbanken –
Teil 3: Inhalt (ISO 26162-3:2023);
Text Deutsch und Englisch**

Management of terminology resources –
Terminology databases –
Part 3: Content (ISO 26162-3:2023);
Text in German and English

Gestion des ressources terminologiques –
Bases de données terminologiques –
Partie 3: Contenu (ISO 26162-3:2023);
Texte en allemand et anglais

Anwendungswarnvermerk

Dieser Norm-Entwurf mit Erscheinungsdatum 2023-04-14 wird der Öffentlichkeit zur Prüfung und Stellungnahme vorgelegt.

Weil die beabsichtigte Norm von der vorliegenden Fassung abweichen kann, ist die Anwendung dieses Entwurfs besonders zu vereinbaren.

Stellungnahmen werden erbeten

- vorzugsweise online im Norm-Entwurfs-Portal von DIN unter www.din.de/go/entwuerfe bzw. für Norm-Entwürfe der DKE auch im Norm-Entwurfs-Portal der DKE unter www.entwuerfe.normenbibliothek.de, sofern dort wiedergegeben;
- oder als Datei per E-Mail an nat@din.de möglichst in Form einer Tabelle. Die Vorlage dieser Tabelle kann im Internet unter www.din.de/go/stellungnahmen-norm-entwuerfe oder für Stellungnahmen zu Norm-Entwürfen der DKE unter www.dke.de/stellungnahme abgerufen werden;
- oder in Papierform an den DIN-Normenausschuss Terminologie (NAT), 10772 Berlin oder Am DIN-Platz, Burggrafenstr. 6, 10787 Berlin.

Die Empfänger dieses Norm-Entwurfs werden gebeten, mit ihren Kommentaren jegliche relevanten Patentrechte, die sie kennen, mitzuteilen und unterstützende Dokumentationen zur Verfügung zu stellen.

Gesamtumfang 55 Seiten

DIN-Normenausschuss Terminologie (NAT)

Inhalt

	Seite
Nationales Vorwort	3
Nationaler Anhang NA (informativ) Literaturhinweise	4
Vorwort	5
Einleitung	6
1 Anwendungsbereich	7
2 Normative Verweisungen	7
3 Begriffe	7
4 Identifizierung von Benennungen	10
4.1 Anforderungen an die Auswahl von Benennungen	10
4.2 Unithood und Termhood	11
4.3 Korpora und Terminologieextraktion	12
5 Sammlung von terminologischen Daten	12
5.1 Anforderungen an Daten	12
5.1.1 Bewertungsverfahren	12
5.1.2 Datenqualität	13
5.1.3 Zweck der Termbank	13
5.1.4 Datenrichtigkeit	13
5.1.5 Gebrauchstauglichkeit	14
5.2 Datenmodell	15
5.3 Datenkategorien	15
5.4 Datenübertragbarkeit	15
6 Validierung der Begriffseintragsqualität	15
6.1 Allgemeine Validierungskriterien	15
6.2 Fehlertypologie und Systemdesign	15
6.3 Fehlertypen	16
6.3.1 Termbank-Spezifizierung und -Pflege	16
6.3.2 Fehlertypen, die der Aufnahme von Begriffseinträgen zugeordnet sind	16
6.3.3 Fehlertypen, die automatisch erzeugten Inhalten zugeordnet sind	16
6.3.4 Fehlertypen, die offenen Datenkategorien zugeordnet sind	17
6.3.5 Fehlertypen, die geschlossenen Datenkategorien zugeordnet sind	17
Anhang A (informativ) Bewertungsmodelle	18
A.1 Scoring-Modelle	18
A.2 Prüfen und Bearbeiten von Begriffseinträgen	18
A.3 Gewichtete Scoringsskalen	18
Anhang B (informativ) Beispiel für einen Begriffseintrag	20
Anhang C (informativ) Beispiel für eine Fehlertypologie	24
Literaturhinweise	30

Bilder

Bild 1 — Termhood	11
-----------------------------	----

Tabellen

Tabelle A.1 — Beispiel eines Antwortbogens für Sprachblöcke	19
Tabelle C.1 — Fehlertypen	24

Nationales Vorwort

Diese Deutsche Norm ist die Übersetzung der Internationalen Norm ISO 26162-3:2023, die im Technischen Komitee ISO/TC 37 „*Language and terminology*“ erarbeitet wurde.

Das zuständige nationale Normungsgremium ist der Arbeitsausschuss NA 105-00-05 AA „Management von Terminologieressourcen“ im DIN-Normenausschuss Terminologie (NAT).

Um Zweifelsfälle in der Übersetzung auszuschließen, ist die englische Originalfassung beigelegt. Die Nutzungsbedingungen für den deutschen Text des Norm-Entwurfes gelten gleichermaßen auch für den englischen Text.

Für die in diesem Dokument zitierten Dokumente wird im Folgenden auf die entsprechenden deutschen Dokumente hingewiesen:

ISO 9001	siehe	DIN EN ISO 9001
ISO 26162-1	siehe	DIN ISO 26162-1
ISO 26162-2:2019	siehe	DIN ISO 26162-2:2020-05

Aktuelle Informationen zu diesem Dokument können über die Internetseiten von DIN (www.din.de) durch eine Suche nach der Dokumentennummer aufgerufen werden.

Nationaler Anhang NA (informativ)

Literaturhinweise

DIN EN ISO 9001, *Qualitätsmanagementsysteme — Anforderungen*

DIN ISO 26162-1, *Management von Terminologieressourcen — Terminologiedatenbanken — Teil 1: Design*

DIN ISO 26162-2:2020-05, *Management von Terminologieressourcen — Terminologiedatenbanken — Teil 2: Software (ISO 26162-2:2019)*

Vorwort

ISO (die Internationale Organisation für Normung) ist eine weltweite Vereinigung nationaler Normungsinstitute (ISO-Mitgliedsorganisationen). Die Erstellung von Internationalen Normen wird üblicherweise von Technischen Komitees von ISO durchgeführt. Jede Mitgliedsorganisation, die Interesse an einem Thema hat, für welches ein Technisches Komitee gegründet wurde, hat das Recht, in diesem Komitee vertreten zu sein. Internationale staatliche und nichtstaatliche Organisationen, die in engem Kontakt mit ISO stehen, nehmen ebenfalls an der Arbeit teil. ISO arbeitet bei allen elektrotechnischen Normungsthemen eng mit der Internationalen Elektrotechnischen Kommission (IEC) zusammen.

Die Verfahren, die bei der Entwicklung dieses Dokuments angewendet wurden und die für die weitere Pflege vorgesehen sind, werden in den ISO/IEC-Direktives, Teil 1 beschrieben. Es sollten insbesondere die unterschiedlichen Annahmekriterien für die verschiedenen ISO-Dokumentenarten beachtet werden. Dieses Dokument wurde in Übereinstimmung mit den Gestaltungsregeln der ISO/IEC-Direktives, Teil 2 erarbeitet (siehe www.iso.org/directives).

Es wird auf die Möglichkeit hingewiesen, dass einige Elemente dieses Dokuments Patentrechte berühren können. ISO ist nicht dafür verantwortlich, einige oder alle diesbezüglichen Patentrechte zu identifizieren. Details zu allen während der Entwicklung des Dokuments identifizierten Patentrechten finden sich in der Einleitung und/oder in der ISO-Liste der erhaltenen Patenterklärungen (siehe www.iso.org/patents).

Jeder in diesem Dokument verwendete Handelsname dient nur zur Unterrichtung der Anwender und bedeutet keine Anerkennung.

Für eine Erläuterung des freiwilligen Charakters von Normen, der Bedeutung ISO-spezifischer Begriffe und Ausdrücke in Bezug auf Konformitätsbewertungen sowie Informationen darüber, wie ISO die Grundsätze der Welthandelsorganisation (WTO, en: World Trade Organization) hinsichtlich technischer Handelshemmnisse (TBT, en: Technical Barriers to Trade) berücksichtigt, siehe www.iso.org/iso/foreword.html.

Dieses Dokument wurde vom Technischen Komitee ISO/TC 37, *Language and terminology*, Unterkomitee SC 3, *Management of terminology resources* erarbeitet.

Eine Auflistung aller Teile der Normenreihe ISO 26162 ist auf der ISO-Internetseite abrufbar.

Rückmeldungen oder Fragen zu diesem Dokument sollten an das jeweilige nationale Normungsinstitut des Anwenders gerichtet werden. Eine vollständige Auflistung dieser Institute ist unter www.iso.org/members.html zu finden.

Einleitung

Manager, Pädagogen und Zuständige für die Unterhaltung von Terminologiedatenbanken führen sowohl in regelmäßigen Abständen wie auch kontinuierlich Bewertungen von Terminologiedatenbanken mit Begriffseinträgen zu unterschiedlichen Zwecken durch:

- Validierung von terminologischen Datensammlungen in Wirtschafts-, Regierungs- und Nichtregierungsbereichen im Zusammenhang mit Qualitätssicherung;
- Bewertung aus pädagogischer Sicht und zusammenfassende Bewertung und Rückmeldung in Ausbildungs- und Schulumgebungen.

ISO 26162-1 und ISO 26162-2 geben Gestaltungsprinzipien und Überlegungen hinsichtlich Software zur Modellierung von Terminologiedatenbanken (Termbanken) vor. ISO 26162-1 legt die allgemeinen Prinzipien für die Gestaltung von Termbanken fest, wie diese in grundlegenden ISO/TC 37-Normen wie ISO 704, in der neben anderen Themen die Grundprinzipien für Begriffseintragsinhalte und -strukturen, Identifizierung von Benennungen, Grundprinzipien für die Modellierung von Begriffssystemen und eine Reihe anderer Bereiche im Zusammenhang mit Terminologearbeit behandelt werden, aufgeführt sind. ISO 26162-1 fördert zudem die Übereinstimmung mit dem terminologischen Metamodell, das in ISO 16642 aufgeführt ist. Das Dokument beschreibt die Rolle, die Datenkategorien bei der Modellierung von terminologischen Daten spielen, und legt Grundprinzipien für die Sicherung und Bewertung der Qualität von Daten, die in Termbanken gespeichert werden, wie Datengranularität, -elementarität und -modellierungsvarianz, fest. Diese Kriterien umfassen wesentliche Bezugsgrößen für die Messung der Qualität und Zuverlässigkeit von terminologischen Daten. ISO 26162-2 setzt die Prinzipien, die in ISO 26162-1 aufgeführt sind, zur Umsetzung von Datenbankgestaltung im Hinblick auf Überlegungen zu Software und Benutzerschnittstellen in Beziehung und gibt daneben Arbeitsabläufe zur pragmatischen Umsetzung in Terminologieverwaltungsumgebungen vor.

Dieses Dokument dient als Leitfaden für die Definition von Verfahren zur Sicherstellung von qualitativ hochwertigen Inhalten in Terminologiedatensammlungen, die dazu entwickelt wurden, Dokumentationsbedürfnisse in unterschiedlichen Bereichen wie beispielsweise Übersetzen, Dolmetschen und Technische Kommunikation zu erfüllen. Die Konformität mit diesem Dokument kann Prozesse stärken, die dazu entwickelt wurden, ein Qualitätsmanagementsystem wie beispielsweise ISO 9001 und die damit zusammenhängenden Auditverfahren in den Bereichen Übersetzen, Dolmetschen oder Technische Kommunikation zu unterstützen. Eine Fehlertypologie wird im Rahmen eines Gesamtbewertungsmodells mit generischen (nicht normierten) Optionen für die Erstellung eines Modells zur Evaluierung von Begriffseinträgen in Abhängigkeit von den Bedürfnissen der Nutzer und Trägerorganisationen vorgelegt.

Anhang A, Anhang B und Anhang C enthalten praxisrelevante Hinweise zu Fehlerbewertungsmethoden. Anhang A beschreibt die Erstellung von Scoring-Modellen, die die in Abschnitt 6 beschriebene Fehlertypologie wiedergeben, wobei Abweichungen bei der Gestaltung in Abhängigkeit von den Bedürfnissen der jeweiligen Organisation möglich sind. Beispielsweise kann ein vorgegebenes Scoring-Modell die Grundlage für einen Antwortbogen bilden, der für Studierende und Auszubildende verwendet wird und der sich voraussichtlich von Antwortbögen, wie sie für Großunternehmen oder nationale Termbanken verwendet werden, unterscheidet.

Anhang B enthält ein Beispiel für einen Begriffseintrag. Anhang C gibt ein Beispiel für ein Bewertungsmodell wieder, das übernommen oder angepasst werden kann, um die Bedürfnisse von Terminologen, Einzelpersonen, die als Freiberufler oder in Unternehmen, Regierungsorganisationen und NGOs arbeiten, zu erfüllen. Die Werte in diesem Bewertungsmodell können verwendet werden, um ein Scoring-Verfahren zu erstellen, wobei sich versteht, dass die tatsächliche Scoring-Praxis entsprechend Spezifikationen und Anwenderbedürfnissen voraussichtlich abweicht.

1 Anwendungsbereich

Dieses Dokument legt inhaltsbezogene Aspekte der Pflege von Terminologiedatenbanken fest. Es gibt Leitlinien zum Inhalt von terminologischen Datensammlungen vor, wobei der Fokus auf der Bewertung der Datenqualität liegt.

Dieses Dokument enthält Leitlinien für Entwickler von Begriffseinträgen, die Interoperabilität und qualitativ hochwertige Inhalte sicherstellen müssen. Es beabsichtigt eine Sicherstellung, dass terminologische Datensammlungen ihrerseits hohe Standards bei der Gestaltungskonformität mit Normen wie ISO 12620-1 und ISO 16642, Datengenauigkeit und Ergebnissen erfüllen. Es beschreibt Prinzipien zur Sicherstellung der Datenqualität (siehe ISO 9001) und zur Bewertung von terminologischen Datensammlungen mit dem Ziel der kontinuierlichen Verbesserung. Dieses Konzept steht im Kontrast zu dem in ISO 23185:2009, das sich auf die Gebrauchstauglichkeit von bestehenden Terminologieressourcen konzentriert.

Dieses Dokument gilt nicht für die Verwaltung von Textkorpora oder für Terminologieextraktionswerkzeuge.

2 Normative Verweisungen

Die folgenden Dokumente werden im Text in solcher Weise in Bezug genommen, dass einige Teile davon oder ihr gesamter Inhalt Anforderungen des vorliegenden Dokuments darstellen. Bei datierten Verweisungen gilt nur die in Bezug genommene Ausgabe. Bei undatierten Verweisungen gilt die letzte Ausgabe des in Bezug genommenen Dokuments (einschließlich aller Änderungen).

ISO 1087, *Terminology work and terminology science — Vocabulary*

ISO 26162-1, *Management of terminology resources — Terminology databases — Part 1: Design*

ISO 26162-2:2019, *Management of terminology resources — Terminology databases — Part 2: Software*

3 Begriffe

Für die Anwendung dieses Dokuments gelten die Begriffe nach ISO 1087 und die folgenden Begriffe.

ISO und IEC stellen terminologische Datenbanken für die Verwendung in der Normung unter den folgenden Adressen bereit:

— ISO Online Browsing Platform: verfügbar unter <https://www.iso.org/obp>

— IEC Electropedia: verfügbar unter <https://www.electropedia.org/>

3.1

Begriffseintrag

Eintrag

terminologischer Eintrag

Teil einer *Terminologiedatensammlung* (3.14), der die terminologischen Daten enthält, die sich auf einen Begriff beziehen

Anmerkung 1 zum Begriff: Ein Begriffseintrag kann Informationen enthalten, die sich auf eine oder mehrere Sprachen beziehen.

Anmerkung 2 zum Begriff: In diesem Dokument wird die Benennung *Eintrag* als Kurzform von *Begriffseintrag* verwendet.

[QUELLE: ISO 30042:2019, 3.5, modifiziert — „Eintrag“ als zweite Vorzugsbenennung aufgeführt, Anmerkungen zum Eintrag hinzugefügt.]

3.2

Bewertungsmodell

Modell zur Analyse von Daten innerhalb eines *Begriffseintrags* (3.1) nach Terminologieprinzipien und vorgegebenen Datenanforderungen, die dem Zweck der *Termbank* (3.11) entsprechen

3.3

Kernstruktur

gemeinsame Struktur und *Datenkategorien* (3.6), die für alle *TBX-Dialekte* (3.5) verwendet werden

Anmerkung 1 zum Begriff: Die Kernstruktur entspricht ISO 16642 (TMF).

[QUELLE: ISO 30042:2019, 3.6]

3.4

Fehlertypologie

systematische Auflistung von Fehlertypen

3.5

TBX-Dialekt

Version der Auszeichnungssprache eXtensible Markup Language (XML) zur Validierung nach der *Kernstruktur* (3.3) von TermBase eXchange (TBX), die genau die *Datenkategorien* (3.6) auf den Ebenen, die durch eine präzise definierte Konfiguration von Datenkategorien vorgegeben sind, erlaubt

Anmerkung 1 zum Begriff: Siehe ISO 30042 hinsichtlich weiterer Einzelheiten.

[QUELLE: ISO 30042:2019, 3.12, modifiziert — Für dieses Dokument vereinfacht, Anmerkung 1 zum Begriff ausgetauscht.]

3.6

Datenkategorie

Klasse formal oder semantisch ähnlicher Datenelemente

BEISPIEL /Wortklasse/, /Fachgebiet/, /Definition/.

Anmerkung 1 zum Begriff: Eine Datenkategorie kann als Verallgemeinerung der Bedeutung eines Datenfeldes in einer Datenbank betrachtet werden.

Anmerkung 2 zum Begriff: Datenkategorienamen werden im Fließtext, wie in diesem Dokument, zwischen Schrägstrichen angegeben, z. B. /Wortklasse/.

[QUELLE: ISO 30042:2019, 3.8]

3.7

Datenintegrität

Übereinstimmung von Datenwerten mit einem vorgegebenen Regelwerk

[QUELLE: ISO/IEC TR 10032:2003, 2.23]

3.8

lexikalische Einheit

Spracheinheit, die dem Wortschatz einer bestimmten Sprache angehört

[QUELLE: ISO 1951:2007, 3.8, modifiziert — Verweisung auf ein Wörterbuch entfernt.]

3.9

Qualitätssicherung

QA, en: quality assurance

Satz von geplanten und systematischen Aktivitäten, die notwendig sind, um sicherzustellen, dass ein *Begriffseintrag* (3.1) die Akzeptanzkriterien auf Grundlage von Terminologieprinzipien und vorgegebenen Datenanforderungen erfüllt

3.10

Spezifikation

Dokument, das die Anforderungen, die von einer *Terminologiedatensammlung* (3.14) zu erfüllen sind, ausführlich aufführt

Anmerkung 1 zum Begriff: Spezifikationen können Verfahren für das Prüfen der Übereinstimmung mit diesen Anforderungen einschließen.

3.11

Termbank

Terminologiedatenbank

Datenbank, die eine *Terminologiedatensammlung* (3.14) umfasst

[QUELLE: ISO 30042:2019, 3.28, modifiziert — „Terminologiedatenbank“ nicht mehr nur als zugelassene Benennung, sondern als zweite Vorzugsbenennung.]

3.12

Termbank-Qualitätsbewerter

Person, die als Terminologe oder Experte des jeweiligen Sachgebiets qualifiziert ist und eine Qualitätsbewertung einer *Terminologiedatensammlung* (3.14) vornimmt

3.13

Termhood

Grad der terminologischen Relevanz einer *lexikalischen Einheit* (3.8)

BEISPIEL „Massengutfrachter“ weist eine stärkere Termhood als „Schiff“ an sich auf. „Maus“ verfügt im Bereich Computeranwendungen über Termhood, jedoch nicht in der Allgemeinsprache.

Anmerkung 1 zum Begriff: Termhood ist sowohl auf einfache Benennungen (die aus Einzelwörtern bestehen) und komplexe Benennungen (die aus mehr als einem Wort oder lexikalischer Einheit gebildet sind) als auch weitere Bezeichnungen wie Eigennamen und Anreden sowie Formeln und Symbole anwendbar.

3.14

Terminologiedatensammlung

Ressource, die aus *Begriffseinträgen* (3.1) mit zugehörigen Metadaten und dokumentarischen Informationen besteht

BEISPIEL TBX-Dokumentinstanz, ISO 1087.

[QUELLE: ISO 30042:2019, 3.29, modifiziert — zweite Vorzugsbenennung „TDC“ entfernt.]

3.15

Unithood

Grad, in dem eine bestimmte Abfolge von Wörtern über ausreichende Kollokationsstärke verfügt, um eine stabile *lexikalische Einheit* (3.8) zu bilden

BEISPIEL „Art-déco-Tisch“ weist eine höhere Unithood als „moderner Tisch“ auf.

Anmerkung 1 zum Begriff: Da sich Unithood aus der Kollokationsbeziehung von Wörtern, die eine bestimmte Wortfolge ergeben, ableiten lässt, ist Unithood nur auf Mehrwortbenennungen anwendbar.

4 Identifizierung von Benennungen

4.1 Anforderungen an die Auswahl von Benennungen

Begriffseinträge sollten die Bedürfnisse ihrer beabsichtigten Zielgruppen und der bestimmungsgemäßen Verwendung sowie organisationsspezifische Anforderungen einschließlich der Anforderungen an Benennungen erfüllen. In ISO 704 werden Prinzipien zur Zuordnung und Analyse von Bezeichnungen erörtert.

Obwohl es nicht möglich ist, universelle Anforderungen für die Auswahl von Benennungen und für den Inhalt von Begriffseinträgen zu erstellen, ist ein wichtiges Ziel einer Termbank in vielen gewerblichen Umgebungen in seinem Wesen präskriptiv: Anwender von Benennungen, die aus irgendwelchen Gründen problematisch sind, weg und zur Verwendung von Vorzugsbenennungen hin zu führen. Somit umfasst die Auswahl von Benennungen sowohl das Identifizieren der Vorzugsbenennungen innerhalb der Organisation sowie das Dokumentieren der entsprechenden synonymen Benennungen, die vermieden werden sollen. In diesem Kontext sollten Umsetzer neben Vorzugsbenennungen sowohl veraltete und obsoletere Benennungen als auch nebenrangige Benennungen, die gleichwohl in wichtigen Unternehmenskontexten wiederholt verwendet werden, dokumentieren.

Verbreitete Kriterien zum Aufnehmen einer Benennung in die Termbank schließen ein:

- Personensicherheit: Wichtige Benennungen, deren fehlerhafte Verwendung die Sicherheit von Kunden oder Mitarbeitern gefährden könnte. Werden falsche Benennungen verwendet, kann dies zu nicht bestimmungsgemäßer Verwendung des Produkts führen, wodurch das Produkt beschädigt oder in manchen Fällen der Benutzer einem Risiko ausgesetzt sein könnte. Die gleichen Bedenken treffen auf Benennungen zu, die in Dokumenten zu Regierungsverordnungen verwendet werden, die den Zugang von Bürgern zu wichtiger Gesundheitsversorgung, Interaktion mit dem Justizsystem, Arbeitsschutz und -gesundheit und weiteren Ressourcen und persönlichen Dienstleistungen betreffen.

BEISPIEL 1 Benennungen im Zusammenhang mit pharmazeutischen Verordnungen, Sicherheitsausrüstung auf einer Ölplattform, Zusammenbau und Instandhaltung von Luftfahrzeug-Fahrwerken oder Kommunikation zwischen Schulen und Eltern.

- Unternehmenssicherung: Benennungen, die das Unternehmen oder die Organisation schützen. Eine Verwendung falscher Benennungen könnte zu Haftungsklagen oder dem Verlust von geistigem Eigentum führen.

BEISPIEL 2 Benennungen im Zusammenhang mit behördlichen Vorschriften zu einer Branche oder Patenten, die einem Unternehmen gehören.

- Unternehmensidentität: Benennungen dieser Kategorie geben Begriffe wieder, die von einem Unternehmen geschaffen wurden. Ihre korrekte Verwendung schützt die Marke und die Markenzeichen und stärkt den Marktanteil. Bei der Entscheidung für markenrelevante Benennungen und Handelsnamen überprüfen weltweit tätige Unternehmen zunehmend mögliche Übersetzungen auf ihre Eignung. Sicherzustellen, dass Benennungen die Stellung des Unternehmens als Branchenführer stärken und es aus der Masse der Wettbewerber hervorheben, gewinnt an Bedeutung. Dies betrifft Benennungen, die mit Produkten oder Dienstleistungen in Zusammenhang stehen oder die außerhalb ihrer gewöhnlichen Bedeutung verwendet werden.

BEISPIEL 3 Produktnamen und registrierte Markennamen; Komponenten patentierter Produkte; Namen von Merkmalen oder Dienstleistungen unterhalb der Markenebene.

- Sachgebiet: Benennungen innerhalb einer Branche oder eines Sachgebiets. Dies schließt Benennungen aus vertikalen Branchen oder Bereichen ein, in denen sich Experten (außerhalb einer festgelegten Organisation) über eine Gruppe von Benennungen einig sind.

BEISPIEL 4 Fachgebiete wie Buchhaltung, automobiles Kerngeschäft, Biologie.

- Organisationen des Öffentlichen Dienstes: Benennungen, die für eine bestimmte Zielgruppe verständlich sind.

BEISPIEL 5 Schriftstücke, die in einer Sprache für Empfänger mit geringeren Ressourcen geschrieben sind, müssen gelegentlich Terminologie verwenden, die sich von den im jeweiligen akademischen oder brancheninternen Umfeld von Fachleuten verwendeten Benennungen unterscheidet.

- Pragmatische und gebietsbezogene Sachverhalte: Alle Benennungen, denen eine Dokumentation oder Standardisierung zuträglich ist, einschließlich häufig verwendeter Benennungen oder schwieriger Benennungen, bei denen eine Dokumentation und Steuerung dieser Benennungen Fehler oder Nachfragen verringert. Wenn Produkte lokalisiert oder Dokumente in andere Sprachen übersetzt werden, deckt diese Kategorie auch Benennungen ab, die bei der Übersetzung eine Herausforderung darstellen können oder denen eine Standardisierung in einer oder mehreren Sprachen zuträglich wäre. Bei manchen Sprachen ist ein höheres Maß an Terminologiedokumentation vorteilhafter als bei anderen.

4.2 Unithood und Termhood

Wie in Bild 1 dargestellt ist, umfasst „Termhood“ die Schnittmenge von Unithood, Verwendung innerhalb des Korpus der Trägerorganisation und dem Zweck der Terminologiedatensammlung (siehe Literaturhinweis [16]).

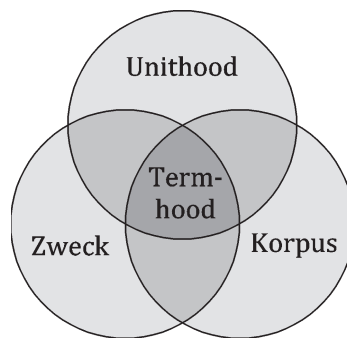


Bild 1 — Termhood

„Unithood“ bezieht sich auf Kollokationsbeziehungen zwischen Wörtern und ist nicht auf aus Einzelwörtern bestehende lexikalische Einheiten anwendbar. Die Festlegung, ob eine bestimmte aus mehreren Wörtern bestehende Wortfolge eine stabile lexikalische Einheit umfasst, die so häufig innerhalb des Organisations-Korpus auftaucht, dass sie eine Kollokation darstellt, ist für die Auswahl von Benennungen wichtig. Aufgrund ihrer vergleichsweise stabilen syntagmatischen Strukturen tendieren Mehrwort-Einheiten mit starker Unithood dazu, wichtige Kommunikationsstrukturen zu bilden, die in gewerblichen Umgebungen Konsistenz erfordern. Obwohl in gewerblichen und professionellen Umgebungen Mehrwortbenennungen vorherrschen, schließt dies die Dokumentation von Einwortbenennungen, die innerhalb des Korpus der Organisation über eine ausreichende Prominenz (Häufigkeit und/oder Verteilung) verfügen und dem Zweck der Terminologiedatensammlung dienlich sind, nicht aus.

Eine aus mehreren Wörtern bestehende lexikalische Einheit, die die Kriterien für Unithood erfüllt, fungiert als Benennung, wenn sie einen identifizierbaren Begriff innerhalb des jeweiligen textuellen und organisatorischen Kontextes bezeichnet. Um das Prinzip der Begriffsorientierung einzuhalten, muss jeder Begriff in einer Termbank als eigenständiger Begriffseintrag dokumentiert werden.

BEISPIEL Manchmal erfordert ein Wort in einer Sprache mehrere Begriffseinträge. So bezeichnet beispielsweise das englische Wort (das in manchen Kontexten als Benennung fungieren kann) „river“ zwei unterschiedliche Begriffe, die im Französischen durch „fleuve“ (Fluss, der in ein Meer oder einen Ozean fließt) und „rivière“ (Fluss, der in einen anderen Fluss oder einen See fließt) wiedergegeben werden. Die deutsche Benennung „Abstandsbolzen“ umfasst zwei begriffliche Referenzen, die im Englischen wie folgt unterschieden werden: „distance rivet“ und „spacer bolt“. Diese beiden Begriffe würden in einer Termbank zwei eigenständige Begriffseinträge erfordern.

4.3 Korpora und Terminologieextraktion

In gewerblichen Umgebungen stellt das Identifizieren von Benennungen üblicherweise einen Prozess dar, der durch den textualen Kontext, in dem die Benennungen auftreten, geprägt ist. Es wird empfohlen, für die Identifizierung von Benennungskandidaten Korpora zu verwenden. Eine Korpusanalyse informiert über die Tiefe oder Breite der Abdeckung des Sachgebiets, ohne die Termbank mit unnötigen Einträgen zu überfrachten. Korpora können aus einer beliebigen Art von Schriftmaterial bestehen, darunter Marketingmaterial, Produktdokumentation, interne Mitteilungen, Transkripte, zweisprachige Übersetzungsspeicher oder sonstige Sammlungen von Textmaterial, das von der Trägerorganisation erstellt wurde. Ein organisationsweites Korpus liefert Nachweise über die Häufigkeit des Auftretens bestimmter Benennungskandidaten sowie über ihre Verteilung innerhalb unterschiedlicher Arten von Textmaterial. Sowohl Häufigkeit als auch Verteilung stellen Kriterien für die Entscheidung dar, ob ein Kandidat die Kriterien für Termhood erfüllt.

BEISPIEL 1 Gewisse selten vorkommende lexikalische Einheiten können für eine Organisation wichtige Benennungen darstellen, wenn sie innerhalb unterschiedlicher Textsorten (Marketing, Vertrieb, Schulung, Onlinetexte, Benutzerhandbücher usw.) vorkommen oder sie trotzdem wichtige Begriffe wiedergeben.

Verschiedene Werkzeuge und Technologien sind verfügbar, um die Extraktion lexikalischer Einheiten zum Zweck der Identifizierung von Benennungen zu unterstützen (siehe auch ISO 26162-2 sowie ISO 12616-1 für zusätzliche Informationen zu Werkzeugen für die Termextraktion).

ANMERKUNG Momentan ist zusätzlich ISO 5078 [3] zur Terminologieextraktion in Entwicklung.

Zwei Schlüsselfaktoren, die bei der Auswahl relevanter Begriffe und Benennungen zu berücksichtigen sind, umfassen:

- Marketing: wichtige Begriffe zur Wiedererkennung von Marken, die eine Organisation von ihren Wettbewerbern abheben;

BEISPIEL 2 Begriffe und damit zusammenhängende Benennungen, die patentierten und markengeschützten Produkten oder Prozessen zugeordnet sind.

- Kundenzufriedenheit: Bereiche, in denen Benennungen und Begriffe in der Vergangenheit zu Problemen geführt haben und die Qualitätssicherungsprobleme in der Vergangenheit oder sonstige ernsthafte Risikokriterien widerspiegeln können.

BEISPIEL 3 Besondere Themen, die zu einer höheren Anzahl an Anrufen beim Kundendienst oder Kundenreklamationen geführt haben.

5 Sammlung von terminologischen Daten

5.1 Anforderungen an Daten

5.1.1 Bewertungsverfahren

Das Bewertungsverfahren, das in diesem Dokument vorgeschlagen wird, wurde entwickelt, um die Qualität einer Termbank auf Grundlage der Bewertung der enthaltenen Begriffseinträge und den diesen Begriffen zugeordneten Benennungen sowie die Richtigkeit der in Zusammenhang stehenden Daten und ihre Gebrauchstauglichkeit zu messen. Die Richtigkeit und Gebrauchstauglichkeit werden im Hinblick auf die Übereinstimmung mit dem Termbank-Datenmodell und entsprechenden Spezifikationen beurteilt.

Qualitätsdaten erfüllen die Anforderungen für den beabsichtigten Zweck (siehe ISO 8000-1). Der Zweck muss deutlich ausgedrückt sein und muss die Grundlage für die Spezifizierung der Datenanforderungen umfassen. Datenanforderungen müssen an den Bedürfnissen der Trägerorganisation ausgerichtet sein und müssen für Nutzer und Beitragende der Termbank zugänglich gemacht werden. Datenrichtigkeit und Gebrauchstauglichkeit können als Funktion der Übereinstimmung mit Spezifikationen gemessen werden.

5.1.2 Datenqualität

Daten müssen den bewährten Verfahren, die in ISO 26162-1 und ISO 26162-2 beschrieben sind, entsprechen und müssen:

- festgelegte Anforderungen an die Terminologiedatensammlung erfüllen;
- auf ein stabiles Datenmodell verweisen;
- explizit definierte Datenkategorien verwenden, die mit dem Datenmodell konsistent sind;
- wie in ISO 26162-2 spezifiziert übertragbar sein.

5.1.3 Zweck der Termbank

Der Zweck der Terminologiedatensammlung muss bezogen auf die Bedürfnisse der Organisation spezifiziert werden. Eine Termbank, die in erster Linie dazu verwendet wird, Autoren- oder Übersetzungsprozesse zu unterstützen, erfordert anders geartete Einträge als solche, die für eine Termbank ausgewählt werden, die zum Zweck der Optimierung von Suchergebnissen oder für ontologische Modellierung entwickelt wurde, wobei sich diese Zwecke nicht notwendigerweise gegenseitig ausschließen. Zum Beispiel:

- Diskurs-orientierte Termbanken bieten oft eine Vielzahl an Begriffseinträgen, die ein breites Spektrum an Sachgebieten widerspiegeln. Beispielsweise kann ein Produktionsunternehmen neben grundlegenden produkt- und verfahrensrelevanten Begriffen auch für das Personalmanagement, die Buchhaltung oder das Gebäudemanagement häufig verwendete Begriffe dokumentieren.
- Produkt- und verfahrensrelevante Einträge können mit einem zentralen Produktkatalog oder mit Computersystemen, die für Logistik und Prozesssteuerung entwickelt wurden, eng verlinkt werden.
- Informationsmanagementsysteme und ontologie-orientierte Systeme können mit offenen Datenverweisen innerhalb des Unternehmens oder sogar mit offenem Zugang zum semantischen Web entwickelt werden.

5.1.4 Datenrichtigkeit

5.1.4.1 Datenrichtigkeit stellt eine Funktion der Datenvalidierung nach der Definition in ISO 1087:2019, 3.6.6 dar. Sie kann mit Folgendem in Zusammenhang stehen:

- Inhalt einzelner Datenfelder (siehe 5.1.4.2);
- Beziehungen zwischen mehreren Datenfeldern innerhalb eines Begriffseintrags (siehe 5.1.4.3); oder
- Beziehungen zwischen mehreren Begriffseinträgen (siehe 5.1.4.4 und „Konsistenzprüfung“ in ISO 1087:2019, 3.6.8).

5.1.4.2 Daten, die innerhalb eines Begriffseintrags bereitgestellt werden:

- müssen den Spezifikationen für jedes Datenfeld, wie diese von der Organisation entsprechend den Anforderungen und Zwecken der Termbank festgelegt wurden, entsprechen; dieses Kriterium gibt die in ISO 1087:2019, 3.6.11 und 3.6.10 aufgeführten Prüfungen von „Format“ und „Vollständigkeit“ ziemlich genau wieder;

BEISPIEL Ein bestimmtes Datenmodell kann die folgenden Regeln haben: „eine Definition muss nach den Regeln in ISO 704 formuliert sein“; „der Inhalt einer Benennungsanmerkung muss organisationsspezifischen Leitlinien entsprechen“; „nach den Leitlinien der TBX-Basic-Spezifizierung ist ein Kontext erforderlich“. (Siehe Literaturhinweis [15].)

- muss die Konventionen, die für die entsprechende Sprache angemessen sind, z. B. Geschlecht, Rechtschreibformen, Schriftpräferenzen, Register, Formalien, einhalten.

5.1.4.3 Hinsichtlich der Angemessenheit des Inhalts von Datenfeldern innerhalb eines Eintrags im Hinblick auf einander:

- müssen alle Benennungen, die in die /Benennung/-Felder eines Eintrags eingetragen werden, den jeweiligen Begriff wiedergeben;
- müssen alle Benennungen, die Synonyme oder Äquivalente innerhalb des Begriffs darstellen, in eigene /Benennung/-Felder eingetragen werden und dürfen nicht in anderen Feldern versteckt werden, wo sie für Filter oder Suchroutinen nicht zugänglich sind;
- müssen Definitionen, wenn sie in verschiedenen Sprachen existieren, denselben Begriff definieren;
- müssen Werte in Datenfeldern, die mit Benennung, Sprache und Begriff zusammenhängen, jeweils für die Benennung, Sprache und den Begriff richtig sein; so ist beispielsweise die Zuordnung /Wortklasse/ = „Substantiv“ nur dann richtig, wenn sie einem Substantiv zugeordnet wird.
- Synonyme und Äquivalente, die einem bestimmten Begriff zugeordnet sind, müssen innerhalb eines einzigen Eintrags aufgeführt sein; Duplikateinträge müssen zusammengeführt oder entfernt werden (siehe „Prüfung auf doppelte Einträge“, ISO 1087:2019, 3.6.7).
- Homophone, die unterschiedliche Begriffe wiedergeben, müssen in unterschiedlichen Begriffseinträgen behandelt werden; wenn die gleiche lexikalische Einheit sowohl ein Substantiv als auch ein Verb wiedergibt, müssen die Benennungen in unterschiedlichen Begriffseinträgen wiedergegeben werden.

5.1.4.4 Hinsichtlich der Konsistenz von Daten, die über mehrere Begriffseinträge verteilt sind:

- müssen explizite Beziehungen zwischen mehreren Begriffseinträgen korrekt identifiziert werden (siehe „Konsistenzprüfung“ in ISO 1087:2019, 3.6.8);
- muss der Inhalt einzelner Datenfelder in allen Einträgen konsistent sein;
- müssen Verlinkungen intakt gehalten werden, d. h. wenn ein Eintrag entfernt wird, müssen bestehende Verlinkungen ebenfalls entfernt werden, wie in ISO 26162-2:2019, 4.7 aufgeführt ist.

5.1.5 Gebrauchstauglichkeit

Um gebrauchstauglich zu sein, müssen terminologische Daten in die richtigen Datenfelder eingetragen werden (z. B. befinden sich Definitionen in /Definition/-Feldern, Anmerkungen in /Anmerkung/-Feldern).

Abhängig vom terminologischen Arbeitsablauf müssen für jeden Eintrag oder Teile von Einträgen die aktuellen Lebenszyklusphasen in Übereinstimmung mit Systemspezifikationen angegeben werden (z. B. unbearbeitet, vorläufig bearbeitet, abgeschlossen).

Daten sollten eine Klassifizierung von Unterabschnitten der Terminologiedatenbank beispielsweise wie folgt ermöglichen:

- Kategorisierung von Sachgebieten innerhalb der jeweiligen Fachrichtung;
- Kunden oder Projekte innerhalb der Organisationsabläufe;
- Sicherheitsanforderungen;

BEISPIEL 1 Beschränkung der Verteilung aufgrund von Sicherheitsrisiken durch Einordnen von Begriffen als öffentlich oder vertraulich.

- Anforderungen hinsichtlich geistiger Eigentumsrechte.

BEISPIEL 2 Beschränkungen, die für eine Benennung vorgegeben werden, um das Recht einer Organisation auf exklusive Verwendung der Benennung zu schützen (z. B. Markenzeichen, Dienstleistungsmarke).

5.2 Datenmodell

Das Datenmodell muss die Bedürfnisse der Organisation erfüllen und den Zweck der Termbank widerspiegeln.

Jeder Begriffseintrag muss die Mindestanzahl an Datenfeldern für Gebrauchstauglichkeit verfügbar machen, was bedeuten kann, dass Datenfelder eingeschlossen sein können, die nur unter bestimmten Umständen verwendet werden.

BEISPIEL 1 Ein Datenfeld kann für /transfer comment/ eingefügt werden, würde aber nur in Fällen verwendet, in denen Äquivalenzbeziehungen eine Erklärung erfordern.

BEISPIEL 2 Eine Benennung innerhalb eines Begriffseintrags wird nicht als /Vollform/ markiert, wenn dieser Begriff nicht auch mit einer /Abkürzung/ wiedergegeben wird.

Wenn eine Terminologiedatensammlung als mit einem bekannten Datenmodell wie einem TBX-Dialekt konform identifiziert wurde, muss die Struktur mit diesem Modell übereinstimmen.

5.3 Datenkategorien

Begriffseinträge müssen explizit definierte Datenkategorien verwenden, die mit dem angegebenen Datenmodell konsistent sind. Anforderungen an einzelne Datenkategorien werden von der Organisation entsprechend ihren Bedürfnissen und dem Zweck der Termbank festgelegt.

Die Namen von Datenkategorien müssen die spezifizierten Datennamen im angegebenen Datenmodell einhalten oder es müssen eindeutige Leitfäden zur Zuordnung von lokalen Namen und Werten zu Werten, die im angegebenen Datenmodell verwendet werden, bereitgestellt werden. Gibt beispielsweise eine Terminologiedatensammlung einen TBX-Dialekt als ihr Datenmodell an, müssen die Datenkategorien diesem Dialekt entsprechen oder diesem Dialekt zugeordnet werden (siehe ISO 30042).

ANMERKUNG Die Definitionen von Datenkategorien sind in Verweisung [14] zu finden.

5.4 Datenübertragbarkeit

Die Terminologiedatensammlung sollte übertragbar und von anderen Anwendungen aus zugänglich sein, wie dies in ISO 26162-2:2019, Abschnitt 4 vorgegeben ist.

6 Validierung der Begriffseintragsqualität

6.1 Allgemeine Validierungskriterien

Die Qualität von Begriffseinträgen wird durch ein Bewerten des Inhalts im Hinblick auf Systemanforderungen validiert.

6.2 Fehlertypologie und Systemdesign

Im Hinblick auf Konsistenz und Interrater-Reliabilität wird empfohlen, Fehlertypologien zu verwenden, die mit dem Termbank-Datenmodell konsistent sind. Dieses Dokument legt ausführliche Kriterien für die Bewertung von Begriffseinträgen fest und stellt ein veranschaulichendes Beispiel einer solchen Fehlertypologie bereit, das in Anhang C für den TBX-Basic-Dialekt dargestellt ist. Alle mit Begriffseinträgen zusammenhängenden Fehlertypologien müssen die Spezifikationen der jeweiligen Termbank widerspiegeln. So muss in TBX-Basic entweder eine Definition oder ein Kontext vorhanden sein, die Entwickler der Termbank können jedoch vorgeben, dass in einer bestimmten Instanzierung Definitionen zwingend und Kontexte optional sind (siehe ISO 30042). TBX-konforme Dialekte sollten vorsehen, dass Begriffsbeziehungen und Sachgebietsfelder nach ISO/TS 24634:2021 ausgedrückt werden.

Entwicklung und Inhalt von Termbanken kann entsprechend den Bedürfnissen einer Organisation sowie entsprechend kulturellen, linguistischen und regionalen Bedürfnissen variieren. Dementsprechend ist es äußerst wahrscheinlich, dass auch tatsächliche Bewertungs- (Scoring-) Schemata variieren.

6.3 Fehlertypen

6.3.1 Termbank-Spezifizierung und -Pflege

Die Sammlung von Begriffseinträgen in der Terminologiedatenbank muss die Spezifizierungen des jeweiligen Datenmodells erfüllen und muss die Kriterien für Termhood wie in Abschnitt 4 beschrieben erfüllen.

Die Termbank muss regelmäßig aktualisiert werden, um für bestehende Sachgebiete relevant zu bleiben, neue Sachgebiete zu unterstützen und bekannte Problembereiche anzugehen. Die Bewertung von mehrsprachigen Termbanken muss auf Termbank-Qualitätsbewerter verteilt werden, die der jeweiligen Sprache mächtig sind und die Bedürfnisse von Übersetzern, technischen Redakteuren und des Sachgebiets kennen.

Die Fehlertypologie in Anhang C kategorisiert Fehlertypen für:

- a) Fehler hinsichtlich der Datenintegrität des gesamten Begriffseintrags;
- b) Fehler auf der Begriffsebene des Eintrags;
- c) Fehler auf Sprachebene;
- d) Fehler auf Benennungsebene.

6.3.2 Fehlertypen, die der Aufnahme von Begriffseinträgen zugeordnet sind

Eine Nichteinhaltung der Termbank-Spezifikationen kann den gesamten Begriffseintrag betreffen, wie dies durch die Eintragsdaten-Integritätsprobleme, die in der beispielhaften Fehlertypologie in Anhang C dokumentiert sind, dargelegt ist. Eine systematische Reaktion auf solche Fehler kann Folgendes einschließen:

- Verschmelzen von Einträgen (wenn für denselben Begriff mehrere Einträge existieren);
- Teilen von Einträgen (wenn mehr als ein Begriff innerhalb eines einzelnen Eintrags behandelt wird);
- Entfernen von Einträgen, die dem Zweck der Termbank nicht dienlich sind.

6.3.3 Fehlertypen, die automatisch erzeugten Inhalten zugeordnet sind

Üblicherweise wird die nachstehende Gruppe von Datenfeldern vom Terminologieverwaltungssystem automatisch erzeugt:

- Eindeutige Kennung;
- Erstellungsdatum;
- Erstellt von;
- Zuletzt geändert am;
- Zuletzt geändert von.

Eindeutige Kennungen können auf der Begriffseintragungsebene (Eintragskennung) oder der Benennungsebene (Benennungskennung) angegeben werden. Die weiteren Elemente können an beliebigen Stellen innerhalb des Eintrags, an denen die Entwickler eine Dokumentation von Verantwortlichkeit und Datum wünschen, auftauchen. Obwohl diese Elemente aus Gründen der Vollständigkeit und Konformität mit den üblicherweise verwendeten Datenmodellen in die Fehlertypologie aufgenommen wurden, ist es unwahrscheinlich, dass es während

ihrer Erzeugung zu Fehlern kommt. Sollten Fehler auftreten, liegt die Lösung wahrscheinlich auf der Systemebene und ist im Hinblick auf die einfache Datenpflege nicht notwendigerweise lösbar.

6.3.4 Fehlertypen, die offenen Datenkategorien zugeordnet sind

Die Datenfelder /Benennung/, /Definition/ und /Kontext/ werden verwendet, um die wichtigsten terminologischen Informationen eines Begriffseintrags zu dokumentieren (siehe ISO 704:2022, Abschnitt 6 und Abschnitt 7 und ISO 26162-1:2019, 4.3). Diese sind Freitextfelder ohne formale Einschränkungen. Dementsprechend unterliegen sie einer größeren Palette an potentiellen Fehlern als die meisten anderen Felder. Zusätzlich können Fehler, die in diesen Feldern auftreten, in manchen Fällen Missverständnisse bei den Nutzern oder fehlerhafte Zuordnungen von Benennungen in Ausgangstexten oder von Äquivalenten in übersetzten Texten verursachen.

Fehler, die offenen Datenkategorien zugeordnet sind, sollten in allen zugeordneten Bewertungsmodellen entsprechend gewichtet werden. Das Feld /Anmerkung/ ist ebenfalls offen und kann Freitext enthalten, betrifft aber üblicherweise die Gesamtqualität eines Eintrags nicht im gleichen Umfang wie die Felder /Benennung/, /Definition/ und /Kontext/. Dementsprechend unterliegt /Anmerkung/ nicht notwendigerweise dem gleichen Maß an Gewichtung.

Eine Definition von Spezifikationen für die Informationen, die in jedem dieser Datenfelder zugelassen sind, unterstützt die Fehlervermeidung. Spezifikationen müssen angeben, in welcher Form Benennungen in der Termbank eingetragen werden (Großschreibung, kanonische Form, Verwendung von Klammern) und das Format von Definitionen und Konventionen für Informationen, die in Verwendungshinweisen hinterlegt werden, festlegen.

Um sicherzustellen, dass Informationen keine Rechtschreibfehler aufweisen, ist die Einschätzung der Genauigkeit von Textfeldern äußerst nützlich. Vorteilhaft ist die Möglichkeit, offene Datenfelder nach Folgendem durchsuchen zu können:

- doppelte Leerzeichen;
- eingebettete harte Zeilenumbrüche;
- versteckter Code;
- Großschreibung von Benennungen;
- Konsistenz von Wortklasse und Definition;
- Definitionen, Kontexteinträge oder Beispiele mit besonders kurzen oder langen Wortfolgen;
- persönliche Angaben wie Telefonnummern;
- Duplikate.

6.3.5 Fehlertypen, die geschlossenen Datenkategorien zugeordnet sind

Alle verbleibenden Datenkategorien werden üblicherweise als Werte in Auswahllisten konfiguriert. Obwohl es möglich ist, während der Erstellung oder Änderung eines Begriffseintrags falsche Werte auszuwählen, sind Werte in Auswahllisten üblicherweise weniger anfällig für Varianzen und bieten somit ein geringeres Risiko aus Sicht einer Risikoanalyse. Wichtige Datenkategorien in dieser Gruppe umfassen /Sachgebiet/ und /Wortklasse/. Letztere ist wesentlich, insbesondere in Fällen, in denen die Form von Verben, Substantiven und Adjektiven identisch sein kann. Eine korrekte Zuordnung des Sachgebiets kann für Filterzwecke, Sicherstellung der Anwendbarkeit von Definitionen und Zuordnung von Äquivalenten in mehreren Sprachen von großer Wichtigkeit sein.