



Historischer Thesaurus

Fortlaufende Statistik

Version 1.0.0

Stand 10. Juni 2026

thesaurus@herder-institut.de

Inhalt

Statistik Historischer Thesaurus 2026/1 (Juni 2026)	2
Anhang.....	4
Erläuterungen zu den Tabellen.....	5
Abkürzungsverzeichnis.....	7
Changelog.....	8

Statistik Historischer Thesaurus 2026/1 (Juni 2026)

Stand 10. Juni 2026

A. Umfang, Übersetzungstiefe

A1: Umfang. Begriffsdatensätze, Deskriptoren, Bezeichnungen

Verfügbare Begriffsdatensätze	3.852
Begriffsdatensätze System gesamt	4.574
Bezeichnungen gesamt	35.413
davon Deskriptoren	13.392
davon Synonyme / Quasisynonyme	22.021

A2: Mehrsprachigkeit. Hauptsprachen

	<i>Anzahl Begriffsdatensätze</i>	<i>Abdeckungsrate</i>
Deutsch	3.852	100 %
Englisch	3.798	98,60 %
Französisch	1.610	41,80 %
Polnisch	1.258	32,66 %

A3: Mehrsprachigkeit. Nebensprachen (mit mind. 30 Datensätzen)

	<i>Anzahl Begriffsdatensätze</i>	<i>Abdeckungsrate</i>
Russisch	422	10,96 %
Ukrainisch	368	9,55 %
Latein	71	1,84 %
Hebräisch	42	1,09 %
Tschechisch	36	0,93 %
Jiddisch	31	0,80 %

A4: Strukturelemente:

Facetten	6
Hierarchieelemente	53
Begriffe als Nichtdeskriptor	232
Node Label	111
Kategorie-/ConceptGroup-Datensätze	41

B. Mappings / Crosskonkordanzen

<i>Gruppe</i>	<i>Zielvokabular</i>	<i>Einträge gesamt</i>	<i>Abdeckungsrate*</i>
A	GND	3.580	92,94 %
A	Wikidata	3.565	92,55 %
A	AAT	3.445	89,43 %
A	WNK	3.433	89,12 %
A	RAMEAU	919	23,86 %
A	DBN	666	17,29 %
B	AGROVOC	17	0,44 %
B	FactGrid	21	0,55 %
B	Iconclass	2	0,05 %
B	LCGFT	5	0,13 %
B	LCNAF	71	1,84 %
B	LCSH	1.241	32,22 %
B	Mazal	43	1,12 %
B	MESH	8	0,21 %
B	NA-ČR	4	0,10 %
B	NALT	1	0,03 %
B	OBG	27	0,70 %
B	PACTOLS	1	0,03 %
B	STW	365	9,48 %
B	TDK	157	4,08 %
B	TGM	1	0,03 %
B	TheSoz	300	7,79 %
B	UNESCO	396	10,28 %

* Entspricht nicht der Übereinstimmungsrate, s. Abschn. Erläuterungen.

Anhang

Erläuterungen zu den Tabellen

Hinweise und Erläuterungen zu einzelnen Tabellenbereichen und -werten:

Verfügbare Begriffsdatensätze

Zur Indexierung freigegebene, geprüfte und mindestens erstredigierte Begriffsdatensätze in den Bearbeitungsständen ‚In Bearbeitung‘ oder ‚Abgeschlossen‘. Maßgebende Größenangabe zum eigentlichen Thesaurus und allen nachfolgenden anteiligen Angaben bzw. Abdeckungsraten (Sprachen, Mappings bzw. Crosskonkordanzen). I. d. R. vollständig verfügbar über Schnittstellen, Vollexporte/Dumps und Viewer.

Begriffsdatensätze System gesamt

Sämtliche Begriffsdatensätze, die im Thesaurus bzw. in der Datenbank verwaltet werden, einschl. unbearbeiteten Kandidaten, abgelehnten Begriffen, internen Wartelisten usf.; sie sind enthalten im RDF-XML-Dump.

Bezeichnungen gesamt

Sämtliche Einzelbezeichnungen, die innerhalb der Begriffsdatensätze des Thesaurus verwaltet werden, einschließlich bevorzugten Bezeichnungen/Deskriptoren, Synonymen, Quasi-Synonymen usf. Erfasst sind auch Bezeichnungen zu Begriffsdatensätzen, die (noch) nicht öffentlich verfügbar sind. Nicht erfasst sind hier die Bezeichnungen von Strukturelementen, einschließlich den *Begriffen als Nichtdeskriptoren*.

Bezeichnungen gesamt: Davon Deskriptoren

Anzahl der Deskriptoren, d. h. der als Schlagwort verwendbaren bevorzugten Bezeichnungen je Sprache. Ein mehrsprachig redigierter Begriffsdatensatz besitzt dementsprechend mehrere Bezeichnungen, die als bevorzugte Bezeichnung je Sprache ausgezeichnet sind. Im Gegensatz zu anderen Vokabularen (wie etwa dem *Art & Architecture Thesaurus*) wird nicht zwischen Deskriptoren und bevorzugten Bezeichnungen unterschieden, beides ist implizit deckungsgleich.

Bezeichnungen gesamt: Davon Synonyme / Quasisynonyme

Anzahl aller Synonyme (Äquivalente) oder Quasisynonyme (Quasi-Äquivalente) innerhalb der Begriffsdatensätze. Eingeschlossen sind *verborgene* Synonyme (bspw. verbreitete Falschschreibungen) sowie die nur punktuell verwendeten und als Bezeichnung hinterlegten Angaben zur Kombination von Begriffen.

Mehrsprachigkeit

Mindestens erstredigierte, öffentlich verfügbare Begriffsdatensätze mit Bezeichnungen in den nachfolgend genannten Sprachen. Größere zahlenmäßige Diskrepanzen bei den Hauptsprachen ergeben sich sowohl aufgrund der inkrementell erfolgenden Redaktion sowie der teils erst nachträglichen Aufnahme bzw. Festlegung von Sprachen als Hauptsprachen des Thesaurus (Französisch, Polnisch).

Strukturelemente

Statistik aller im Thesaurus genutzten, redaktionell in Bearbeitung befindlichen oder abgeschlossenen Strukturelemente – die nicht zur Indexierung verwendet werden dürfen. Eingeschlossen sind Datensätze, die außerhalb des eigentlichen Thesaurus-Trees für die Strukturierung von Anmelde- und Redaktionsbereichen genutzt werden.

Abdeckungsrate (Mehrsprachigkeit)

Anteil der Thesaurusdatensätze, die mehrsprachig verfügbar sind.

Mappings / Crosskonkordanzen: Einträge gesamt

Anzahl der Begriffsdatensätze des Thesaurus, in denen eine Angabe zum Mapping auf das jeweilige Ziel- bzw. Referenzvokabular hinterlegt ist. Größere zahlenmäßige Diskrepanzen bei den obligatorischen Mappings ergeben sich aufgrund der zeitlich verzögerten Festlegung dieser Referenzvokabulare als obligatorische Mappingziele im Kontext der Mehrsprachigkeit.

Wichtig: Die Angabe erfolgt einschließlich aller Mappingeinträge ohne festgestellte Übereinstimmung bzw. ohne erfolgreiche Ermittlung einer sachlichen Entsprechung (!). Der Wert entspricht also NICHT der Übereinstimmung oder inhaltlichen Überlappung des Historischen Thesaurus mit anderen Vokabularen, sondern dokumentiert den Umfang der (erfolgreichen oder erfolglosen) redaktionellen Prüfung insgesamt.

Übersichten der inhaltlichen Übereinstimmungsraten sind künftig ebenfalls vorgesehen.

Mappings / Crosskonkordanzen: Gruppe

- A. Obligatorische Mappings
- B. Fakultative Mappings, die überwiegend nur punktuell und nach Einschlägigkeit hinterlegt werden

Abkürzungsverzeichnis

AAT	Art & Architecture Thesaurus (AAT)
DBN	Deskryptory Biblioteki Narodowej (DBN)
GND	Gemeinsame Normdatei (GND)
LCGFT	LC Genre/Form Terms (LCGFT)
LCNAF	LC Name Authority File (LCNAF)
LCSH	Library of Congress Subject Headings (LCSH)
MESH	Medical Subject Headings (MESH)
NA-ČR	Národní autority ČR (NA-ČR)
NALT	NAL Agricultural Thesaurus (NALT)
OBG	Objektbezeichnungsdatei
RAMEAU	Répertoire d'autorité matière encyclopédique et alphabétique unifié RAMEAU
STW	Standard-Thesaurus Wirtschaft (STW)
TDK	Tezaurus dziedzictwa kulturowego
TGM	Thesaurus for Graphic Materials (TGM)
TheSoz	Thesaurus Sozialwissenschaften (TheSoz)
UNESCO	UNESCO Thesaurus
WNK	Wortnetz Kultur (WNK)

