

Steffi Nitsche, Anke Taube

DNBLab:

Praktische Einführung in die Metadatenanalyse



Inhalt

- 1. Daten, Formate und Bezugswege
- 2. Das DNBLab
- 3. MARC-Daten herunterladen
- 4. MARC-Daten analysieren



Gesetzlicher Auftrag

Sammlung von Schrift, Bild und Ton seit 1913 in und über Deutschland oder in deutscher Sprache veröffentlichten Publikationen

- fast 50 Mio. Werke
- 33 Mio. bibliografische Datensätze (freie Metadaten)
- 16 Mio. digital (Netzpublikationen)
- 3 Mio. mit freiem Zugang



Weitere Bestände

- Deutsches Exilarchiv 1933 1945
- Deutsches Buch- und Schriftmuseum
- Deutsches Musikarchiv

dnb.de/portraet

NETZPUBLIKATIONEN BESTAND 2023



3.561.518

Book on Demand, Hochschulschriften, Musikalien und sonstige



PERIODIKA 11.154.333

> E-Paper, Hefte, Artikel



86.026

insgesamt



WEBSITES

64.381

Zeitschnitte

→ davon 3 Mio. frei zugänglich

2.575.109
Zugang 2023

14.866.258
Bestand insgesamt





Bereits digitalisierter Bestand

- Objekte
- 307.711 Bücher, Zeitschriften etc.
- → davon etwa 130.000 frei zugänglich

ToCs

- 2.697.313 Inhaltsverzeichnisse
 - → alle frei zugänglich

- Musik
- 56.232 Tonträger
 - → davon etwa 1.800 frei zugänglich



dnb.de/metadatenbezugswege

Titeldaten									
Titeldaten									
Format	Datenshop	SRU	ОАІ-РМН	WWW-/SFTP- Server					
MARC 21	х			х	Normdater	1			
MARC21-xml	x	x	x	x					
RDF (RDF/XML)	х	х	х	х	Normdaten				
RDF (JSON-LD)	х			х	Format	Datenshop	SRU	OAI-PMH	WWW-/SFTP- Server
RDF (Turtle)	x			х	MARC 21	x			
CSV	x								х
DNB Casual	x	x	×		MARC21-xml	x	х	х	X
(oai_dc)	^	^	^		RDF (RDF/XML)	×	×	x	x
MODS-xml	x	х	x		RDF (JSON-LD)	x			х
PDF	х			x	RDF (Turtle)	х			х



Welches Exportformat ist für Sie das Richtige?







Deutsche Nationalbibliothek Digitale Dienste Iab@dnb.de

Welches Exportformat ist für Sie das richtige?

Ein guter Einstieg zu Beantwortung dieser Frage finden Sie auf der Seite

https://www.dnb.de/metadatendienste unter dem Punkt: Welche Exportformate sind verfügbar?

Letztlich können nur Sie das entscheiden. Es gibt aber ein paar Kriterien, die Ihnen bei dieser Entscheidung helfen können.

Vollständigkeit

Keines der Exportformate bildet alle in unserer internen Datenbank befindlichen Daten vollständig ab. AMAC 22 bletet die vollständigsten beten, das Format DND, Casual (auch oal, dz. genannt) liefert nur sehr ausgewählte Daten. Eine möglichst große vollständigkeit bletet natürlich die meisten Möglichkeiten, kann aber für Ihren Anwendungsfall durch die schiere Masse an Informationen auch einen Ballast darstellen. Und mit dem Grad der Vollständigkeit steien natürlich auch die:

Komplexität

Komplex kann dabei einerseits die Vielzahl an gelieferten Daten sein, aber auch die Strukturen, in denen diese Informationen oder Daten "verpackt" sind. Nicht umsonst hat die Dokumentation von MARC 21, allein für die Titeldaten der DNB fast 400 Seiten. Lesbarkeit (für Menschen)

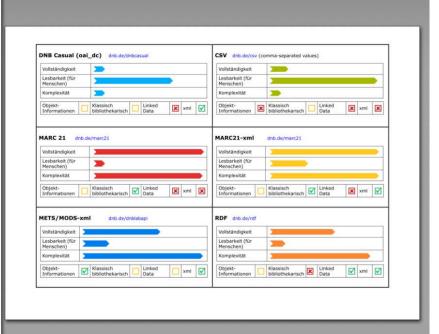
Esportformate sind in enster Linie maschinenlesbar, nur so können sie ihren Zweck erfüllen. Dennoch sit es manchmal notwendig oder hilfriecht, die Metadaten zunächst mit dem menschlichen Auge zu lesen. Das ist durch die unterschiedlichen Strukturen der Formate zum Teil relativ einfach möglich, zum Teil aber nahezu ummöglich. Bei MABC 21 etwa brauchen Sie ein eigenes Hilfsmitzte zur Darstellung (anders bei MARG-zml). Hilfreich für die Leisbarkeit durch Menschen ist zum Beispiel, wenn die Bezeichnung der einzelnen Felder der Kategorien im Klatexte enthalten ist und nicht durch eine Zifferm-Zeichen-Folge order wird. In den RDF-Formaten etwa werden die meisten Inhalte nicht mehr durch ein Lütern angegeben, sonder durch einen Hyperlink. Die beden Formate, die am wenigsten vollständig sind, CSV und oal, dc, sind andererseits am wenigsten komplex und am leichtesten in diesem Sinne lesbar.

Objektinformation

Objektinformation
Ausschließlich eines der von uns angebotenen Formate (METS/MODS) umfasst auch
Informationen zu den beschriebenen digitalen Objekten selbst. Aber bei fast allen unseren
Formaten finden Sie in unseren Metadaten digitaler Objekte einen Link, der zu diesem

Zugang

Nicht jedes Format ist über jeden Bezugsweg verfügbar, Dazu finden Sie auf https://www.dnb.de/metadatenbezugswege einen guten Überblick.







Handout unter dhb.de/dhblabtutorials

Deutsche Nationalbibliothek Digitale Dienste lab@dnb.de



Übersicht: Inhalte in MARC21-Feldern

Die folgende Tabelle bietet einen Überblick über die am haufigsten benötigten und für die Arbeit mit den bibliografischen Daten unserer Titeldaten relevantesten MARC21-Felder¹. Sie ist nicht vollständig! Eine Gesamtübersicht zu MARC21, allen Feldern und deren Bedeutungen für unsere Titeldaten, Bestandsdaten sowie für die Normdaten der GND finden Sie unter https://www.dnb.de/marc21.

Ich suche			Tag	Subfield/ Position
Identifikatoren	IDN	С	001	-
	ISBN	D	020	a
	ISSN	D	022	3
	URN (und weitere Identifikatoren)	D	024	a
	Quelle der Nummer oder des Codes	D	024	2
Verantwortliche und	Personenname (Haupteintragung)	D	100	a
Mitwirkende (bspw. Autoren/Herausgeber) ³	zugehörige Funktionsbezeichnung	D	100	0
HULUTER/HEI BUSGEDER J	Personenname (Nebeneintragung – bspw. Mitverantwortliche)	D	700	а
	zugehörige Funktionsbezeichnung	D	700	e
	Körperschaftsname (Haupteintragung)	D	110	a
	Kongressname (Haupteintragung)	D	111	a
	Körperschaftsname (Nebeneintragung)	D	710	а
	Kongressname (Nebeneintragung)	D	711	3
Titelangabe	Titel	D	245	a
	Zusatz zum Titel	D	245	ь
	Verfasserangabe etc.	D	245	c
Titelvarianten	Hauptsachtitel / Kurztitel	D	246	a
	Zusatz zum Titel	D	246	b
Gesamttitel (bspw. bei	Gesamttitelangabe	D	490	a
mehrbändigen Werken)	Bandnummer/Bandbezeichnung	D	490	V
	Gesamttitel – Einheitstitel (Nebeneintragung)	D	830	a
Einheitstitel	Einheitstitel (Haupteintragung)	D	240	a
	Einheitstitel (Nebeneintragung)	D	730	8
Ausgabebezeichnung	Ausgabe / Auflage	D	250	a
Verlag	(Erzeuger-, Verlags-, Vertriebs-, Herstellername)	D	264	ь

³ Einige MARC-Felder sind wiederholbar und können mehrfach in einem Datensatz enthalten sein. Alle Informationen hierzu finden sich in der Feldbeschreibung der Titeldaten.

Publikationsort	(Entstehungs-, Verlags-, Vertriebs-, Herstellungsort)	D	264	а
Datumsangaben	Erscheinungsdatum (Entstehungs-, Verlags-, Vertriebs-, Herstellungs- oder Copyrightdatum)	D	264	c
	Erscheinungs-/Entstehungsdatum	C	008	07-10
	Abschließendes Erscheinungs- /Entstehungsdatum ⁴	C	008	11-14
	Datum der Ersterfassung	С	008	00-05
Sprache	des vorliegenden Werkes	D	041	а
	des Originals (bei Übersetzungen)	D	041	h
Physische Merkmale	Umfang (bspw. Seitenzahl)	D	300	а
	Andere physische Details	D	300	b
	Dimension (bspw. Höhe in cm)	D	300	0
Klassifikation/Sachgruppe	DDC-Notation	D	082	a
	Vergabestelle	D	082	q
	Ausgabenummer	D	082	2
	Zusätzliche DDC-Notation (alle Felder wie DDC-Hauptsachgruppe – nur, wenn 082 bereits vorhanden)	D	083	Alle Subfields wie 082
	Andere Notationen	D	084	а
	Vergabestelle der Notation	D	084	q
	Quelle der Nummer	D	084	2
Schlagwörter ⁵	Schlagwort - Personenname	D	600	а
	Schlagwort – Körperschaftsname	D	610	a
	Schlagwort - Kongressname	D	611	a
	Schlagwort – Einheitstitel	D	630	a
	Schlagwort - Zeitschlagwort	D	648	a
	Schlagwort - Sachschlagwort	D	650	a
	Schlagwort - Geografischer Name	D	651	8
	Schlagwort - Genre/Formschlagwort	D	655	a
Hochschulschriftenvermerk	alle Informationen als Freitext	D	502	3
oder	Typ (Dissertation, Masterarbeit etc.)	D	502	b
	Universität	D	502	C
	Jahr	D	502	d
Links	URN ⁶	D	856	U
	DOI	D	856	u
	Inhaltsverzeichnis ²	D	856	ш
	Inhaltstext ⁶	D	856	U
	Weitere URIs	D	856	u
	Spezifische Materialangaben	D	856	3
	Allgemeine Anmerkungen	D	856	7

⁴ bei Zeitschriften und Schriftenreihen, alternativ Datum des Originals in Sortierform. Vgl. hierzu die





2 | 2

[&]quot; "C" steht für Controlfield. "D" für Datafield.

³ Relevant sind hier auch jeweils die Unterfelder \$0 (IDN des zugehörigen Normdatensatzes) sowie \$2 (Quelle).

Teit Zeitschriften und Schriftenenben, allernativ Detum des Originals in Sorderform, Vgl. hietzu die Felbleschribung der MAR-Triebsden under Hussyl/men, dink definant 21.
**Palevant sind auch hier zweist die Unterfelder 30 (Dirt des zugehöngen Normdatensattes) sowie 52 (Quelle).
**In Verbindung mit der Angabe "Inhaltsverardenhis" im Unterfeld 53.
**In Verbindung mit der Angabe "Inhaltsverardenhis" im Unterfeld 53.
**In Verbindung mit der Angabe "Inhaltsverardenhis" im Unterfeld 53.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 53.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 53.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 53.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 53.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 53.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 53.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 53.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 54.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 54.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 54.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 54.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 54.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 54.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 54.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 54.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 54.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 54.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 54.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 54.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 54.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 54.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 54.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 54.
**In Verbindung mit der Angabe "Inhaltstert" im Vererde 54.
**In Vererden mit der Angabe "Inhaltstert" im Vererde 54.
**In Vererden mit der Angabe "Inhaltstert" im Vererde 54.
**



? ? ? ?

Welcher Zugang ist für Sie der Richtige?

? ? ? ?

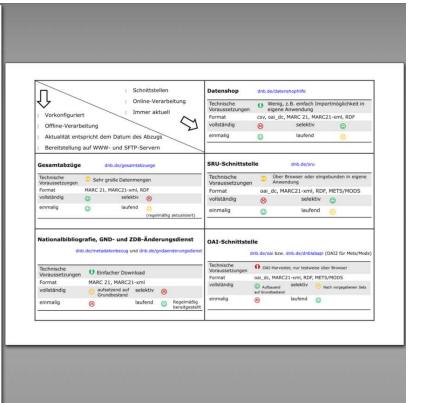




Handout unter dnb.de/dnblabtutorials

LEIPZIG FRANKFURT AM MAIN







Was ist das DNBLab?

- Bibliothekslab der Deutschen Nationalbibliothek
- Start im Sommer 2020
- Zentraler Anlaufpunkt für alle mit Interesse an Text- and Data Mining
- Einführungsveranstaltungen, Workshops und Lehrkooperationen



Was bietet das DNBLab?

- freie digitale Objektsammlungen
- Beschreibungen unserer Datenformate und Schnittstellen
- Online-Coding Tutorials, Erklärvideos, Hand-outs
- Beispielsammlung für das Arbeiten mit Daten
- zentrale Ansprechpartner*innen <u>lab@dnb.de</u>



dnb.de/dnblabsets

DnbLab: Freie digitale Objektsammlungen

Beschreibung	Anzeige im Katalog	Datensets mit Hyperlinks zu Objekten und/oder Bilddateien
Mehr als 2,8 Millionen Inhaltsverzeichnisse von Büchern, die seit 1913 vor allem in Deutschland er- schienen sind; zeitlicher Schwerpunkt sind die Jahre ab 1983 bis heute mit über 6,5 Millionen durchsuchbaren Einzelseiten		Download L ^Z (MARC21-xml)
Mehr als 750.000 Ob- jekte ohne Zugriffsbe- schränkung aus Deutschland	Sammlung im Katalog 🗗	Download L ^Z (MARC21-xml)
Mehr als 325.000 Online- Hochschulschriften ohne Zugriffsbeschränkung aus Deutschland	Sammlung im Katalog L [™]	Download L ^Z (MARC21-xml)
	Mehr als 2,8 Millionen Inhaltsverzeichnisse von Büchern, die seit 1913 vor allem in Deutschland er- schienen sind; zeitlicher Schwerpunkt sind die Jahre ab 1983 bis heute mit über 6,5 Millionen durchsuchbaren Einzelseiten Mehr als 750.000 Ob- jekte ohne Zugriffsbe- schränkung aus Deutschland Mehr als 325.000 Online- Hochschulschriften ohne Zugriffsbeschränkung	Mehr als 2,8 Millionen Inhaltsverzeichnisse von Büchern, die seit 1913 vor allem in Deutschland er- schienen sind; zeitlicher Schwerpunkt sind die Jahre ab 1983 bis heute mit über 6,5 Millionen durchsuchbaren Einzelseiten Mehr als 750.000 Ob- jekte ohne Zugriffsbe- schränkung aus Deutschland Mehr als 325.000 Online- Hochschulschriften ohne Zugriffsbeschränkung Katalog L Sammlung im Katalog L Katalog L



srudump.streamlit.app/

Create DNB data dump

Please enter your search query here	
tit=wespe	
Please select catalogue:	Please select metadata format for output:
O DNB (Titeldaten)	MARC21-xml
○ GND (Normdaten)	○ oai_dc
○ DMA	○ RDFxml
○ ZDB (Zeitschriftendatenbank)	○ mods-xml
Adressdaten (ISIL - und Siegelverzeichnis)	
Start SRU query	
You are currently running a query for: tit=wespe	
Records found: 96	
✓ Query complete	
Download Data	



MARC21-xml

- | MARC = **Ma**chine-**R**eadable **C**ataloguing
- | Metadatenformat für den Austausch von Bibliotheksdaten
- | 1960er-Jahre: Entwicklung durch die Library of Congress
- | 1999: Version MARC 21 (Harmonisierung Vorgängervarianten)
- | 2004: Beschluss zur Übernahme von MARC 21 als einheitliches Austauschformat im deutschsprachigen Bibliothekswesen



MARC21-xml:

- | bibliografisches Metadatenformat
- Daten werden in einem strukturierten und standardisierten Format gespeichert
- | Jeder Datensatz besteht aus mehreren **Feldern**, die spezifische Informationen zu einer Ressource (z. B. Buch, DVD DVD) enthalten
- | Einige Informationen können sich in mehreren Feldern befinden (bspw. Jahreszahlen)



MARC21-xml:

- Ein MARC21-Datensatz besteht immer aus drei Teilen:
 - Leitdaten (Leader): Enthält grundlegende Informationen über den Datensatz (z. B. Länge, Status des Datensatzes).
 - Steuerfelder (Controlfields): Felder, die maschinenlesbare Informationen enthalten, z. B. einen Identifikator oder den Publikationsstatus.
 - Datenfelder (Datafields): Enthalten die eigentlichen bibliografischen Informationen wie Titel, Autor, Verlag, Erscheinungsjahr usw.
 - Jedes **Datafield** besteht aus einer dreistelligen Zahl und kann mehrere **Unterfelder (Subfields)** enthalten, die durch Buchstaben oder Zahlen gekennzeichnet sind.



MARC21-xml:

- MARC21 liegt das Regelwerk Resource Description and Access
 (RDA) zugrunde, d.h. dieses bestimmt
 - | Welche Felder und Informationen in jedem Datensatz vorhanden sein müssen
 - Welche Felder und Informationen optional sind
 - Wie die Informationen in den verschiedenen Feldern "abgelegt" werden (bspw. Regeln zur Erfassung von Autoren / Herausgebern)
 - Welche Felder und Unterfelder wiederholbar sind und welche nicht
 - | Welche Informationen codiert gespeichert werden
- | Wichtig: Erschließung richtet sich immer nach dem jeweils vorliegenden Medienwerk

https://services.dnb.de/sru/dnb?version=1.1&operation=searc hRetrieve&query=1033661937&recordSchema=MARC21xml&maximumRecords=100



Veröffentlichung	en der Landesarchivverwaltung Rheinland-Pfa
	Band 115
Territor	rialisierung mit dem Schwert?
Die F Balduin von	Fehdeführung des Trierer Erzbischofs Luxemburg (1307/08-1354) im Erzstift Trier
	von Julia Eulenstein

	Piptipii
Tag	Inhalt
024	\$a 9783931014865
082	\$8 2/u \$a 943.4313026092 \$q DE-101 \$2 22/ger
100	\$a Eulenstein, Julia \$d 1974- \$e Verfasser \$4 aut \$2 gnd \$0 (DE-588)1033823317 \$0 https://d-nb.info/gnd/1033823317
245	\$a Territorialisierung mit dem Schwert? \$b die Fehdeführung des Trierer Erzbischofs Balduin von Luxemburg (1307/08 - 1354) im Erzstift Trier \$c von Julia Eulenstein. [Rheinland-Pfalz, Landesarchivverwaltung]
264	\$a Koblenz \$b Landeshauptarchiv Koblenz \$c 2012
300	\$a X, 612 S. \$b Kt. \$c 25 cm
490	\$a Veröffentlichungen der Landesarchivverwaltung Rheinland-Pfalz \$v Bd. 115
502	\$a Zugl.: Gießen, Univ., Diss., 2009/2010
651	
653	
653	
653	
655	
810	\$a Landesarchivverwaltung Rheinland-Pfalz \$4 aut \$t Veröffentlichungen der Landesarchivverwaltung Rheinland-Pfalz \$v Bd. 115 \$w (DE-101)011436247 \$w (DE-600)238856-X \$9 3115 \$7 as

https://services.dnb.de/sru/dnb?version=1.1&operation=searchRetrieve&query=982707010&recordSchema=MARC21-xml&maximumRecords=100



Ligroff Page Longton Co.
Dorothea Forstner • Renate Becker
And the second s
LEXIKON
CHRISTLICHER
SYMBOLE
Aktualisierte Neuausgabe
marixverlag

Tag	Inhalt
024	\$a
082	\$8 \$a \$q \$2
100	\$a Forstner, Dorothea \$d 1884-1969 \$e Verfasser \$4 aut \$2 gnd \$8 1\p \$0 (DE-588)1089923899 \$0 https://d-nb.info/gnd/1089923899
245	\$a Lexikon christlicher Symbole \$c Dorothea Forstner; Renate Becker
250	\$a Aktualisierte Neuausg.
264	\$a Wiesbaden \$b Marixverl. \$c 2007
300	\$a 436, [48] S. \$b III. \$c 22 cm
500	\$a Lizenz der VerlAnst. Tyrolia, Innsbruck Literaturverz. S. 408 - 422
650	
650	
653	
689	
700	\$a Becker, Renate \$e Verfasser \$4 aut \$2 gnd \$0 (DE-588)1226987982 \$0 https://d-nb.info/gnd/1226987982
856	\$m V: DE-604 \$q application/pdf \$u https://d-nb.info/982707010/04 \$3 Inhaltsverzeichnis

https://services.dnb.de/sru/dnb?version=1.1&operation=se archRetrieve&query=Adelige+Streitkultur&recordSchema= MARC21-xml&maximumRecords=100





Tag	Inhalt
024	\$a
082	\$8
245	\$a Adelige Streitkultur \$c herausgegeben von Gerd Dethlefs in Verbindung mit Ralf-Peter Fuchs, Ute K. Boonen und Luisa Röhrich
264	\$a [Münster] \$b Aschendorff Verlag \$c [2023]
300	\$a 236 Seiten \$b Illustrationen \$c 25 cm
490	\$a Rhein-Maas \$v Band 12 (2023)
648	
650	
650	
700	\$a Dethlefs, Gerd \$d 1958- \$e Herausgeber \$4 edt \$2 gnd \$0 (DE-588)110569660 \$0 https://d-nb.info/gnd/110569660
830	\$a Rhein-Maas \$v Band 12 (2023) \$w (DE-101)1002736889 \$w (DE-600)2555043-3 \$9 212 \$7 as



Ausführliche Dokumentation: dnb.de/marc21

MARC 21 ■ INHALT • Über MARC 21 • Anwendung von MARC 21 in der Deutschen Nationalbibliothek und der Zeitschriftendatenbank im Format MARC 21 MARC 21 • Arbeitshilfen • Anwenderebene • Version 2.5 Stend: 12. Februar 2010

Handout "Übersicht: Inhalte in MARC21-Feldern": dnb.de/dnblabtutorials

Ich suche		Art des Feldes ²	Tag	Subfield/ Position
Identifikatoren	IDN	С	001	-
	ISBN	D	020	a
	ISSN	D	022	a
	URN (und weitere Identifikatoren)	D	024	a
	Quelle der Nummer oder des Codes	D	024	2
Verantwortliche und	Personenname (Haupteintragung)	D	100	a
Mitwirkende (bspw. Autoren/Herausgeber) ³	zugehörige Funktionsbezeichnung	D	100	е
Autoren/merausgeber j	Personenname (Nebeneintragung – bspw. Mitverantwortliche)	D	700	a
	zugehörige Funktionsbezeichnung	D	700	e
	Körperschaftsname (Haupteintragung)	D	110	a
	Kongressname (Haupteintragung)	D	111	a
	Körperschaftsname (Nebeneintragung)	D	710	a
	Kongressname (Nebeneintragung)	D	711	a
Titelangabe	Titel	D	245	a
	Zusatz zum Titel	D	245	b
	Verfasserangabe etc.	D	245	С



Vielen Dank für Ihre Aufmerksamkeit!

Kontakt:

s.nitsche@dnb.de

a.taube@dnb.de

oder

lab@dnb.de