# Was sind gute Daten?

Stefan Lohrum, KOBV Zentrale im Zuse-Institut Berlin

VÖBB – Workshop zur Datenqualität 10.02.2020, Zentral- und Landesbibliothek Berlin

## Daten: Messen & Erfassen

Beispiel: Fiebermessung

Mit welchem Thermometer?
Wo (Stirn, Ohr, Mund, Achsel, Darm)?

Stirn- und Ohrthermometer messen nicht so exakt wie Fieberthermometer am Darmausgang

Die Technik entwickelt sich immer weiter, doch zeigte eine dänische Studie, dass für Kinder die rektale Messung, d.h. am Darmausgang, mit einem digitalen Fieberthermometer immer noch am exaktesten ist. [...] Wollen Eltern das Fieber im Mund messen, kann dies für kleinere Kinder eine Herausforderung darstellen. [...]

Studie: Morgensen et al: Forehead or ear temperature measure- ment cannot replace rectal measurements, except for screening purposes (in BMC Pediatrics 2018, PMCID: PMC5787302)

Quecksilber-Thermometer





Infrarot-Thermometer



Temperatur Scan am Flughafen



## Messung

- Aufwand
- Kontext
- Fehler

# Routenplanung / Stauwarnung





Verkehrsüberwachung



Verkehrszähler







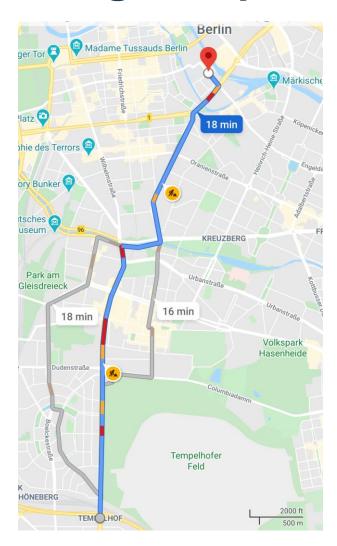
Magnetstreifen



Staumelder

- Erfassung von Verkehrsdaten aus unterschiedlichen Quellen
- Konsolidierung
- Verkehrsnachrichten Verkehrsinformation

# **Google Maps**



#### **Smartphone**

- erfasst Standort Koordinaten (z.B. GPS)
- sendet Information an GoogleKoordinaten + Zeitstempel + ID

#### Google

- Zuordnung anhand der ID
- berechnet Geschwindigkeit  $\frac{\Delta s}{\Delta t}$



- Durchschnittsgeschwindigkeit aller Smartphones an einem Ort
  - Keine eigene Erfassung:Mitnutzung der Standortübermittlung
  - Einheitliches Datenformat Identifier
  - Allein in Deutschland mehr als1.000.000.000 Übermittlungen / Tag

(40 Mio Android Smartphones, 10% in Bewegung, alle 5 Min eine Nachricht)



# Google Maps

# Berlin Madame Tussauds Berlin

16 min

Tempelhofer

Feld

#### **Smartphone**

- erfasst Standort Koordinaten (z.B. GPS)
- sendet Information an Google Koordinaten + Zeitstempel + ID

#### Google





- Keine eigene Erfassung: Mitnutzung der Standortübermittlung
- **Finheitliches Datenformat** Identifier
- Allein in Deutschland mehr als 1.000.000.000 Übermittlungen / Tag

(40 Mio Android Smartphones, 10% in Bewegung, alle 5 Min eine Nachricht)



TEM LHOP

HÖNEBERG

latz 😉

hie des Terrors

ory Bunker 🕮

tsches 🔝

Park am Gleisdreieck

18 min

Dudenstraße

Hasenheide

Märkische

# Daten - Eigenschaften

- Daten werden zu ein bestimmten Zweck erfasst
- Daten haben einen Wert (Aufwand bei der Erfassung)
- Daten haben Struktur
- Daten haben Kontext / Provenienz
- Daten enthalten Fehler
- Daten enthalten Identifikatoren

Große
Datenmengen

Viele Datenquellen

Big Data

Schnelle Verarbeitung

Kosteneffektiv Prozeßautomatisierung

**Big data** is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation. (aus Gartner IT Glossery)

Große
Datenmengen

Viele Datenquellen

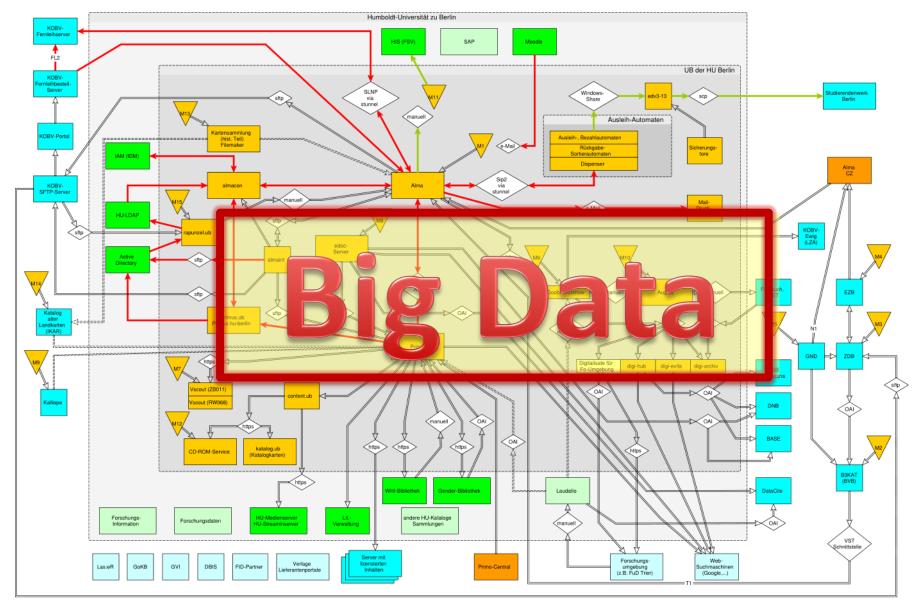
Prozeßauto-

Big Data

Schnelle Verarbei-

in Bibliothe Ren?

**Big data** is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation. (aus Gartner IT Glossery)



Lizenz: CC-BY-SA

KOBV .:

## **Automatisierte Prozesse in Bibliotheken**

#### Verbundkataloge

- "Ansigeln" des eigenen Bestandes
- automatisierte Übernahme von Sacherschließung

## Lieferung der Metainformation durch Verlage / Aggregatoren

- Datenbanken / Archive (z.B. Genios, Munzinger)
- Ebooks
- Patron Driven Aquisition

## Nutzung der Informationen aus den passiven Bestellungen

- in der Fernleihe als Vorschläge für den Bestandsaufbau

### Nutzungsstatistiken

- Automatisierte Zählung der Bibliotheksnutzer/inn/en
- Auswertungen der Ausleihen
- Auswertungen von Recherchezugriffen

## Automatisierte Prozesse in Bibliotheken

### Verbundkataloge

- "Ansigeln" des eigenen Bestandes
- automatisierte Übernahme von Sacherschließung



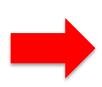
- in der Fernleihe als Vorschläge für den Bestandsaufbau

#### Nutzungsstatistiken

- Automatisierte Zählung der Bibliotheksnutzer/inn/en
- Auswertungen der Ausleihen
- Auswertungen von Recherchezugriffen

# Erschließung – Wozu?

- 1. Beschreibung der Medien im Bestand der Bibliothek
- 2. Auffindbar machen der Medien
- DNB: Erschließung bedeutet, ein Medienwerk ein Buch, eine Zeitschrift, eine Website, eine Karte oder Musikalie durch Metadaten so zu beschreiben, dass es in einem Online-Katalog oder einer Datenbank gefunden wird.

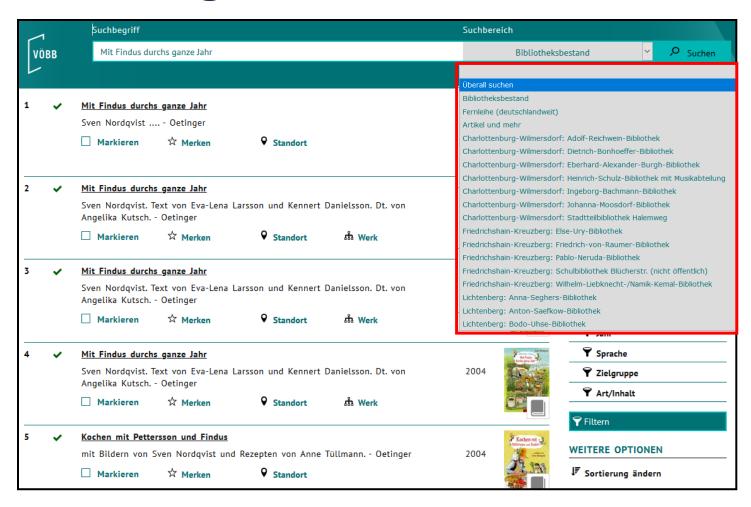


Wie werden Medien im VÖBB gesucht und gefunden ?

Welche Daten sind dazu erforderlich?

https://www.dnb.de/DE/Professionell/Erschliessen/erschliessen\_node.html

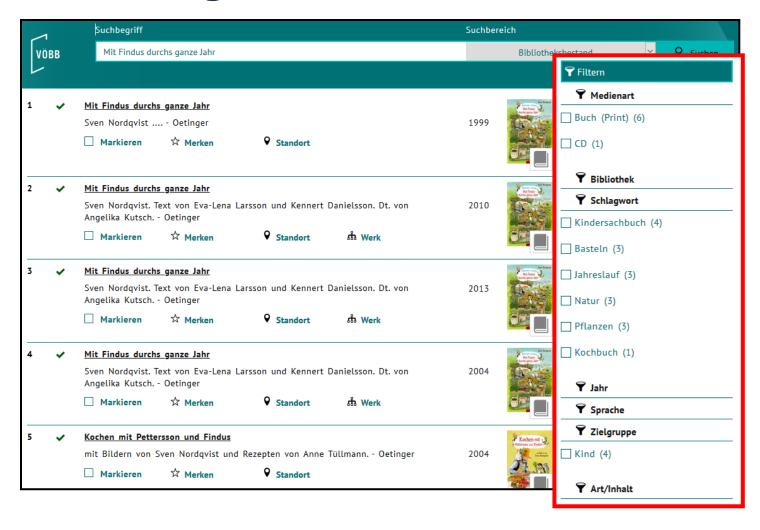
# Suchmöglichkeiten im VÖBB Portal



#### Auswahl

Quellen

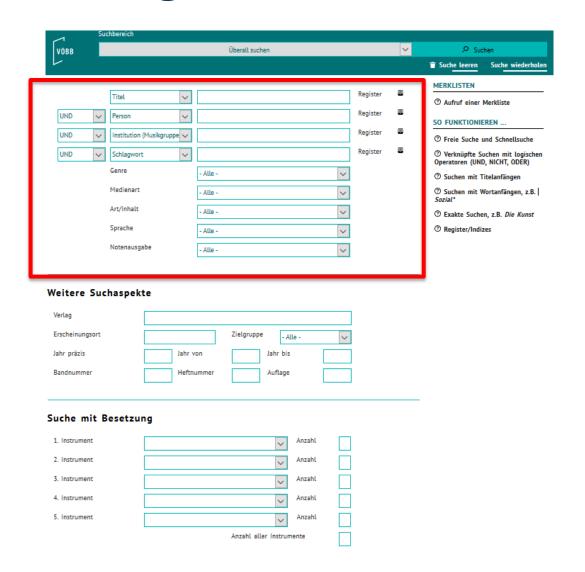
# Suchmöglichkeiten im VÖBB Portal



#### Auswahl

- Quellen
- **►** Filter

# Suchmöglichkeiten im VÖBB Portal



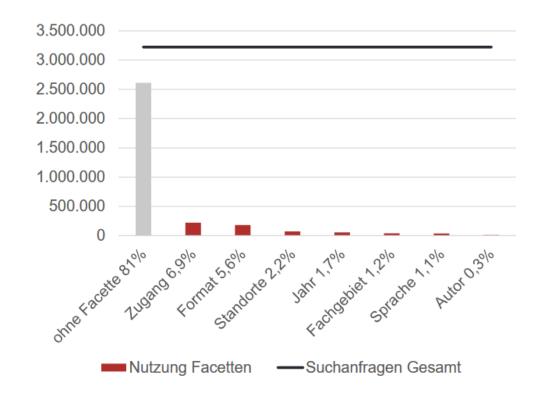
#### Auswahl

- Quellen
- Filter
- Erweiterte Suche

## Nutzung im Portal der UB Leipzig

**STATISTIK** 

NUTZUNG FACETTEN



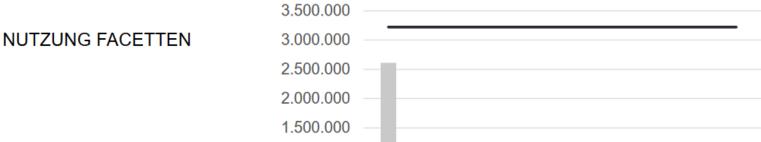
Quelle: Björn Muschal: Beobachtungen und Gedanken zur Anforderung an bibliographische Metadaten, Vortrag auf der Fortbildungsveranstaltung "Metadaten zwischen Autopsie und Automatisierung" der dbv-Kommission Erwerbung und Bestandsentwicklung an der ULB Düsseldorf am 13. März 2017

URL: https://www.bibliotheksverband.de/fileadmin/user\_upload/Kommissionen/Kom\_ErwBest/Tagungen/Erwkomm\_Fortbild\_Ddorf2018\_Muschall.pdf

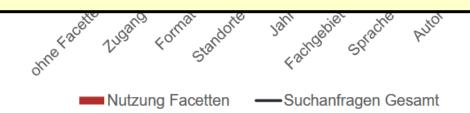
KOBV .:::

## Nutzung im Portal der UB Leipzig





# Analyse: Wie suchen die Nutzer im VÖBB ?



Quelle: Björn Muschal: Beobachtungen und Gedanken zur Anforderung an bibliographische Metadaten, Vortrag auf der Fortbildungsveranstaltung "Metadaten zwischen Autopsie und Automatisierung" der dbv-Kommission Erwerbung und Bestandsentwicklung an der ULB Düsseldorf am 13. März 2017

URL: https://www.bibliotheksverband.de/fileadmin/user\_upload/Kommissionen/Kom\_ErwBest/Tagungen/Erwkomm\_Fortbild\_Ddorf2018\_Muschall.pdf

## Daten im VÖBB: Zwei Beispiele

#### Zeitungsartikel aus der Genios Datenbank



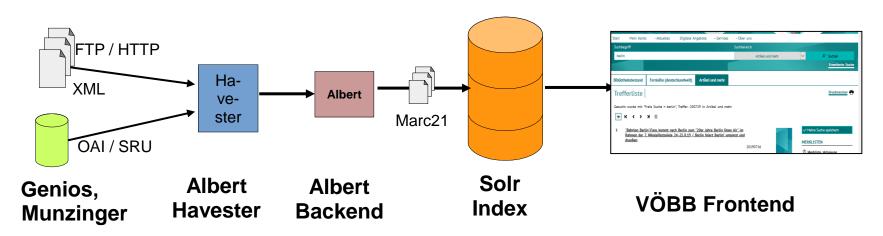
#### Kinderbuch



## Datenobjekt aus der Genios Datenbank

```
<Document TD="77046347-AACHEN" TDNAME="NO" DB="AAN">
<Abstract>n.n.</Abstract>
<Authors><Author>n.n.</Author></Authors>
<Descriptors><Descriptor>n_n./Descriptor>
                                                                                            Datums-
<Date>20200205</pate>
<Issue>n.n.</Issue>
                                                                                            Facette
<ISSN>n.n.</ISSN>
<ISBN>n.n.</ISBN>
<Subtitle>Schwarz-Blaue Funken feiern Biwak und Kostümfest an einem Wochenende</Subtitle>
<Section>Lokales</Section>
<Column>n.n.</Column>
<Lead>n.n.</Lead>
<Headline>n.n.
<Series-Title>n.n.</Series-Title>
<Language>n.n.</Language>
                                                                                            Zeitungs-
<Page>18</Page>
<Publication-Title>Aachener Nachrichten
                                                                                            Facette
<Source>AAN</Source>
<Title>Handgemachter Fastelovvend, der von Herzen kommt</Title>
<Volume>n.n.</Volume>
<Year>2020</Year>
                                                                                            Identifier
<Identifier>AAN 77046347-AACHEN</Identifier>
<Publisher>Aachener Nachrichten</Publisher>
<Copyright>Alle Dokumente genießen nach § 4 und § 87 a ff UrhG urheberrechtlichen
          Schutz. Eine Bearbeitung oder Vervielfältigung ist nicht zulässig.
          <BR/&gt;www.genios.de</Copyright>
</Document>
```

# **Datenfluss – Genios/Munzinger Daten**



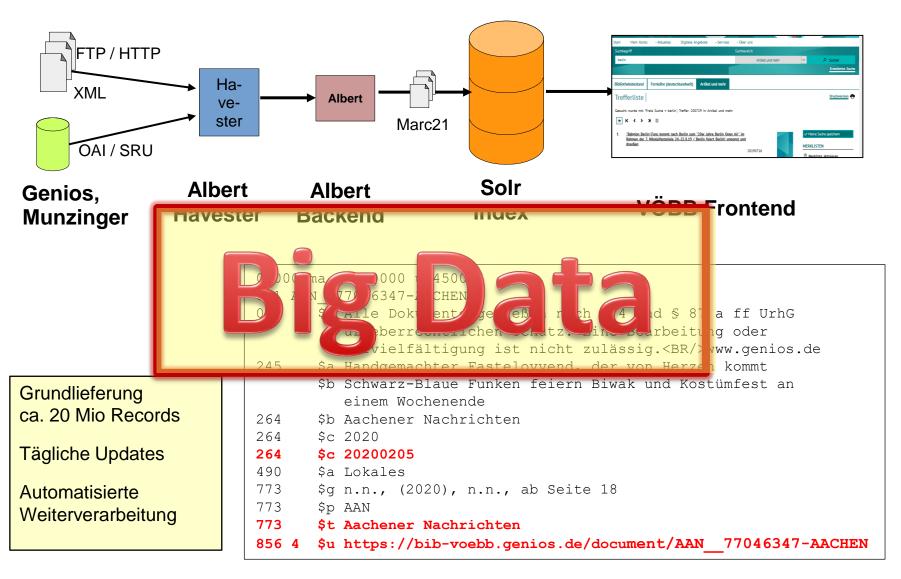
Grundlieferung ca. 20 Mio Records

Tägliche Updates

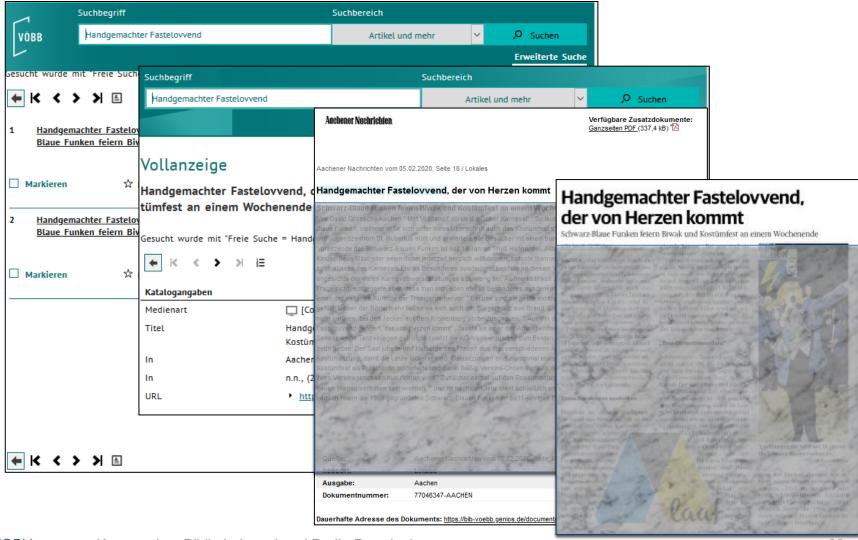
Automatisierte Weiterverarbeitung

```
00000nma a2200000 u 4500
001 AAN 77046347-AACHEN
       $i Alle Dokumente genießen nach § 4 und § 87 a ff UrhG
017
          urheberrechtlichen Schutz. Eine Bearbeitung oder
          Vervielfältigung ist nicht zulässig. < BR/>www.genios.de
       $a Handgemachter Fastelovvend, der von Herzen kommt
245
       $b Schwarz-Blaue Funken feiern Biwak und Kostümfest an
          einem Wochenende
264
       $b Aachener Nachrichten
264
       $c 2020
       $c 20200205
264
490
       $a Lokales
       $g n.n., (2020), n.n., ab Seite 18
773
773
       $p AAN
773
       $t Aachener Nachrichten
       $u https://bib-voebb.genios.de/document/AAN 77046347-AACHEN
856 4
```

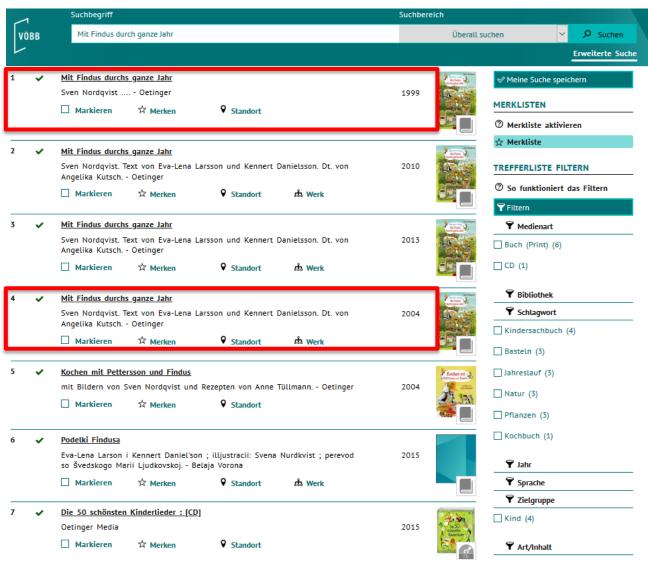
# Datenfluss – Genios/Munzinger Daten



## Präsentation im VÖBB - Portal



# Monographie: Titel: Mit Findus durch ganze Jahr ISBN: 3-7891-6912-9



#### Facetten:

GranulareErschließungerforderlich

#### Frage:

Wieso ist der Treffer 1999 nicht mit dem Werk (2004, 2010, 2013) verbunden?

KOBV .::: Kooperativer Bibliotheksverbund Berlin-Brandenburg

## Datensatz 1: VÖBB-ID 06007650

```
001 ZLB06007650
003 DE-602
005 20180704220833.0
00571nam a2200193uu 4500
007 tu
008 990331s1999
       $a 3789169129 $9 3-7891-6912-9 $c fest geb. : 25,00
020
035 $a (DE-101)956079822
035 $a (DE-599)KBVZLB0600765
     $a DE-602
040
041
       $a ger
049
       $a DE-609;06007650
245 10 $a Mit Findus durchs ganze Jahr $c Sven Nordqvist ...
250
       $a 1
260 3 $a Hamburg $b Oetinger $c ca. 1999
300
       $a ca. 62 Seiten $b zahlr. Ill. (farb.)
500
       $a Aus dem Schwed, übers.
700 1 $0 (DE-588)115632468 $0 (DE-609)00001955 $a Nordqvist, Sven $4 oth
       $a Nodokuvisuto, Suven
909
909
       $a Norudokuvisuto, Suven
```

DNB-Nummer

- Datenübernahme aus der DNB
- Wenige, aber korrekte Daten (ISBN + Erscheinungsjahr)

# Datensatz 2: VÖBB-ID 06007650 – Anzeige Portal

Katalogangaben

Zitierlink

Medienart

Land

| [Buch]

Art/Inhalt Kindersachbuch

Verfasser 

Nordqvist, Sven

Titel Mit Findus durchs ganze Jahr / Sven Nordqvist. Text von Eva-Lena Larsson und Kennert

Danielsson. Dt. von Angelika Kutsch

Werk Nordgvist, Sven <1946->. Nordgvist, Sven <1946->. Pyssla med Findus

Weitere Titel Pyssla med Findus

Person Larsson, Eva-Lena [Textverfasser/in]

Veröffentlichung Hamburg: Oetinger, 2004

Umfang / Format 58 Seiten : zahlr. Ill. ; 31 cm

Sprache Deutsch

Deutschland

ISBN 3-7891-6912-9

Genre Antolin 4. Klasse

Schlagwortkette 

Pflanzen ; Jahreslauf ; Kindersachbuch

Natur ; Jahreslauf ; Kindersachbuch

Basteln

Verbundsystematik 

• 4.3/Spo Sport. Spiel. Basteln. Heimwerken (bis 8 Jahre)

4.3/Spo 951,1 Basteln. Basteln mit und für Zielgruppen. Einzelne Zielgruppen. Kinder

und Jugendliche (bis 8 Jahre)

Inhalt Kater Findus hält den alten Pettersson rund ums Jahr in Atem. Ständig will er etwas ko-

chen, basteln oder untersuchen: Blaubeerkuchen backen, Borkenschiffchen basteln, Steck-

linge setzen. (Deutsch)

Besitzende Bibliotheken 

Standortkarte

#### Facetten:

Granulare Erschließung erforderlich

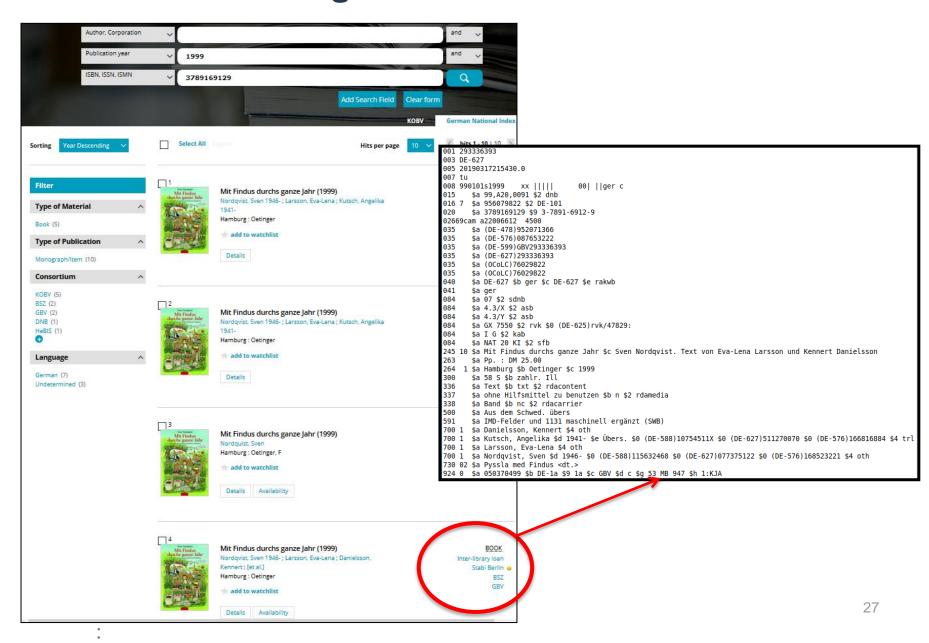
#### Frage:

Werden noch weitere Informationen erfasst, die nicht angezeigt werden?

## Datensatz 2: VÖBB-ID 06007650

```
001 ZLB06191173
003 DE-602
                                                                                                Eigenkatalogisat
005 20191112221354.0
007 tu
008 060725s2004
                  ||||||||ger||
01569nam a2200373uu 4500
                                                                                                  Personen
      $a 3789169129 $9 3-7891-6912-9 $\square$13.99
020
035
      $a (DE-599)KBVZLB06191173
                                                                                                  mit GND IDs
040
      $a DE-602
                                                                                                  und Relationen
041
      $a ger
044
      $c XA-DE
049
      $a DE-609:06191173
100 1 $a Nordqvist, Sven $0 (DE-609)00001955 $0 (DE-588)115632468 $4 aut
240 10 $0 (DE-609)16229316 $a Pyssla med Findus
245 10 $a Mit Findus durchs ganze Jahr $c Sven Nordqvist. Text von Eva-Lena Larsson und Kennert Danielsson. Dt. vo
Angelika Kutsch
246 13 $a Pyssla med Findus
                                                                                                 Jahresangabe
250
      $a 1
260 3 $a Hamburg $b Oetinger $c 2004
                                                                                                 2004
      $a 58 Seiten $b zahlr. Ill. $c 31 cm
300
      $a Kater Findus hält den alten Pettersson rund ums Jahr in Atem. Ständig will er etwas kochen, basteln oder
520
untersuchen: Blaubeerkuchen backen, Borkenschiffchen basteln, Stecklinge setzen.
650 7 $a Basteln $0 (DE-609)00003024 $0 (DE-588)4004668-0 $2 gnd
650 7 $a Jahreslauf $0 (DE-609)00101236 $0 (DE-588)4027996-0 $2 gnd
650 7 $a Natur $0 (DE-609)00002853 $0 (DE-588)4041358-5 $2 and
650 7 $a Pflanzen $0 (DE-609)00702595 $0 (DE-588)4045539-7 $2 gnd
                                                                                                 Granulierte
655 7 $a Kindersachbuch $0 (DE-609)00003026 $2 local
                                                                                                 Sacherschließung
689 00 $0 (DE-588)4045539-7 $D s $a Pflanzen
689 01 $0 (DE-588)4027996-0 $D s $a Jahreslauf
689 02 $A s $a Kindersachbuch $5 DE-101
689 10 $0 (DE-588)4041358-5 $D s $a Natur
689 11 $0 (DE-588)4027996-0 $D s $a Jahreslauf
689 12 $A s $a Kindersachbuch $5 DE-101
689 20 $0 (DE-588)4004668-0 $D s $a Basteln $5 DE-101
700 1 $0 (DE-609)03020958 $a Larsson, Eva-Lena $4 lyr
      $a Nodokuvisuto, Suven
909
      $a Norudokuvisuto, Suven
909
```

## Datenanreicherung aus Verbunddatenbanken / GVI



## Datensatz aus dem K10Plus / Staatsbibliothek

```
ASB-
001 293336393
003 DE-627
                                                                                                              Systematik
005 20190317215430.0
007 tu
008 990101s1999
                                00| ||ger c
                   xx |||||
015
       $a 99,A20,0091 $2 dnb
016 7 $a 956079822 $2 DE-101
                                                                                                                    RVK-
020
       $a 3789169129 $9 3-7891-6912-9
02669cam a22006612 4500
                                                                                                                    Notation
035
       $a (DE-478)952071366
035
       $a (DE-576)087653222
035
       $a (DE-599)GBV293336393
035
       $a (DE-627)293336393
035
       $a (0CoLC)76029822
035
       $a (0CoLC)76029822
040
       $a DE-627 $b ger $c DE-627 $e ra
                                                                                                           Personen
041
       $a ger
084
       $a 07 $2 sdnb
                                                                                                           mit GND IDs
084
       $a 4.3/X $2 asb
084
       $a 4.3/Y $2 asb
                                                                                                           und Relationen
       $a GX 7550 $2 rvk $0 (DE-625)rvk/47829:
084
084
       $a I G $2 kab
084
       $a NAT 20 KI $2 sfb
245 10 $a Mit Findus durchs ganze Jahr $c Sven Nordqvist. Text von Eva-Lena Larse
                                                                                 und Kennert Danielsson
263
       $a Pp. : DM 25.00
    1 $a Hamburg $b Oetinger $c 1999
264
300
       $a 58 S $b zahlr. Ill
336
       $a Text $b txt $2 rdacontent
       $a ohne Hilfsmittel zu benutzen $b n $2 rdamedia
337
338
       $a Band $b nc $2 rdacarrier
500
       $a Aus dem Schwed. übers
591
       $a IMD-Felder und 1131 maschinell erg
700 1 $a Danielsson, Kennert $4 oth
700 l $a Kutsch, Angelika $d 1941- $e Übers. $0 (DE-588)10754511X $0 (DE-627)511270070 $0 (DE-576)166816884 $4 trl
700 1 $a Larsson, Eva-Lena $4 oth
700 l $a Nordqvist, Sven $d 1946- $0 (DE-588)115632468 $0 (DE-627)077375122 $0 (DE-576)168523221 $4 oth
730 02 $a Pyssla med Findus <dt.>
924 0 $a 050370499 $b DE-1a $9 1a $c GBV $d c $g 53 MB 947 $h 1:KJA
```

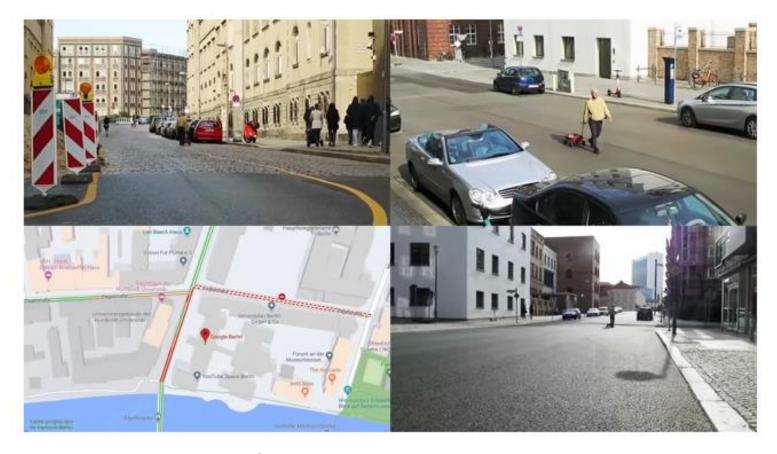
## **Brauchen wir gute Daten?**

- ► Ja!
- Big Data ist kein Voodoo!
- Garbage in Garbage out!

## Was sind gute Daten?

- Sie sind konsistent, gleiche Information im gleichen Kontext / Kategorien ("Cataloger's choice" ist ein Problem)
- Sie enthalten Identifier Diese sind essentiell zur Verknüpfung mit anderen Beständen
- Sie werden tatsächlich benötigt
- Sie lassen sich effizient erfassen
- Wir können mit den Datenfehlern umgehen

## Google Maps: Auch Google kann sich irren



Der Berliner Künstler **Simon Weckert** sorgt mit 99 geborgten Handys in einem Bollerwagen für einen "Stau".

Quelle: YouTube: https://youtu.be/k5eL\_al\_m7Q

## Quellennachweis & Nutzungsrechte

Diese Folien werden nach der Open Access Leitlinie des Kooperativen Bibliotheksverbund Berlin-Brandenburg (ZIB-Report 17-54; URN: urn:nbn:de:0297-zib-65303) unter der Lizenz CC-BY-SA 4.0 veröffentlicht. Davon ausgenommen sind die Abbildungen, Photos sowie Screenshots auf den Folien 2, 3, 9, 17, 18, 22, 23, 27 und 30. Soweit nicht anders vermerkt liegen hier alle Rechte bei den jeweiligen Urhebern.

Folie 2 und 3: Alle Abbildungen und Photos dienen ausschließlich der Illustration und sind sog. "schmückendes Beiwerk". Sie stammen u.a. von der Seite der Verkehrsinformationszentrale Berlin, https://www.vmzberlin.com

Folie 9: Der abgebildete Datenfluss an Universitätsbibliothek der Humboldt Universität Berlin wurde freundlicherweise von Dr. Michael Voß unter der Lizenz CC-BY-SA zur Verfügung gestellt.

Folie 17: Die Statistik stammt aus dem Vortrag von Björn Muschal: Beobachtungen und Gedanken zur Anforderung an bibliographische Metadaten unter https://www.bibliotheksverband.de/fileadmin/user\_upload/Kommissionen/Kom\_ErwBest/Tagungen/Erwkomm\_Fortbild\_Ddorf2018\_Muschall.pdf

Folie 18: Die Abbildungen stammen aus der Genios Datenbank https://bib-voebb.genios.de/document/AAZ\_\_77046347-AACHEN sowie der Homepage der Verlagsgruppe Oetinger, https://www.oetinger.de/buch/mit-findus-durchs-ganze-jahr/9783789169120

Folie 22: Bei Abbildung und Text "Handgemachter Fastelovvend der von Herzen kommt" sind alle Rechte der Medienhaus Aachen GmbH vorbehalten.

Folien 23 und 27: Die Abbildungen der Cover stammen aus Trefferlisten im Verbundkataloges des VÖBB, http://www.voebb.de

Folie 30: Die Abbildungen stammen aus den Google Maps Hacks by Simon Weckert, zu finden auf YouTube unter https://www.youtube.com/watch?v=OnfR3gLlalc

## Vielen Dank für Ihr Interesse

