



Catalog enrichment à la Linked Open Data

License

2



This presentation – inclusive the graphics made by the author, are licensed CC0:
<https://creativecommons.org/about/cc0>

Pictures from <http://www.istockphoto.com/> at slides 5, 7, 8 and 41 are licensed CC-BY-ND:
<http://creativecommons.org/licenses/by-nd/3.0/de/>

Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. <http://lod-cloud.net/>

- Catalog enrichment
 - Definition
 - Technique
 - Matching
 - Linking
- Implementation demo
- Conclusion

- Catalog enrichment
 - Definition
 - Technique
 - Matching
 - Linking
- Implementation demo
- Conclusion



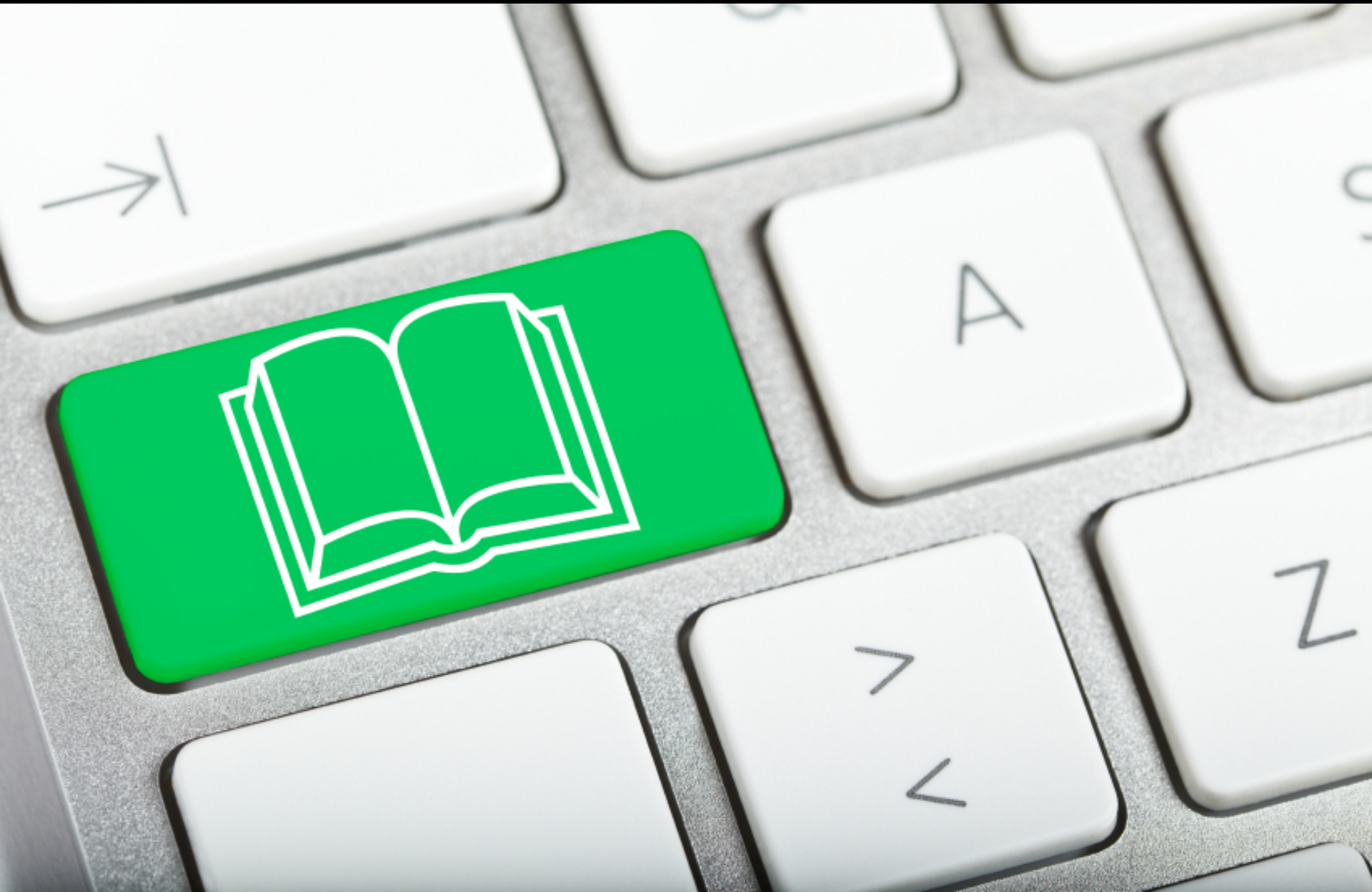
Catalog enrichment ?

Catalog enrichment: definition

6

- Any addendum to the records:
 - links to fulltexts/webpages/...
 - subjects, tags, recensions
 - covers
 - ...
- The source of the addendum does not matter (users, libraries, companies...)
- New features: only indirect

„INSTANT GRATIFICATION“





→ Catalog enrichment

- Definition
- Technique
- Matching
- Linking

→ Implementation demo

→ Conclusion

Catalog enrichment: methods

10

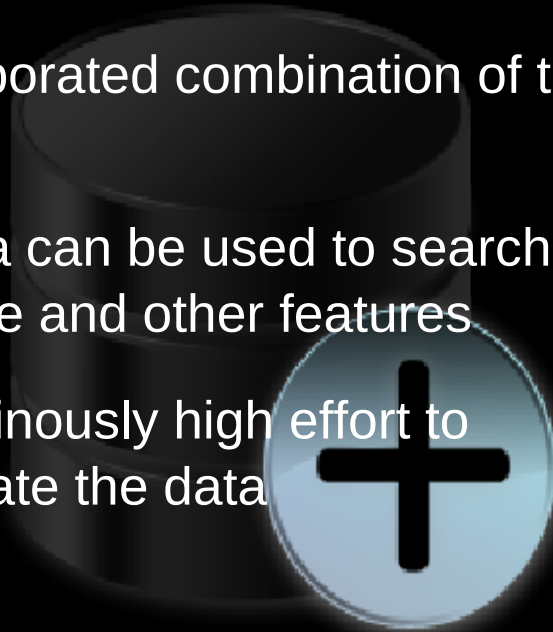


Source of the pictures :<http://findicons.com/about>

database vs. mashup

locale DB:

- + elaborated combination of the data
- + data can be used to search and browse and other features
- continuously high effort to integrate the data



dynamic mashup:

- + data always up-to-date
- + relatively easy to integrate the data
- needs (performant) API
- no search etc.



Bildquelle: <http://findicons.com/about>

RDF based storing with SPARQL endpoint:

- Easy to add data
- Open to be used by customer
- Self-describing data
- SPARQL is a (too?) powerful API

- Catalog enrichment
 - Definition
 - Technique
 - Matching
 - Linking
- Implementation demo
- Conclusion



Source of the picture: <http://www.flickr.com/photos/jhsum-commons/4419490136/>

- triple store with SPARQL Endpoint: 4store
- open data from the hbz union catalog
- 16 M records \Leftrightarrow 1 B Triple
- links to:

5.500 Projekt Gutenberg

12.000 DBpedia

70.000 b3kat

200.000 Dewey Decimal Class.

270.000 DNB Nationalbiografie

420.000 OCLC

1.250.000 Open Library

700.000 ZDB

800.000 LOC Iso-639-2

22.000.000 gnd authority file

32.000.000 lobid-organisations

- Silk
- Culturegraph
- Google-refine
- Hadoop
- ...

Matching algorithms

17

- depending on the data
 - Interesting data reside „elsewhere“
 - => other cataloging rules
- DBpedia example:
 - Creator, ISBN etc. are often missing => only title
 - constraints:
 - german DBpedia
 - category:Literarisches_Werk ,
category:Lexikon,_Enzyklopädie

Problem: disambiguation

18

- matching is to blurry
- Post processing:
 - Allow only bundle with same creator

Bundle having the same creator

19

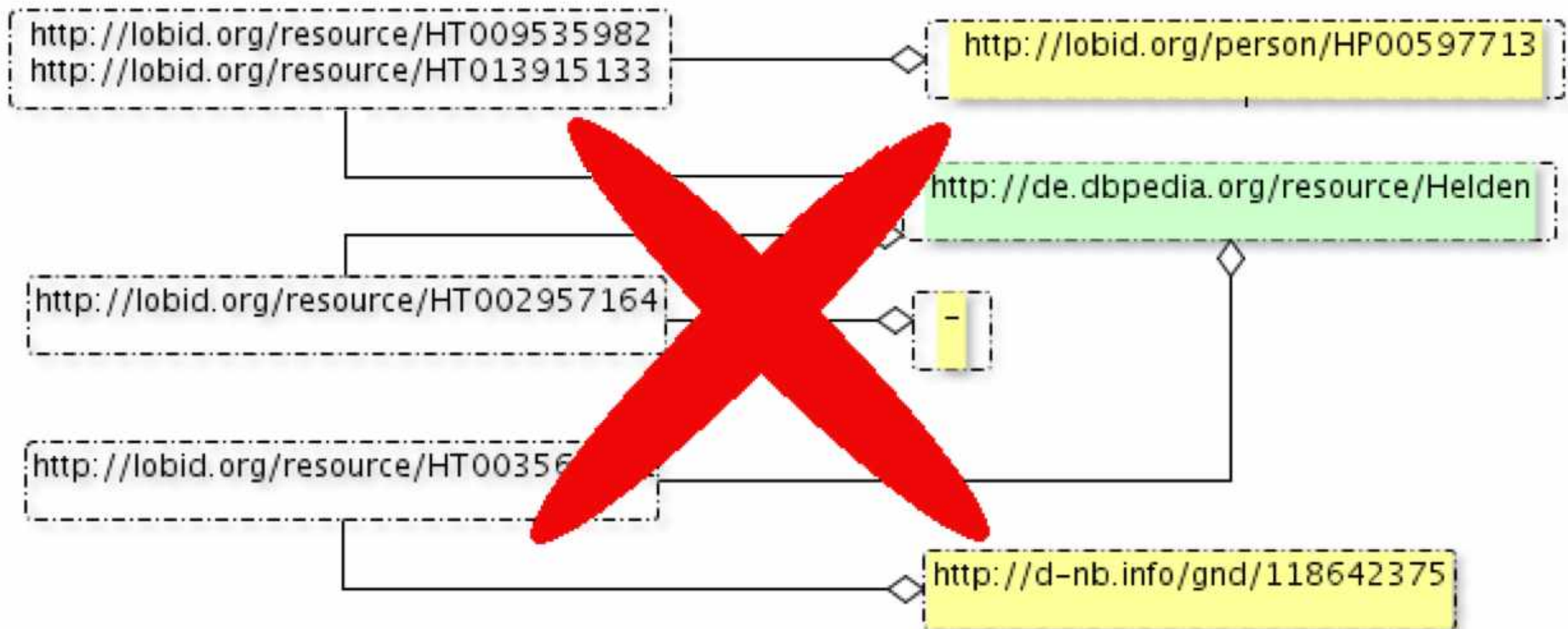
http://de.dbpedia.org/resource/Die_heilige_Johanna_der_Schlachth%C3%B6fe

<http://lobid.org/resource/HT000050080>
<http://lobid.org/resource/HT000050081>
<http://lobid.org/resource/HT002659047>
<http://lobid.org/resource/HT004741084>
<http://lobid.org/resource/HT006212600>
<http://lobid.org/resource/HT008109255>
<http://lobid.org/resource/TT000706133>
<http://lobid.org/resource/TT000706142>
<http://lobid.org/resource/TT001365832>

<http://d-nb.info/gnd/118514768>

Bundle having different creators

20





LOW-HANGING FRUIT

Kai Schreiber, „Reiche Ernte“ 7. August 2005 via Flickr CC BY-SA 2.0

- Catalog enrichment
 - Definition
 - Technique
 - Matching
 - Linking
- Implementation demo
- Conclusion

triplification

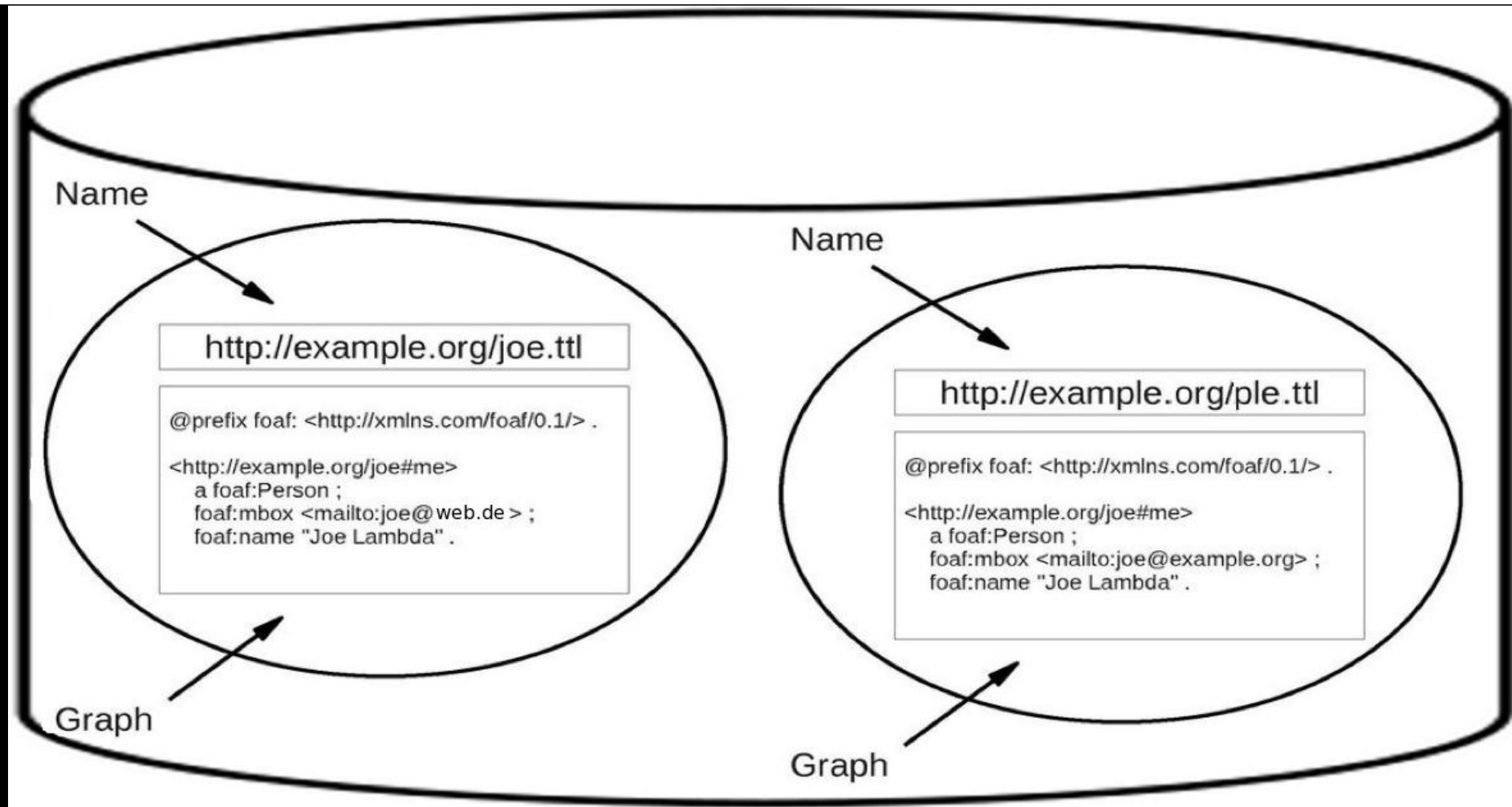
23

- Find predicates or mint them yourself
 - rdrel:workManifested
 - => Triple:
`<lobid-resource> <rdrel:workManifested> <dbpedia-resource>`

- What is the license ?
- Import triples into the SPARQL-Endpoint
 - own „named graph“ has advantages:
 - Easily removable/changeable
 - Provenience is stored
 - Query specific named graphs

Named Graphs

25



What we achieved

26

- **12.000** „sure“ links to **4.000** DBpedia resources => 4.000 new „Work“-levels (21.000 discarded links)
 - average size of a bundle: 3
- links to freebase: **3.000**
- **0.1** % enrichment

What we achieved

27

- **5.500** links zu **400** Project Gutenberg ressourcen (fulltexts in different formats)
 - => **0.05%** enrichment
- **1.200.000** links to the work level of the Open Library
 - => **12.5%** enrichment

What we achieved

28

Sir Tim Berners Lee:



Source of picture: <http://www.w3.org/DesignIssues/LinkedData.html>



LOW-HANGING FRUIT

Kai Schreiber, „Reiche Ernte“ 7. August 2005 via Flickr CC BY-SA 2.0

What we achieved

30

DBpedia example:

„Die Heilige Johanna der Schlachthöfe“



<http://lobid.org/resource/HT000050081>

Titel	Die Heilige Johanna der Schlachthöfe
Autor	http://d-nb.info/gnd/118514768
Illustrator	http://d-nb.info/gnd/118636170
Erscheinungsjahr	1968
Typ	< http://purl.org/ontology/bibo/Book >
Typ	< http://purl.org/vocab/frbr/core#Manifestation >
Sprache	http://id.loc.gov/vocabulary/iso639-2/deu
Ausgabe	1. Aufl.
Erscheinungsort	Leipzig
Verlag	Reclam
Format	print
Umfang	150 S., Ill.

[http://de.dbpedia.org/resource/Die heilige Johanna](http://de.dbpedia.org/resource/Die_heilige_Johanna)

Umfang	http://lobid.org/item/HT000050081%3A12B633
Typ	< http://purl.org/vocab/frbr/core#Item >
Zugehöriger Titel	http://lobid.org/resource/HT000050081
Besitzer	Universitäts- und Stadtbibliothek Köln, Hauptabteilung
Signatur	12B633
Weitere Information	http://www.ub.uni-koeln.de/usbportal?service=search&db=UBKSLNP&query=HBZID%3DHT000050081

Weitere Information http://193.30.112.134/F/?func=find-c&ccl_term=IDN%3DHT000050081

edit

dbpedia-owl:originalLanguage	<ul style="list-style-type: none"> dbpedia:German_language
dbpedia-owl:premiereDate	<ul style="list-style-type: none"> 1959-04-30 (xsd:date)
dbpedia-owl:premierePlace	<ul style="list-style-type: none"> dbpedia:Deutsches_Schauspielhaus dbpedia:Hamburg
dbpedia-owl:premiereYear	<ul style="list-style-type: none"> 1959-01-01 00:00:00 (xsd:date)
dbpedia-owl:settingOfPlay	<ul style="list-style-type: none"> Chicago
dbpedia-owl:subjectOfPlay	<ul style="list-style-type: none"> Capitalism and religion
dbpprop:genre	<ul style="list-style-type: none"> dbpedia:Drama dbpedia:Epic_theatre dbpedia:Non-Aristotelian_drama
dbpprop:name	<ul style="list-style-type: none"> Saint Joan of the Stockyards
dbpprop:origLang	<ul style="list-style-type: none"> dbpedia:German_language
dbpprop:place	
dbpprop:premiere	
dbpprop:setting	
dbpprop:subject	
dbpprop:wikiPage	
dbpprop:writer	
dcterms:subject	<ul style="list-style-type: none"> category:Plays_by_Bertolt_Brecht category:Plays_based_on_the_Legend_of_Jeanne_d'Arc category:1959_plays category:Works_relating_to_Chicago
rdf:type	<ul style="list-style-type: none"> owl:Thing dbpedia-owl:WrittenWork schema:CreativeWork yago:1959Plays dbpedia-owl:Work dbpedia-owl:Play yago:PlaysByBertoldBrecht
rdfs:comment	<ul style="list-style-type: none"> Die heilige Johanna der Schlachthöfe (auch: Die heilige Johanna der Stockyards, deutsch: Die heilige Johanna der Schlachthöfe) ist ein episches Theaterstück von Bertolt Brecht und sein erstes Bühnenstück. Es erzählt die Geschichte der Johanna Dark, die den ausgesperrten Arbeitern auf den Stockyards den Glauben bringt. Es wurde zwischen 1929 und 1931, während der Zeit der Weimarer Republik, geschrieben. Es ist eine Adaption von Brechts Stück Die heilige Johanna der Schlachthöfe (1929). In dieser Version der Geschichte von Jeanne d'Arc, transformiert Brecht die Geschichte in die 20. Jahrhundert Chicago. Die heilige Johanna der Schlachthöfe (1929) ist eine pièce de théâtre du dramaturge allemand Bertolt Brecht. Es wurde zwischen 1929 und 1931, während der Zeit der Weimarer Republik, geschrieben. Es ist eine Adaption von Brechts Stück Die heilige Johanna der Schlachthöfe (1929). In dieser Version der Geschichte von Jeanne d'Arc, transformiert Brecht die Geschichte in die 20. Jahrhundert Chicago. Die heilige Johanna der Schlachthöfe (1929) ist eine pièce de théâtre du dramaturge allemand Bertolt Brecht. Es wurde zwischen 1929 und 1931, während der Zeit der Weimarer Republik, geschrieben. Es ist eine Adaption von Brechts Stück Die heilige Johanna der Schlachthöfe (1929). In dieser Version der Geschichte von Jeanne d'Arc, transformiert Brecht die Geschichte in die 20. Jahrhundert Chicago. Die heilige Johanna der Schlachthöfe (1929) ist eine pièce de théâtre du dramaturge allemand Bertolt Brecht.
rdfs:label	<ul style="list-style-type: none"> Saint Joan of the Stockyards
owl:sameAs	
foaf:name	
foaf:page	<ul style="list-style-type: none"> http://en.wikipedia.org/wiki/Saint_Joan_of_the_Stockyards
is dbpedia-owl:wikiPageRedirects of	dbpedia:St_Joan_of_the_Stockyards

Saint Joan of the Stockyards

NO IMAGE

Saint Joan of the Stockyards (German: Die Heilige Johanna der Schlachthöfe) is a play written by the German modernist playwright Bertolt Brecht between 1929 and 1931, after the success of his musical The Threepenny Opera and during the period of his radical experimental work with the Lehrstücke. It is based on the musical that he co-authored with Elisabeth Hauptmann, Happy End (1929). In this version of the story of Joan of Arc, Brecht transforms... [More](#)

 [Read article at Wikipedia](#)

Author: [Bertolt Brecht](#)

Genre: [Drama](#)

Also known as: [St. Joan of the Stockyards](#)



Bertolt Brecht

Bertolt Brecht

February 1898

theatre practiti

Written Work

[breçt] (listen); born Eugen Berthold Friedrich Brecht (he was a German poet, playwright, and theatre director. An tury, Brecht made equally significant contributions to...

What we achieved

34

Open Library example:

„With reference to reference“

With reference to reference [1 edition](#)

Edit

 Last edited by [Open Library Bot](#)
 28. April 2010 | [History](#)
By [Catherine Z. Elgin](#)

SUBJECTS

[Reference \(Philosophy\)](#)

PEOPLE

[Nelson Goodman](#)

Excerpts

A theory of reference should identify and characterize the relations between a language (or, more broadly, a symbol system) and its objects, and explain the ways the language functions in or contributes to our understanding of those objects.

added anonymously.

 There is only 1 edition record, so we'll show it here... • [Add edition?](#)

With reference to reference

Catherine Z. Elgin ; with a foreword by Nelson Goodman.

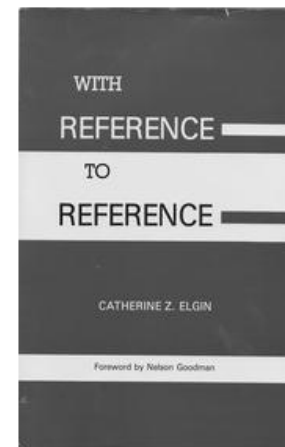
Published **1983** by [Hackett Pub. Co.](#) in [Indianapolis, Ind.](#)Written in [English](#).

First Sentence

A theory of reference should identify and characterize the relations between a language (or, more broadly, a symbol system) and its objects, and explain the ways the language functions in or contributes to our understanding of those objects.

Table of Contents

	Foreword	1
	Preface	3
I	Problem and Project	5
1.	The Subject	5
2.	Roads not taken	10
II	On Denoting	19

[Manage Covers](#)

Read

No readable version available.



Borrow

[Physical copy, local](#) WorldCat

Buy

[Alibris](#)[Amazon](#)[AbeBooks](#)[Biblio.com](#)[Book Depository](#)[Powells](#)

Linking Example: LODUM

36

FACTS

about <http://data.uni-muenster.de/context/cris/publication/41562>

>> [show simple property table](#)

Textdetektive und LEKOLEMO – Ansätze zur Förderung von

Abstract: Es werden zwei Programme vorgestellt, die darauf abzielen, durch die Vermittlung von Textdetektiven und sechsten Klassen vorgesehen, während das Programm LEKOLEMO zur Förderung von Kleingruppentrainings für Schüler aus siebten Klassen entwickelt wurde. Beide Programme betonen die Notwendigkeit einer Förderung der Selbstregulation der kognitiven und metakognitiven Lesestrategien und der instruktionalen Rahmenbedingungen.

Subject: undefined

Issued: 2007

Isbn: 978-3-17-019404-5

Publisher: <http://data.uni-muenster.de/context/cris/publication/publisher/E8B553975FF8>

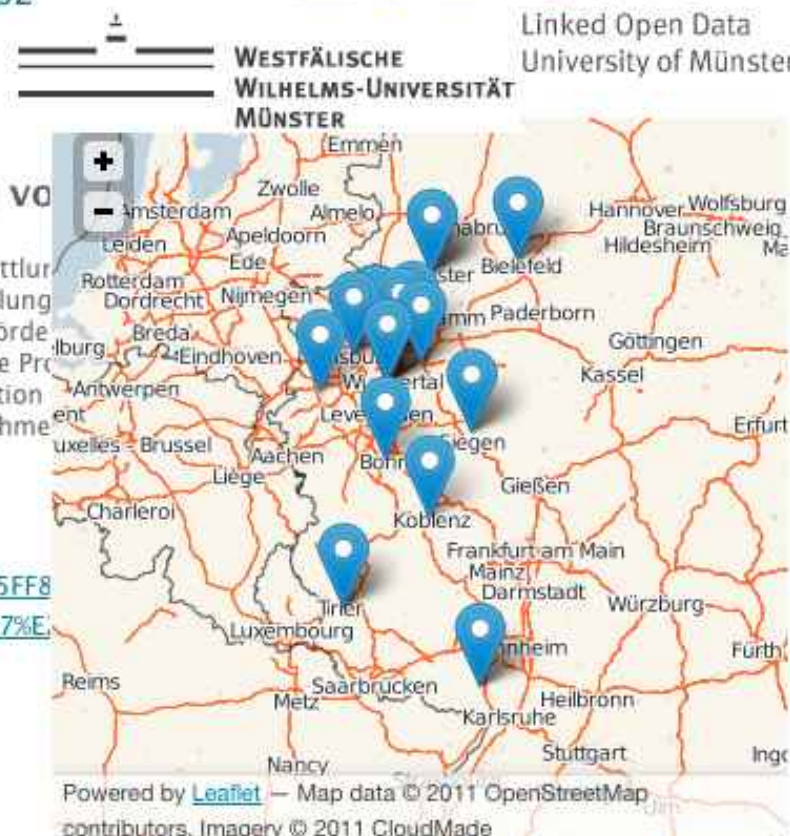
Download: <http://search.ebscohost.com/login.aspx?direct=true&db=pdx&AN=0198157%E>

Author(s)

- [Souvignier, Elmar](#)

LODUM

Linked Open Data
University of Münster



Integration into the catalog

37

- What is *allowed* ?
- What *should* be integrated, what not?
- Human readable presentation of the links/URIs
- (some) data *should* be indexed locally (e. g. to be able to search)
- ...

- Catalog enrichment
 - Definition
 - Technique
 - Matching
 - Linking
- Implementation demo
- Conclusion

Implementation demo

39

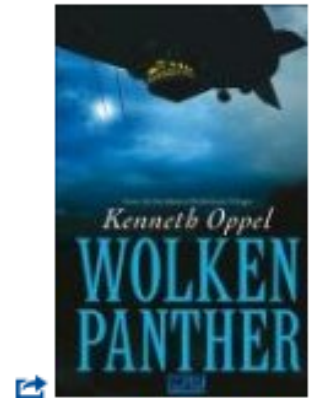
001. [Wolkenpanther](#)

Bibliographische Angaben

Angaben zum Inhalt

Bestand / Besitz

Titel:	Wolkenpanther / Kenneth Oppel. Aus d. Engl. von Anja Hansen-Schmidt
Einheitssachtitel:	Airborn
Autor(in):	Oppel, Kenneth ; Hansen-Schmidt, Anja
Link:	Wikipedia Seite zum Werk
Ausgabe:	[Nachdr.] - Weinheim : Beltz & Gelberg, 2005
Umfang:	549 S., 211 mm x 137 mm
ISBN:	3-407-80941-7



+ [merken](#) [drucken](#) [exportieren](#) [Verfügbarkeit prüfen](#)

Implementation demo

40

001. [Wolkenpanther](#)

Bibliographische Angaben

Angaben zum Inhalt

Bestand / Besitz



Schlagwörter:

- Abenteuer
- Jugend
- Luftschiff
- Überfall
- Seeräuber
- Flucht

Zusammenfassung: Wolkenpanther ist ein Roman des kanadischen Schriftstellers Kenneth Oppel. Erschienen ist es unter dem Originaltitel Airborn 2004. In Deutschland kam es 2005 bei Beltz und Gelberg heraus, übersetzt von Anja Hansen-Schmidt. (via Wikipedia)

+ [merken](#) [drucken](#) [exportieren](#) [Verfügbarkeit prüfen](#)

- Catalog enrichment
 - Definition
 - Technique
 - Matching
 - Linking
- Implementation demo
- Conclusion





N.Y. PLAYGROUND

Everything that's possible with LOD could also be achieved without LOD.

It's just *easier* with LOD.

Ad astra ? Ad data !

To boldly go where no data has gone before.

Source of the picture: <http://hubblesite.org/gallery/album/star/pr2006050d>

Hubble
Heritage

NASA, ESA, and The Hubble Heritage Team (STScI/AURA) • Hubble Space Telescope ACS • STScI-PRC06-50b

Open source

46



<http://sourceforge.net/projects/culturegraph/>



<http://4store.org/>

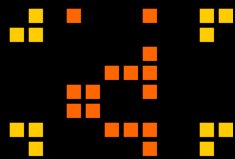


<https://github.com/lobid/>

Silk

<https://www.assembla.com/spaces/silk>

Thank you !



Pascal Christoph
christoph@hbz-nrw.de

semweb@hbz-nrw.de

list of references

- KiM: Empfehlungen zur Öffnung bibliothekarischer Daten
<https://wiki.d-nb.de/pages/viewpage.action?pageId=45419980>
- Till Kreutzer (2010): Open Data – Freigabe von Daten aus Bibliothekskatalogen
<http://www.hbz-nrw.de/dokumentencenter/veroeffentlichungen/open-data-leitfaden.pdf>
- Adrian Pohl (2010): *Open Data im hbz-Verbund*. Erschienen in: ProLibris. 3. Preprint:
http://www.hbz-nrw.de/dokumentencenter/produkte/lod/aktuell/pohl_2010_open-data.pdf
- Tim Berners Lee's talk of Open Data (2010): <http://www.youtube.com/watch?v=3YcZ3Zqk0a8>
- Jansen / Christoph: Dynamische Kataloganreicherung auf Basis von Linked Open Data
http://de.slideshare.net/h_jansen/dynamische-kataloganreicherung-auf-basis-von-linked-open-data
- Blog post: First results using SILK to link to DBpedia
<https://wiki1.hbz-nrw.de/display/SEM/2012/05/03/First+results+using+SILK+to+link+to+DBpedia>
- Blog post: 1.2 M links to Open Library
<https://wiki1.hbz-nrw.de/display/SEM/2012/05/23/1.2+M+links+to+Open+Library>
- Oliver Flimm (2010): LOD und die Open Library <http://de.slideshare.net/flimm/lod-openlibrary20100512>
- Directory of data „thedatahub“ aka CKAN: <http://www.thedatahub.org/>
- 49 bibliographic data sources as LOD <http://thedatahub.org/group/bibliographic?tags=lod>