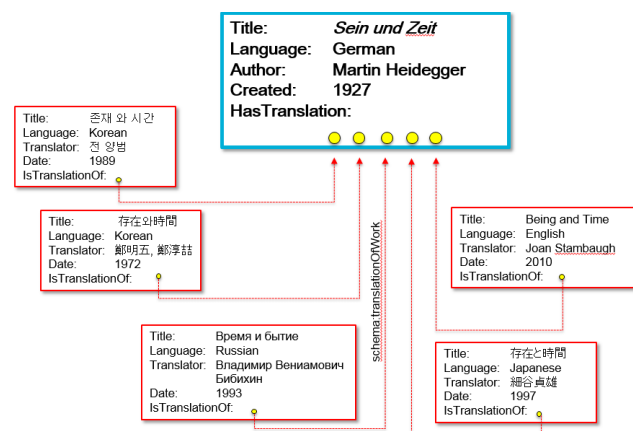


International Conference on Dublin Core and Metadata Applications 2017
Washington DC, 26 October 2017

Using the Semantic Web to Improve Knowledge of Translations

Karen Smith-Yoshimura

OCLC Research



Challenges

- Different writing systems, different transliterations
- Metadata not always good enough
- Linked data opportunities



<https://www.nasa.gov/image-feature/africa-and-europe-from-a-million-miles-away>

Why focus on translations?

Why focus on translations?

For Developers:

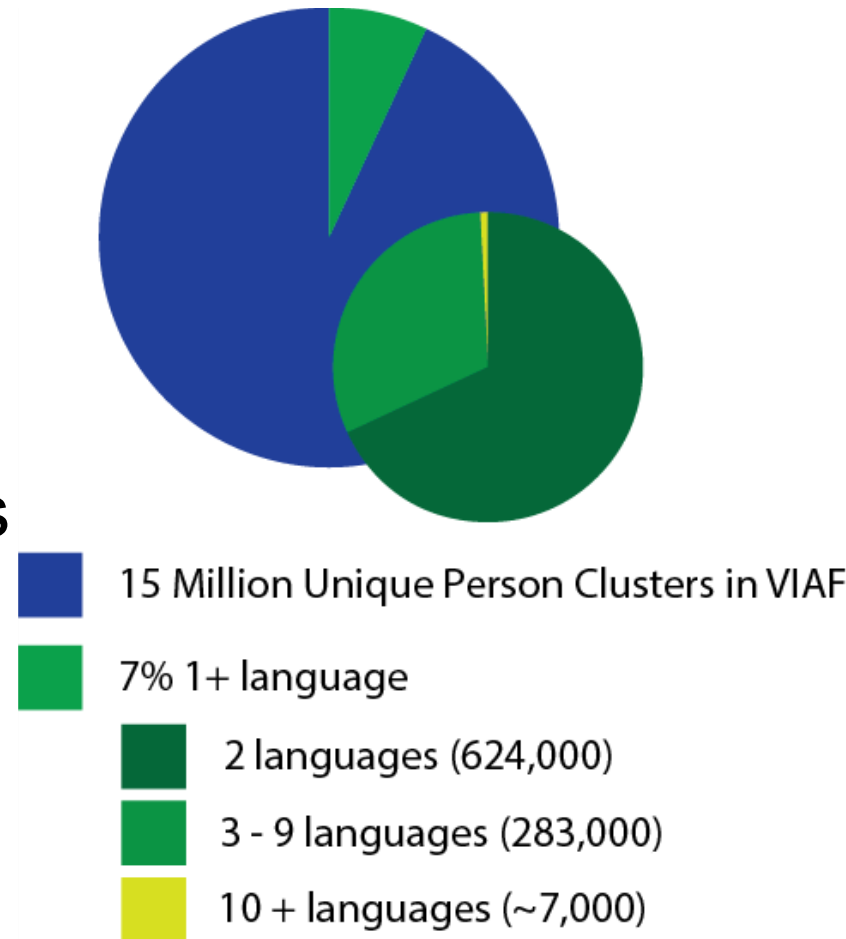
Present information
in the preferred
language &
script of the user

For Academics:

Understand
information sharing
across cultures

Translations

Leo Tolstoy: 97 languages
Rabindranath Tagore: 93
Homer: 84 languages
Mahatma Gandhi: 52 languages
Isaac Bashevis Singer: 52
Najīb Maḥfūz: 47 languages
Cao Xueqin: 27 languages
Murasaki Shikabu: 21 languages



[HangingTogether](#), 2013-11-12
[By January 2017, person clusters
have increased to 45 million]

Ιλιάδα

The Iliad

紅樓夢

Dream of the Red Chamber

زقاق المدق

Midaq Alley

ঘরে বাইরে

The Home and the World

דער בעל-תשובה

The Penitent

Война и миръ

War and Peace

સત્યના પ્રયોગો અથવા આત્મકથા

The Story of My Experiments with Truth

[Gandhi autobiography]

源

氏

物

語

The Tale of Genji

Title: *Sein und Zeit*
Language: German
Author: Martin Heidegger
Created: 1927

Title: 존재 와 시간
Language: Korean
Translator: 전 양범
Date: 1989
IsTranslationOf: ●

Title: 存在와時間
Language: Korean
Translator: 鄭明五, 鄭淳喆
Date: 1972
IsTranslationOf: ●

Title: Время и бытие
Language: Russian
Translator: Владимир Вениамович Биbihин
Date: 1993
IsTranslationOf: ●

Title: Being and Time
Language: English
Translator: Joan Stambaugh
Date: 2010
IsTranslationOf: ●

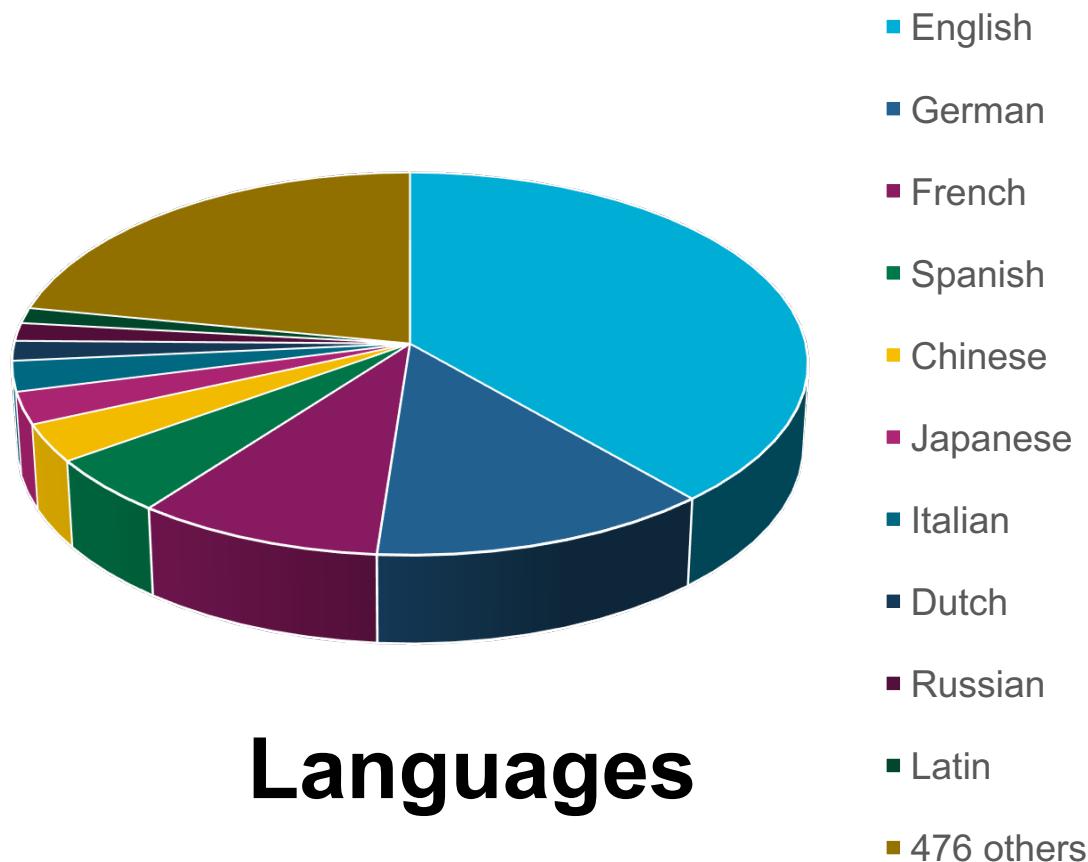
Title: 存在と時間
Language: Japanese
Translator: 細谷 貞雄
Date: 1997
IsTranslationOf: ●

schema:translationOfWork

WorldCat today



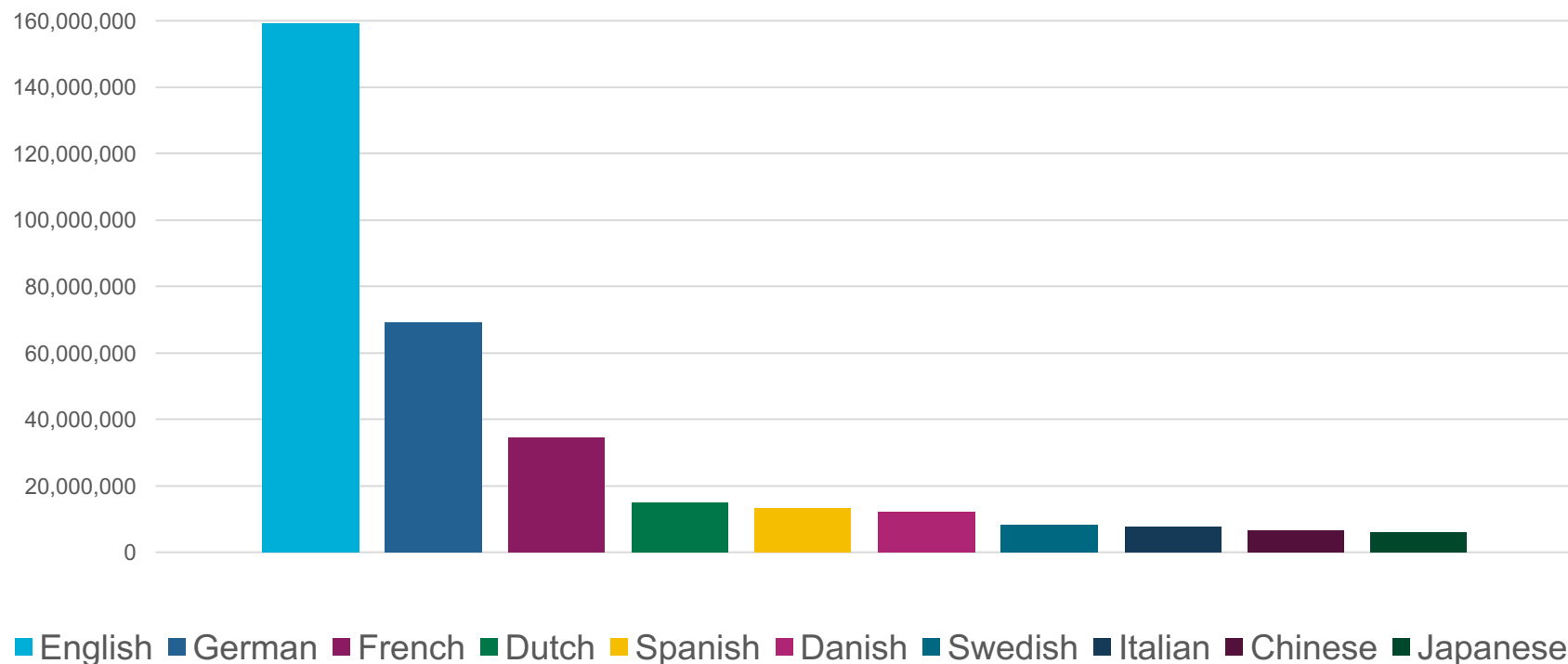
- Resources in nearly all languages
- More than 2.5 billion holdings contributed by libraries worldwide
- More than half the database is for works not in English



Languages

April 2017

Language of cataloging



Language of cataloging and subject headings

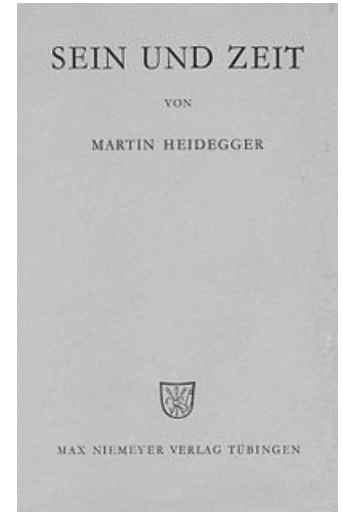
Sein und Zeit by Martin Heidegger

Filosofía alemana [[@es](#), Spanish]

Fundamentalontologie.

Ontologie. [[@de](#), German]

哲学思想 [[@zh](#), Chinese]



001 469621251
 009 16467
 040 BDF \$e ncafnor \$b fre \$c BDF \$d FRMCM
 015 00432988 \$2 bnf
 016 7 FRBNF391946830000009 \$2 FrPBN
 020 2013222718 (br)
 020 9782013222716 (br)
 029 0 BDF \$b FRBNF391946830000009
 029 1 FRMCM \$b FRMCM1/321183
 082 14 843
 100 1 Loti, Pierre, \$d 1850-1923. \$4 aut \$0 (FrPBN)11913405
 245 10 Pêcheur d'Islande / \$c Pierre Loti.
 260 [Paris] : \$b Hachette jeunesse, \$c 2004 \$e (27-Evreux : \$f Impr. Hérissé)
 300 348 p. : \$b carte, couv. ill. en coul. ; \$c 18 cm.
 490 0 Le livre de poche, Jeunesse, Classique
 520 Après avoir vécu à Paris, Gaud, une belle femme, revient vivre en Bretagne et tombe amoureuse de Yann, marin sur le bateau "la Marie" qui navigue dans la mer du Nord, connue pour ses dangers, vers l'Islande. Yann n'aime que la mer... Ce classique est à la fois un récit de voyage et un roman d'amour.
 760 0 \$t Le Livre de poche. Jeunesse. Classique, \$d 2004 \$x 1761-8673 \$w (FrPBN)39034252
 760 0 \$t Le Livre de poche. Jeunesse \$x 0223-7091 \$g 1142 \$w (FrPBN)34227359
 760 0 \$t Le Livre de poche, \$d 2004 \$x 0248-3653 \$w (FrPBN)34228560
 856 42 \$3 Notice et cote du catalogue de la Bibliothèque nationale de France \$u <http://catalogue.bnf.fr/ark:/12148/cb39194683s>
 886 1 \$2 intermrc \$a 009 \$b a-f-j-----
 886 2 \$2 intermrc \$a 680 \$b 0- \$a 809 \$b 809

Language of cataloguing = fre

Leveraging language of cataloging

040 THC \$b eng \$c THC \$d OCLCQ \$d OCLCG \$d VOV \$d IDA \$d LTU \$d OCLCQ
 043 e-ic---
 050 4 PQ2472 \$b .P413 1969
 082 04 844.08105552
 100 1 Loti, Pierre, \$d 1850-1923.
 245 10 Iceland fisherman / \$c Pierre Loti ; translated by W.P. Baines.
 260 London : \$b Dent ; \$a New York : \$b Dutton, \$c 1969.
 300 vii, 242 p. ; \$c 19 cm.
 490 1 Everyman's library ; \$v 920
 500 Translation of Pêcheur d'Islande.
 504 Includes bibliographical references (vi).
 520 "Iceland Fisherman (1886) was Loti's seventh novel and first considerable success. An impressionist writer of pervasive melancholy and great descriptive power, he served for many years in the French Navy, where he learned to know the sea in all its moods and to observe at first hand the scenic contrasts which are an outstanding feature of his books. In this tale, the hero, Yann, sturdy but wayward, and the girl Gaud, simple and complex by turns, reflect the ever-changing atmosphere of the element by which they live. There is also light, color and sound in Loti's magnificent pictures of the Icelandic seas, pictures which Laurence Bisson has described as 'a system of suggestion and evocation not without resemblance to the Symbolist technique.'
 650 0 Fishers \$z Iceland \$v Fiction.
 650 0 French fiction \$y 19th century \$v Translations into English.
 651 0 Iceland \$v Fiction.
 830 0 Everyman's library ; \$v 920.

Language of cataloguing
= eng

001 469621251
 009 16467
 040 BDF \$e ncafnor \$b fre \$c BDF \$d FRMCM
 015 00432988 \$2 bnf
 016 7 FRBNF391946830000009 \$2 FrPBN
 020 2013222718 (br)
 020 9782013222716 (br)
 029 0 BDF \$b FRBNF391946830000009
 029 1 FRMCM \$b FRMCM1/321183
 082 14 843
 100 1 Loti, Pierre, \$d 1850-1923. \$4 aut \$0 (FrPBN)11913405
 245 10 Pêcheur d'Islande / \$c Pierre Loti.
 260 [Paris] : \$b Hachette jeunesse, \$c 2004 \$e (27-Evreux : \$f Impr. Hérissé)
 300 348 p. : \$b carte, couv. ill. en coul. ; \$c 18 cm.
 490 0 Le livre de poche, Jeunesse, Classique
 520 Après avoir vécu à Paris, Gaud, une belle femme, revient vivre en Bretagne et tombe amoureuse de Yann, marin sur le bateau "la Marie" qui navigue dans la mer du Nord, connue pour ses dangers, vers l'Islande. Yann n'aime que la mer... Ce classique est à la fois un récit de voyage et un roman d'amour.
 760 0 \$t Le Livre de poche. Jeunesse. Classique, \$d 2004 \$x 1761-8673 \$w (FrPBN)39034252
 760 0 \$t Le Livre de poche. Jeunesse \$x 0223-7091 \$g 1142 \$w (FrPBN)34227359
 760 0 \$t Le Livre de poche, \$d 2004 \$x 0248-3653 \$w (FrPBN)34228560
 855 42 \$3 Notice et cote du catalogue de la Bibliothèque nationale de France. \$u http://catalogue.bnf.fr/ark:/12148
 886 1 [schema:description](#) "Après avoir vécu à Paris, Gaud, une belle femme, revient vivre en Bretagne et tombe amoureuse de Yann, marin sur le bateau "la Marie" qui navigue dans la mer du Nord, connue pour ses dangers, vers l'Islande. Yann n'aime que la mer... Ce classique est à la fois un récit de voyage et un roman d'amour."
 886 2

Language of cataloguing = fre

@fr

[schema:description](#) ""Iceland Fisherman (1886) was Loti's seventh novel and first considerable success. An impressionist writer of pervasive melancholy and great descriptive power, he served for many years in the French Navy, where he learned to know the sea in all its moods and to observe at first hand the scenic contrasts which are an outstanding feature of his books. In this tale, the hero, Yann, sturdy but wayward, and the girl Gaud, simple and complex by turns, reflect the ever-changing atmosphere of the element by which they live. There is also light, color and sound in Loti's magnificent pictures of the Icelandic seas, pictures which Laurence Bisson has described as 'a system of suggestion and evocation not without resemblance to the Symbolist technique.'"
 @en

@en

001 469621251
 040 THC \$b eng \$c THC \$d OCLCQ \$d OCLCQ \$d VOV \$d IDA \$d LTU \$d OCLCQ
 650 0 Fishers \$z Iceland \$v Fiction.
 650 0 French fiction \$y 19th century \$v Translations into English.
 651 0 Iceland \$v Fiction.
 830 0 Everyman's library ; \$v 920.

Many languages in WorldCat written in non-Latin scripts

Language	% with Original Script	% with Transliteration Only
Chinese	92%	8%
Japanese	88%	12%
Russian	12%	88%
Hebrew	77%	23%
Arabic	54%	46%
Thai	67%	33%
Korean	76%	24%
Greek	19%	81%
Persian	31%	69%
Ukrainian	8%	92%
Bulgarian	25%	75%
Hindi	1%	99%
Yiddish	61%	39%
Tamil	25%	75%
Urdu	11%	89%

홍길동전

Ιλιάδα

زقاق المدق

Война и миръ

דער בעל-תשובה

紅樓夢

พระอภัยมณี

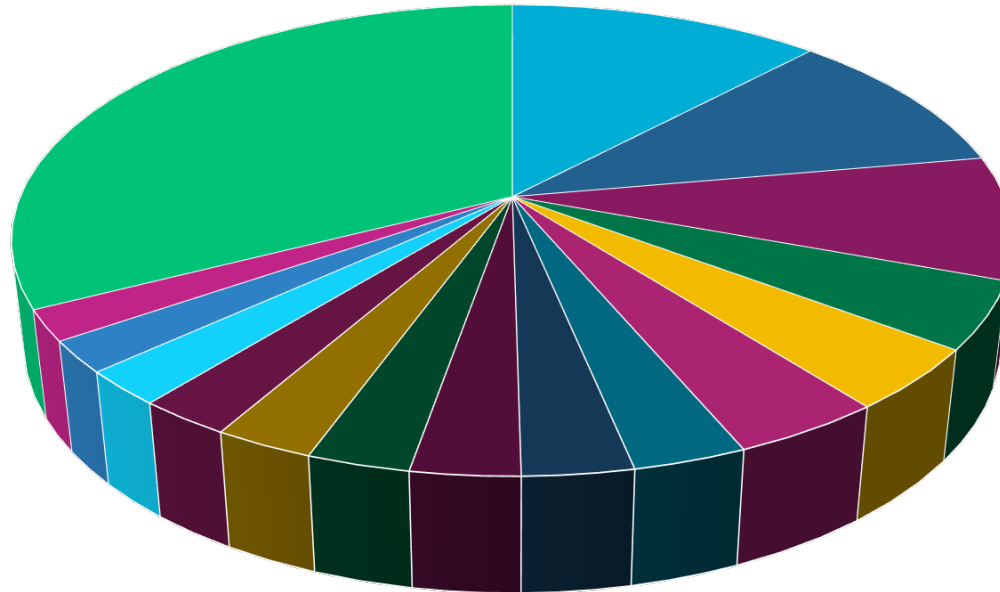
源

氏

物

語

Wikipedia languages



English

Dutch

Spanish

Japanese

Cebuano

French

Waray-Waray

Portuguese

Swedish

Russian

Polish

Chinese

German

Italian

Vietnamese

All other 279 languages

As of 2017-05-31

Multilingual Linked Data Dataset

The Grand Design by Stephen Hawking and Leonard Mlodinow

81 translations in 24 languages

Les mots et les choses: Une archéologie des sciences humaines

by Michel Foucault

293 translations in 20 languages

Pêcheur d'Islande by Pierre Loti

494 translations in 34 languages

Principia philosophiae by René Descartes

889 translations in 14 languages

Sein und Zeit by Martin Heidegger

570 translations in 33 languages

Multilingual linked data dataset

Data extracted
from 4,073
WorldCat
records

Enhanced by
data from
Wikidata



SPARQL ENDPOINT

http://localhost:3030/multilingual/query

CONTENT TYPE (SELECT)

JSON

CONTENT TYPE (GRAPH)

Turtle

7

SELECT DISTINCT ?wd ?label (COUNT(DISTINCT ?translation) AS ?translationCount) (COUNT(DISTINCT ?translationLanguage) AS ?numTranslationLanguages) (COUNT(?geoLocation) AS ?numGeolocations)

8

WHERE {

9

?wd wdt:P243 ?ocn;

10

wdt:P364 ?originalLanguage;

11

rdfs:label ?label.

12

?originalLanguage wdt:P218 ?originalLanguageCode.

13

FILTER(LANG(?label) = ?originalLanguageCode)

14

15

?wc library:oclcnum ?ocn.

16

17

OPTIONAL {

18

?wc dct:mediator ?geoLocation

19

}

QUERY RESULTS

Table

Raw Response

Showing 1 to 5 of 5 entries

Search:

Show

All

entries

	wd	label	translationCount	numTranslationLanguages	numGeolocations
1	wd:Q93773	"The Grand Design"@en	"80"^^xsd:integer	"24"^^xsd:integer	"5229"^^xsd:integer
2	wd:Q3412135	"Pêcheur d'Islande"@fr	"493"^^xsd:integer	"34"^^xsd:integer	"1592"^^xsd:integer
3	wd:Q1195384	"Principia philosophiae"@la	"888"^^xsd:integer	"14"^^xsd:integer	"2223"^^xsd:integer
4	wd:Q404567	"Sein und Zeit"@de	"566"^^xsd:integer	"33"^^xsd:integer	"2669"^^xsd:integer
5	wd:Q1215306	"Les Mots et les Choses"@fr	"293"^^xsd:integer	"20"^^xsd:integer	"1454"^^xsd:integer

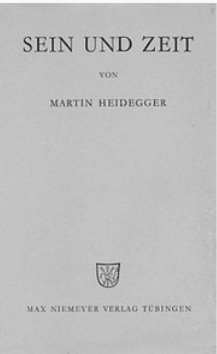
Showing 1 to 5 of 5 entries

1 876191532
8 19970719s1997 it 000 0 ita d
40 [0] __ [1\$a] ITCPE [2\$b] ita [3\$c] ITCPE
20 [0] __ [1\$a] 883040943X

20 [0] __ [1\$a] 9788830409439	1 9788304389
82 [0] 14 [1\$a] 111 [29\$2] 19	8 20150720e201405uuit b 001 0 ita d
100 [0] 1_ [1\$a] Heidegger, Martin. [31\$4] aut	40 [0] __ [1\$a] STF [2\$b] eng [5\$e] rda [3\$c] STF [4\$d] OCLCA
245 [0] 10 [1\$a] Essere e tempo / [3\$c] di Martin Heidegger ;	20 [0] __ [26\$z] 9788830415447
250 [0] __ [1\$a] 12. ed.	41 [0] 1_ [1\$a] ita [8\$h] ger
260 [0] __ [1\$a] Milano : [2\$b] Longanesi, [3\$c] stampa 1997	50 [0] _4 [1\$a] B3279.H48 [2\$b] S4316 2014
300 [0] __ [1\$a] XXXIX, 558 p. ; [3\$c] 21 cm.	100 [0] 1_ [1\$a] Heidegger, Martin, [4\$d] 1889-1976, [5\$e] author.
490 [0] 0_ [1\$a] I marmi [22\$v] 155	240 [0] 10 [1\$a] Sein und Zeit. [12\$I] Italian. [6\$f] 2014
500 [0] __ [1\$a] Copyr. 1976.	245 [0] 10 [1\$a] Essere e tempo / [3\$c] di Martin Heidegger ; a cura di Franco Volpi sulla versione di Pietro Chiodi con le glosse a margine dell'autore ; traduzione di Pietro Chiodi.
653 [0] 0_ [1\$a] Essere.	250 [0] __ [1\$a] 7. edizione.
700 [0] 1_ [1\$a] Chiodi, Pietro. [31\$4] aut	264 [0] _1 [1\$a] Milano : [2\$b] Longanesi, [3\$c] maggio 2014.
	300 [0] __ [1\$a] 632 pages ; [3\$c] 22 cm.
	336 [0] __ [1\$a] text [29\$2] rdacontent
	337 [0] __ [1\$a] unmediated [29\$2] rdamedia
	338 [0] __ [1\$a] volume [29\$2] rdacarrier
	490 [0] 1_ [1\$a] I marmi ; [22\$v] v. 182
	500 [0] __ [1\$a] Translation of: Sein und Zeit / Martin Heidegger.
	504 [0] __ [1\$a] Includes bibliographical references (pages 553-579) and index.
	650 [0] _0 [1\$a] Ontology.
	650 [0] _0 [1\$a] Space and time.
	650 [0] _7 [1\$a] Ontology. [29\$2] fast [27\$0] (OCoLC)fst01045995
	650 [0] _7 [1\$a] Space and time. [29\$2] fast [27\$0] (OCoLC)fst01127622
	700 [0] 1_ [1\$a] Volpi, Franco, [4\$d] 1952-2009, [5\$e] editor.
	700 [0] 1_ [1\$a] Chiodi, Pietro, [5\$e] translator.

- No original language
- No original title
- Chiodi translator, not author

Being and Time



Cover of the first edition

Author Martin Heidegger

Original title *Sein und Zeit*

Translator 1962: John Macquarrie and Edward Robinson
1996: Joan Stambaugh

Country Germany

Language German

Subject Hermeneutics, phenomenology

Published 1927 (in German)
1962: SCM Press
1996: State University of New York Press
2008: Harper Perennial Modern Thought

Pages 589 (Macquarrie and Robinson translation)
482 (Stambaugh translation)

ISBN 0-631-19770-2 (Blackwell edition)
978-1-4384-3276-2 (State University of New York Press edition)

Followed by *Kant and the Problem of Metaphysics*

All “translationOfWork”

Label



Εἶναι και Χρόνος

Instead of WorldCat's
Einei kai chronos

Wikipedia (30 entries) edit

ar	الوجود والزمان
cs	Bytí a čas
da	Væren og tid
de	Sein und Zeit
el	Εἶναι και Χρόνος
en	Being and Time
eo	Esto kaj tempo
es	El ser y el tiempo
et	Olemine ja aeg
fa	هستی و زمان
fi	Oleminen ja aika
fr	Être et Temps
fy	Sein und Zeit
he	הוויה וזמן
hu	Lét és idő
it	Essere e tempo
ja	存在と時間
kk	Болмыс және уақыт
ko	존재와 시간
la	Sein und Zeit
nl	Zijn en Tijd
no	Væren og tid
pa	ਵਜੂਦ ਅਤੇ ਵਕਤ
pt	Ser e Tempo
ru	Бытие и время
simple	Being and Time
tr	Varlık ve Zaman
zh_min_nan	Chûn-chāi kap Sî-kan
zh_yue	存在與時間
zh	存在與時間



Markup for the semantic web

Original Work (in Chinese)

<<http://worldcat.org/entity/work/id/1215997>>

a schema:CreativeWork;

schema:creator <<http://viaf.org/viaf/102266649>> ; # "Gao, Xingjian"

schema:inLanguage "zh";

schema:name "靈山"@zh-hant.

Translated Work (in English)

<<http://worldcat.org/entity/work/id/145209748>>

a schema:CreativeWork;

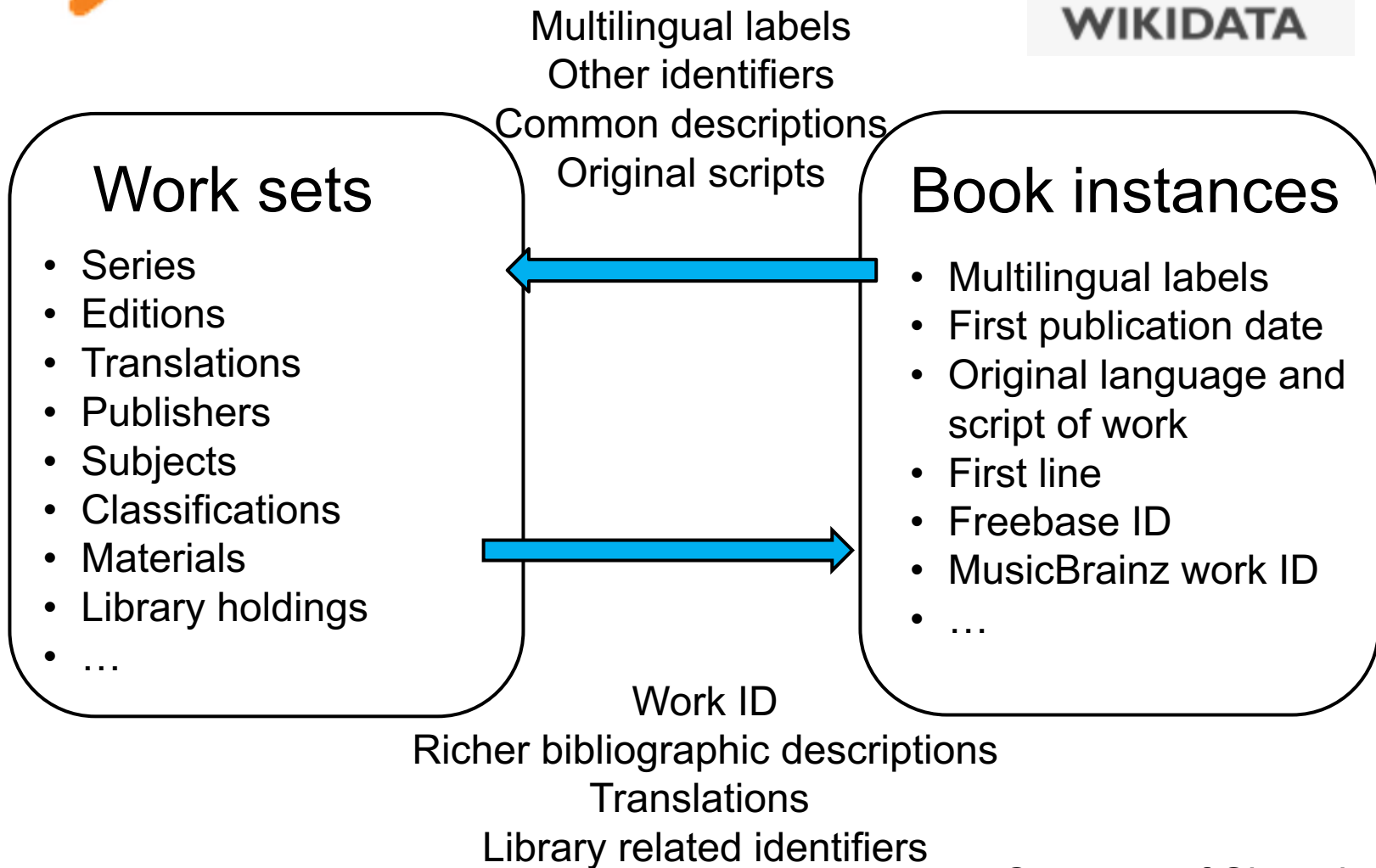
schema:creator <<http://viaf.org/viaf/102266649>> ; # "Gao, Xingjian"

schema:translator <<http://viaf.org/viaf/81663420>> ; # "Lee, Mabel"

schema:inLanguage "en";

schema:name "Soul Mountain"@en ;

schema:translationOfWork <<http://worldcat.org/entity/work/id/1215997>>



Courtesy of Shenghui Wang

Meeting the challenges

- WorldCat has more translations than any other resource.
- We can tag data elements from different languages of cataloging.
- We can ingest data from other linked data sources to present information in the preferred language and script of the user and associate translations to the original work.

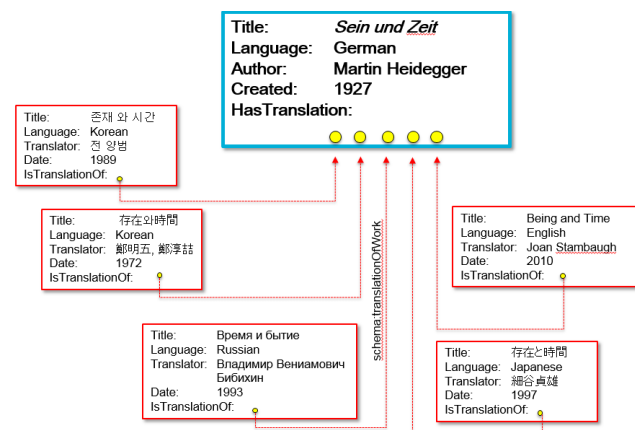


Thank you!

Karen Smith-Yoshimura

smithyok@oclc.org

@KarenS_Y



Together we make breakthroughs possible.