The Missing Link – A Vocabulary Mapping Effort in Economics

Andreas Oskar Kempf, Joachim Neubert, Manfred Faden
ZBW – Leibniz Information Centre for Economics –
German National Library of Economics
14th European Networked Knowledge Organization Systems (NKOS) Workshop
September 18th 2015

ZBW Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

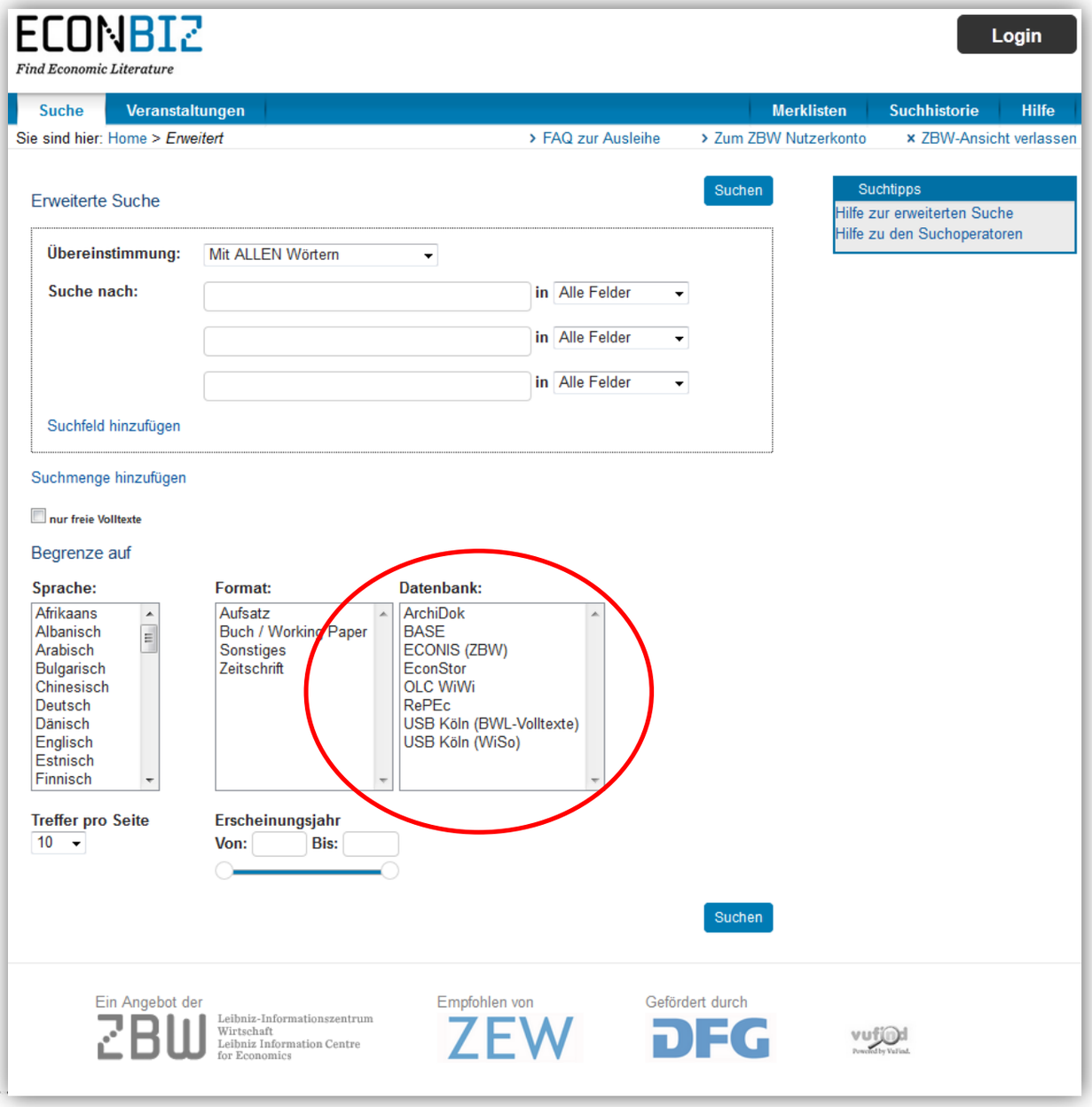
The ZBW is a member of the Leibniz Association.

Introduction

Why do we do vocabulary mappings in general?

- Mappings enable an integrated search in a distributed search environment.

- Mappings translate search terms into the vocabulary of the target KOS.



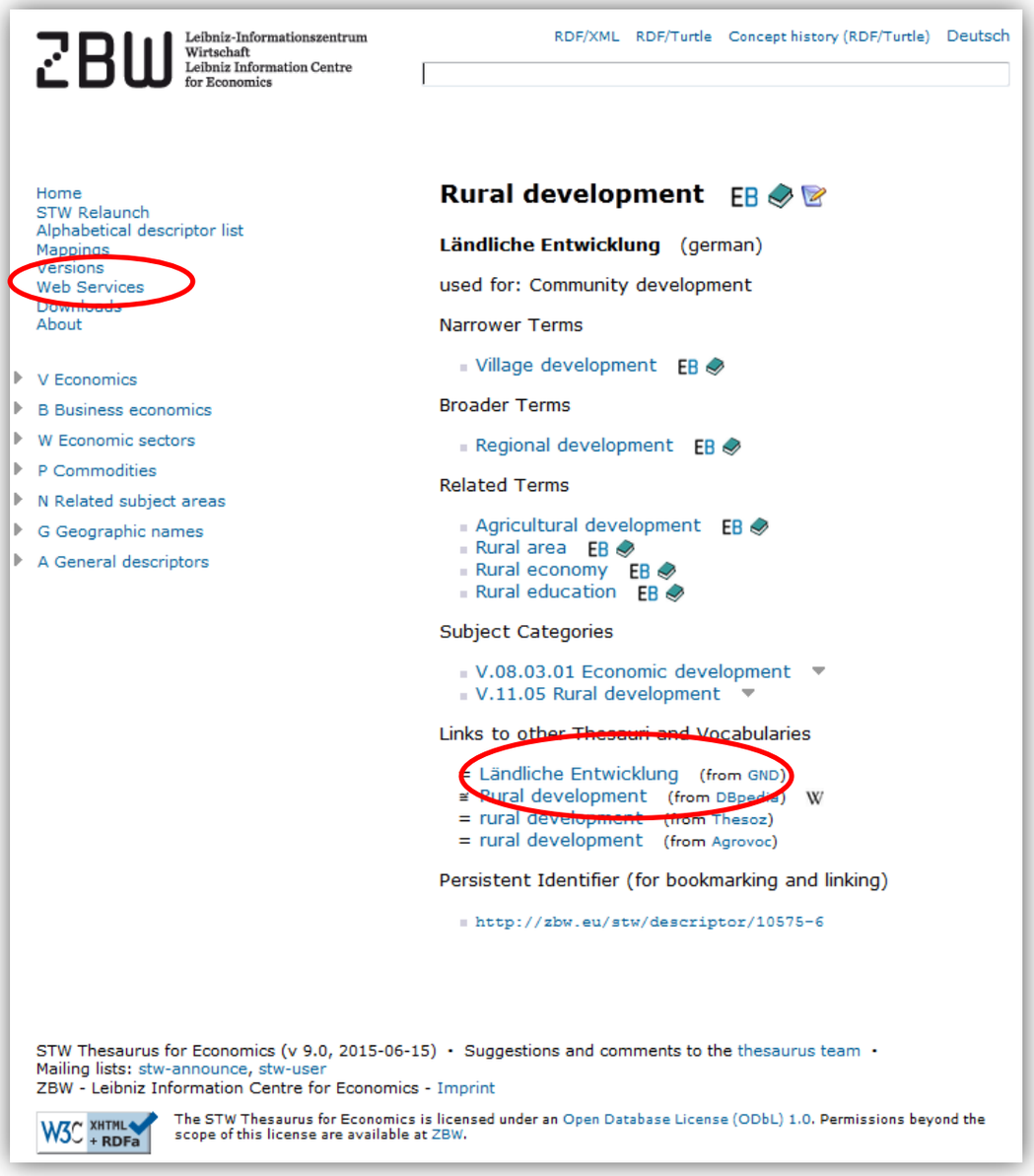
CC BY NC SA

Introduction

For what reason did we at ZBW do mappings in the past?

- … to offer an integrated search space for our search portal for economics EconBiz,
  - ➢ e. g. Integrated Authority File

- … to link the STW with other vocabularies for the development of semantic web applications.



CC BY NC SA

Introduction

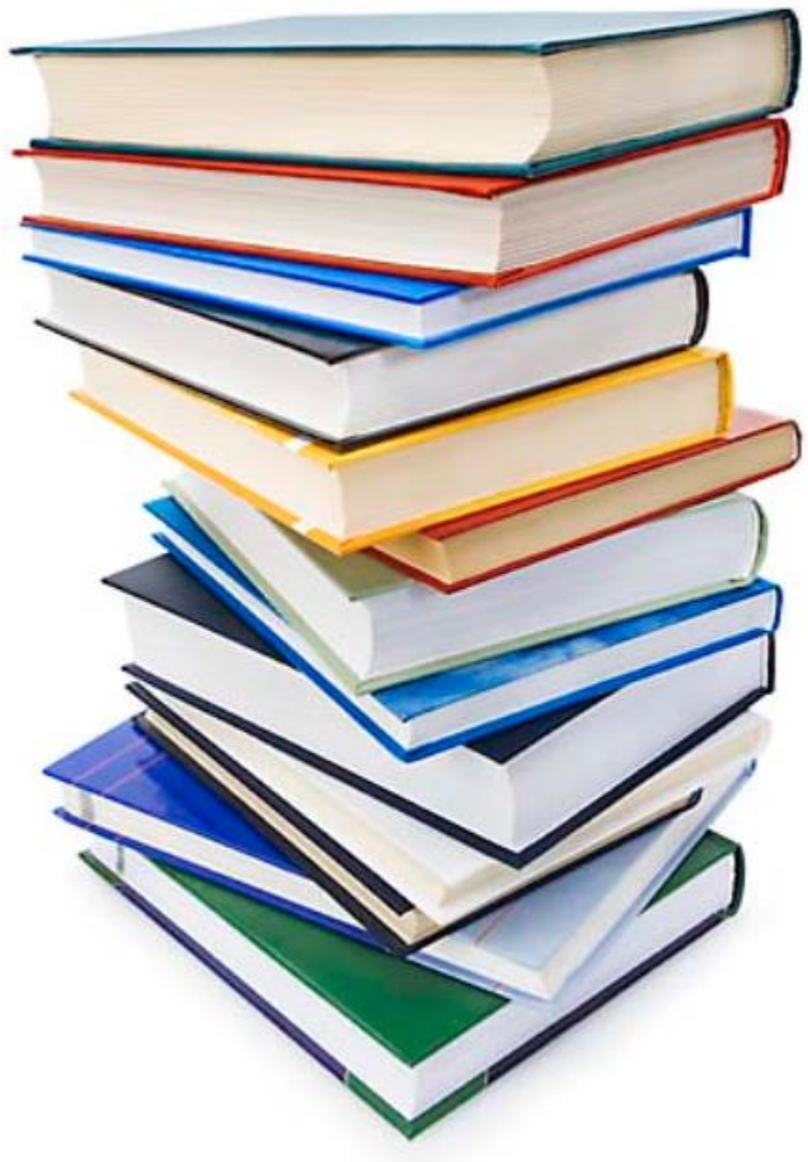
What is new about the current mapping effort?

Context:

– … increasing numbers of publications and decreasing personnel resources.

➢ … complementary approaches to conventional subject indexing are needed,

➢ i. a. reuse of user-generated content.



CC BY NC SA

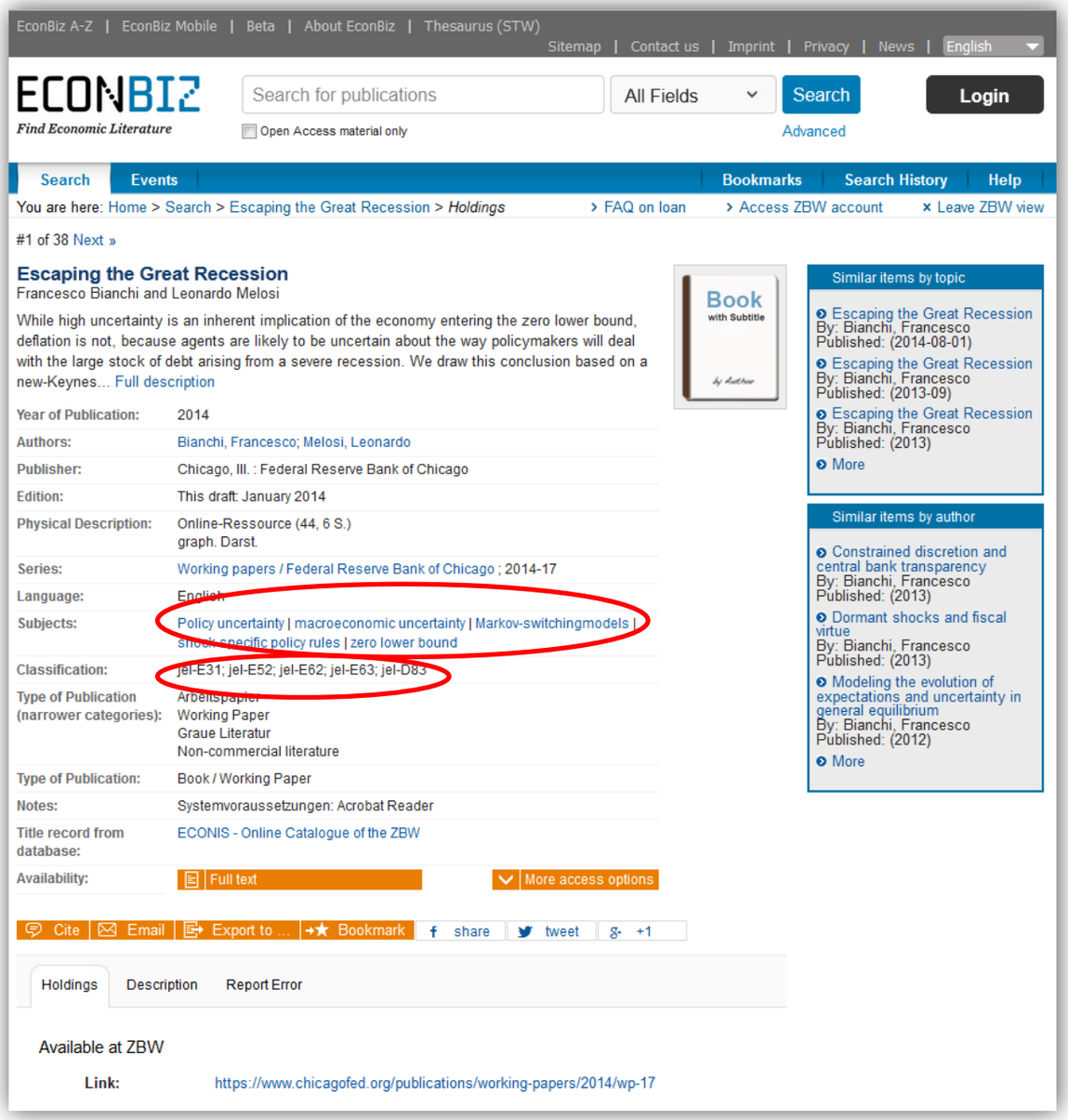
Introduction

Current reuse scenario of user-generated content at ZBW:

Regarding working paper series:

– Verbal subject indexing: inclusion of **author keywords** into bibliographic records if available.

– Classificatory subject indexing: inclusion of **JEL classes** into bibliographic records if available.



CC BY NC SA

Introduction

Future reuse scenario for a JEL – STW (systematic display) mapping effort:

– … building on the fact that economists are usually quite familiar with the JEL classification codes.

– … animate economists to use STW subject headings in order to provide a more fine-grained content description with a standardized vocabulary.

STW –
Thesaurus for Economics
(subject category system)

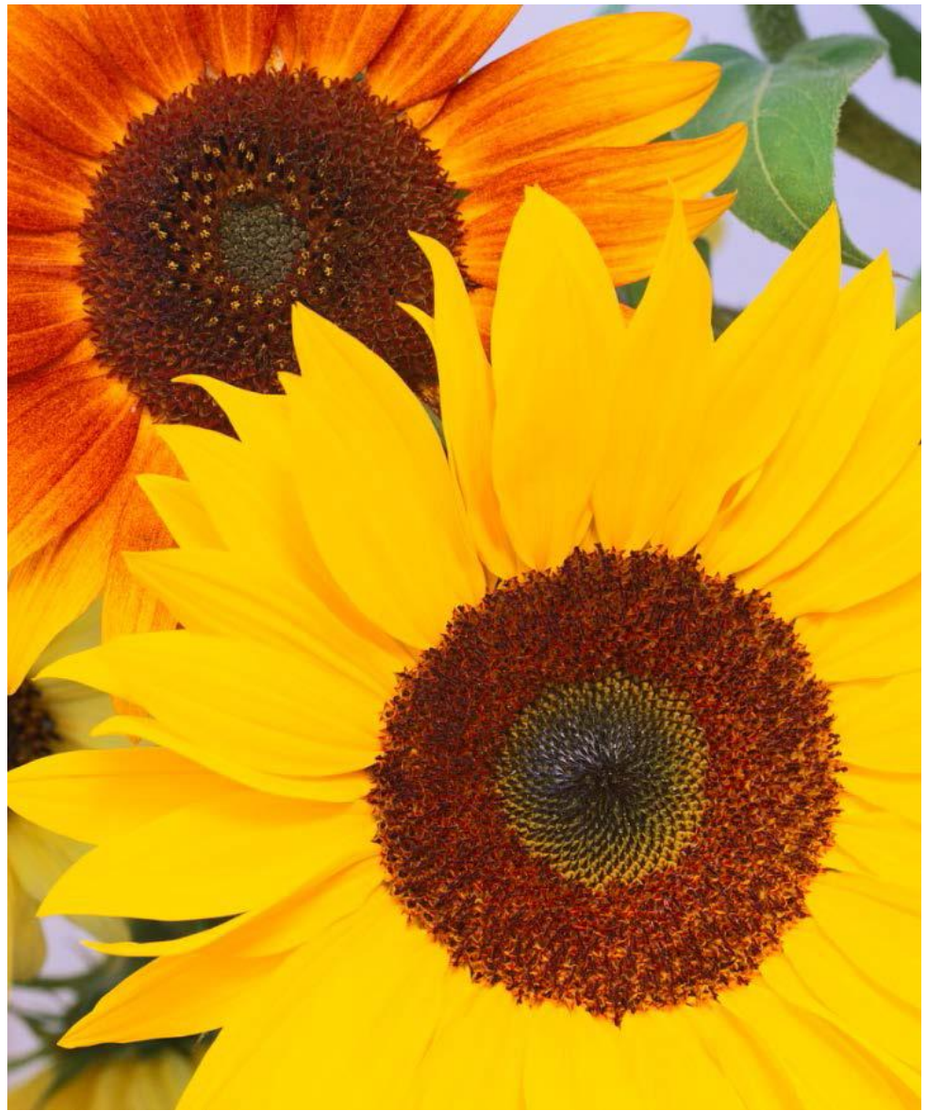
JEL –
Journal of Economic Literature Classification System

CC BY NC SA

Research question

Regarding the use case we have in mind to what extent is a useful mapping between both KOS possible?

Dealing with this question on the one hand includes a **theoretical reflection** on the structure of both KOS. On the other hand it includes the presentation of a specific **iterative semi-automatic mapping approach**.



CC BY NC SA

Outline

- Introduction
- Knowledge organization systems in economics
- Definition of interoperability and structural models for mapping
- Mapping process
- Empirical examples
- Results
- Conclusion and future outlook

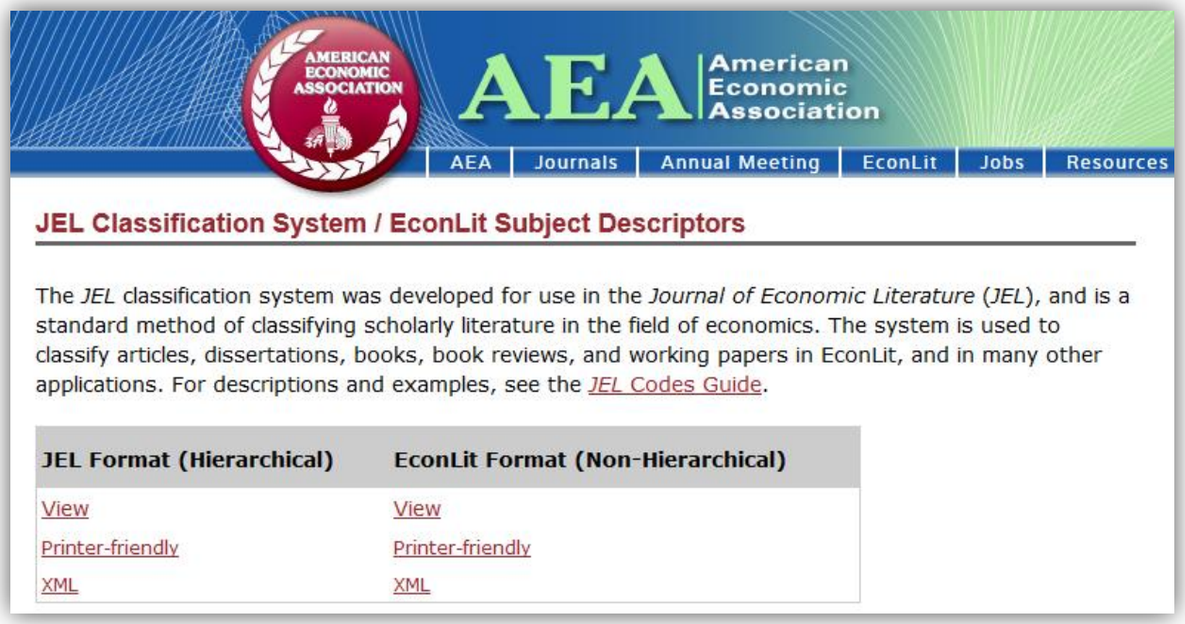
CC BY NC SA

JEL Classification System

Institutional background:

– It is published by the American Economic Association (AEA), which publishes the American Economic Review and maintains the searchable database EconLit.

– The AEA Executive Committee regularly reports on changes of JEL classes in the American Economic Review.



CC BY NC SA

JEL Classification

Scope:

– It represents an **Anglo-American understanding of economics** mainly focusing on (national) economics [ger.: VWL].

Structural characteristics:

– It is a **precombined classification system** with a monohierarchical structure and polydimensional ordering principles.

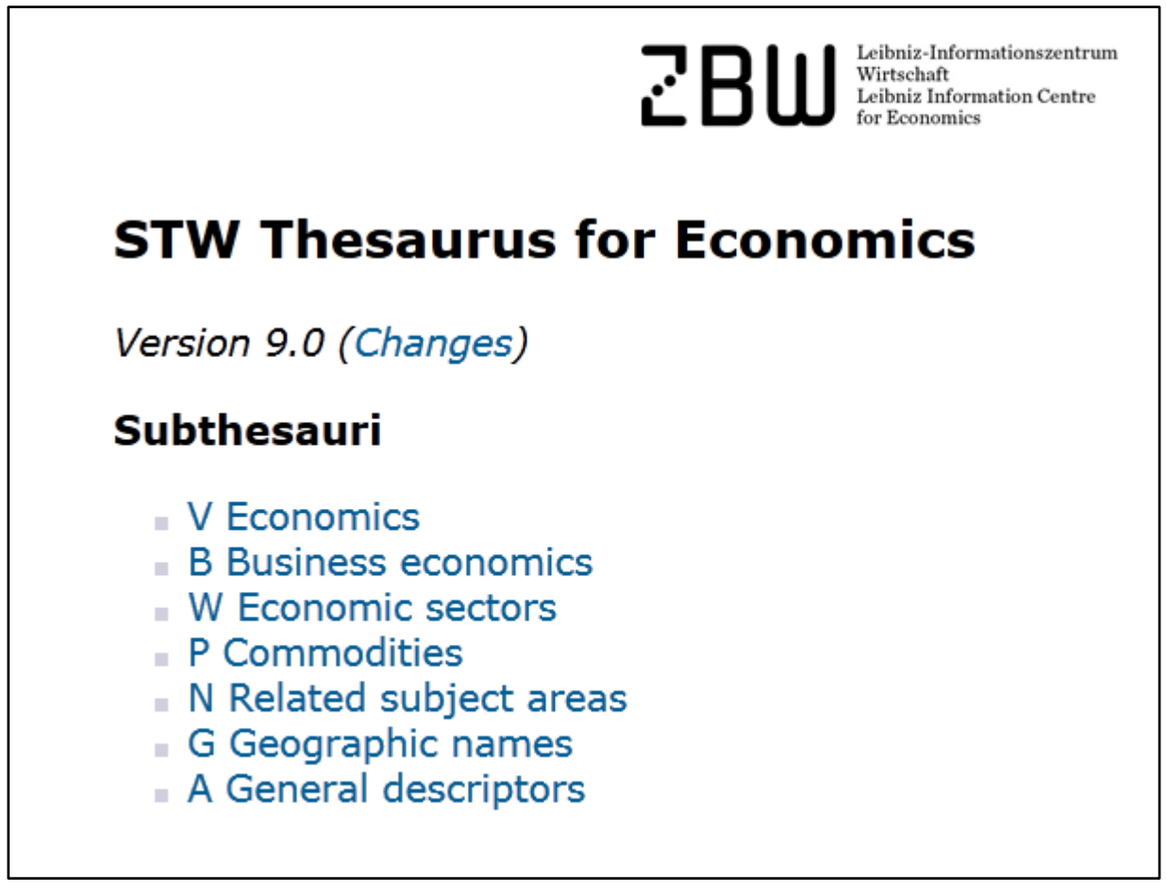
STW Thesaurus for Economics

Institutional background:

– Developed in cooperation thanks to a project funded by the Ministry for Economy in the 1990s.

Scope:

– It covers all economics-related subject areas and, on a broader level, the most important related subjects (e.g. social sciences).



CC BY NC SA

STW Thesaurus for Economics

Structural characteristics:

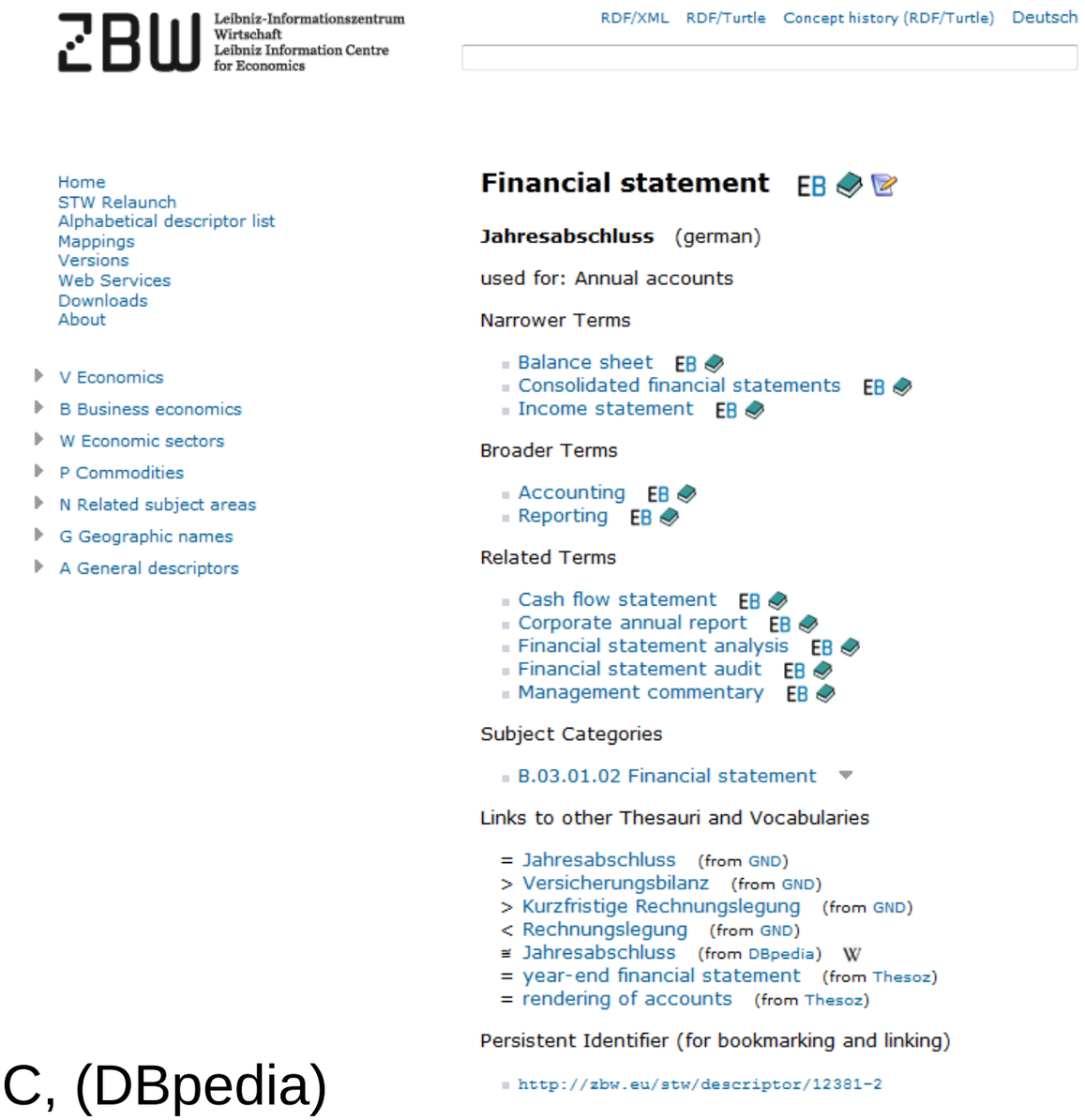
– STW is a **polyhierarchica**l bilingual thesaurus.

Types of relations:

– **equivalent relations**, including synonyms and quasi-synonyms (UF),
– **hierarchical relations**, including broader (BT) and narrower terms (NT)
– **associate relations**, including related terms (RT)

Links to other vocabularies:

– Mappings to GND, TheSoz, AGROVOC, (DBpedia)



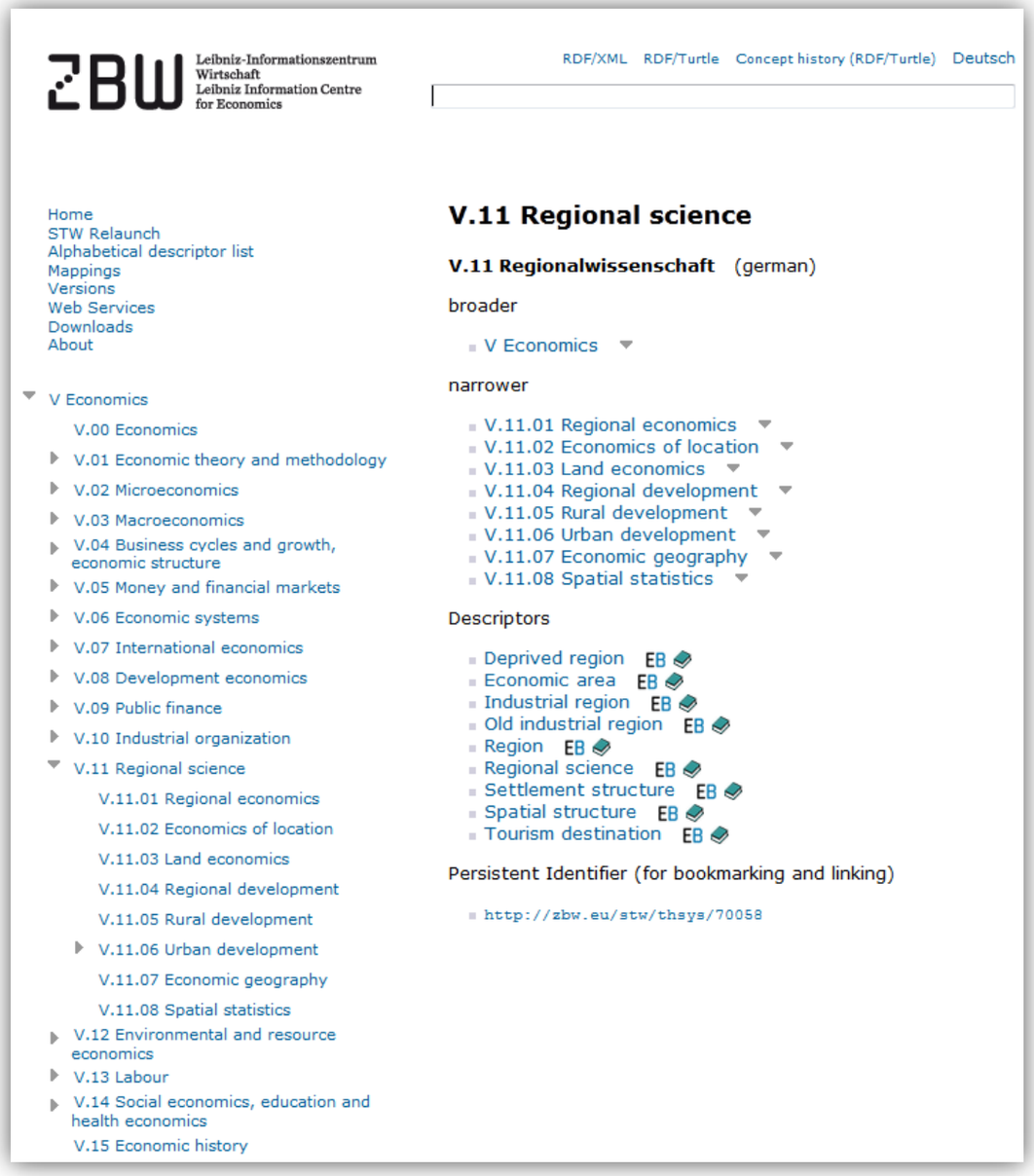
CC BY NC SA

STW subject categories

Structural characteristics:

– The STW subject categories (in total 497) constitute a monohierarchical structure with polydimensional – for subthesaurus V + B – consistently subject-specific ordering principles for vertical and horizontal subdivision.

	Subthesaurus V	Subthesaurus B
1st level	1	1
2nd level	15	10
3rd level	62	38
4th level	43	21
Total	121	70



CC BY NC SA

JEL Classification vs. STW Subject Categories

	JEL Classification	STW Subject Categories
Definition	Class (ISO 25964-2: 3.10, „concept (3.17) or group of similar or related concepts (3.17) (sic!) used as a division or subdivision in a classification scheme (3.12).“)	Concept group (ISO 25964-2: 3.18, „group of concepts selected by some specified criterion…“)
Scope	Domain-specific (USA, UK)	Domain-specific (GER > international) Here: Restriction to the subthesauri: V: Economics and B: Business economics.
Purpose	All-embracing systematization of a discipline.	Systematization of the thesaurus vocabulary
Structural characteristics	- Precombined classification - Monohierarchical - Polydimensional ordering principles	- Monohierarchical - Polydimensional ordering principles

– Because of the **structural heterogeneity** between the two vocabularies mapping relations for the most part are not expected to be relations of full equivalence. Rather they are presumed to oftentimes consist of inexact equivalent relations.

CC BY NC SA

Outline

- Introduction
- Knowledge organization systems in economics
- Definition of interoperability and structural models for mapping
- Mapping process
- Empirical examples
- Results
- Conclusion and future outlook

CC BY NC SA

Definition of interoperability

ISO 25964: Thesauri and interoperability with other vocabularies

Developed by an international working group (2008-2013)

- Part 1: Thesauri for information retrieval (published 2011)
  Contains guidelines for establishing monolingual and multilingual thesauri.
- Part 2: Interoperability with other vocabularies (published 2013)
  Deals with mappings between thesauri and other types of vocabularies for information retrieval.

CC BY NC SA

Definition of interoperability

ISO 25964-2:2013(E)

3.38
interoperability
ability of two or more systems or components to exchange information and to use the information that has been exchanged.

NOTE Vocabularies can support interoperability by including mappings to other vocabularies, by presenting data in standard formats and by using systems that support common computer protocols.

3.40
mapping, gerund (verbal noun)
process of establishing relationships between the concepts (3.17) in one vocabulary and those of another

3.41
mapping, noun
(product of mapping process) relationships between a concept (3.17) in one vocabulary and one or more concepts (3.17) in another

CC BY NC SA

Two different types of vocabularies

Structural unity:
The mapped vocabularies have the same structure.
The equivalence of the concepts of such vocabularies is expressed by their identical structural position in the vocabulary. All the relationships of the concepts correspond to each other (e.g. multilingual thesauri of public institutions)

Structural disunity:
The mapped vocabularies do not have the same structure.
Equivalence of concepts has nothing to do with their position in the vocabularies. The mapping process produces either exact equivalence pairs or inexact equivalent pairs.

Different types of equivalences:
(Real) exact equivalence: =EQ
Inexact equivalence: ~EQ (e.g. voc.have emerged from different cultural communities)
Partial equivalence: The concept is broader: BM („Broader Mapping“)
The concept is narrower: NM („Narrower Mapping“)
The concepts are somehow related: RM („Related Mapping“).

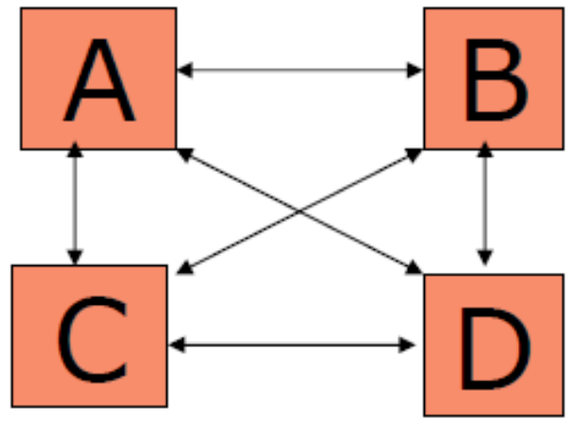
CC BY NC SA

Structural models for mapping

ISO 25964-2:2013(E)

Three different structural models for mapping across vocabularies
Model 1: Structural unity (6.2)
„All the participating vocabularies share exactly the same structure of hierarchical and associate relationships between concepts…“
(e.g. multilingual thesauri)

Model 2: Direct-linked (6.3)
The direct-linked model addresses linkages betweent two or more vocabularies that do not share the same structure. As well as differing in scope, language and structure, the vocabularies may include other types of vocabulary (classification scheme, name authority list, etc.) .



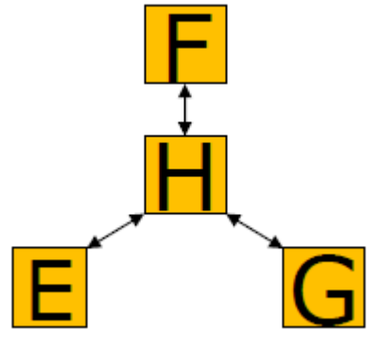
CC BY NC SA

Structural models for mapping

Model 3: Hub structure (6.4)

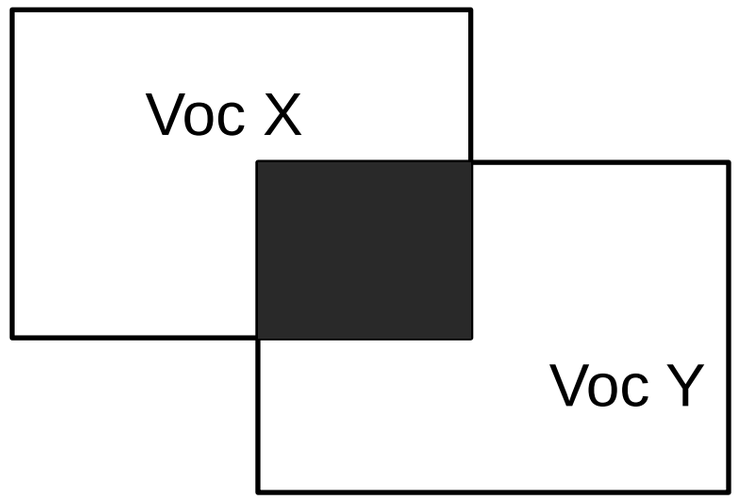
ISO 25964-2:2013(E)

One vocabulary is designated as „hub“, or conprehensive structure to which each of the other vocabularies is mapped as „satellite“. The concepts of the different vocabularies are only mapped to the concepts of the one vocabulary which has the role of a hub. This model is appropriate if there is one vocabulary with a dominating position.



Model 4: Selective Mapping (6.5)

In cases where there is only small overlap expected, it could be unnecessary to map the vocabularies comprehensively.



Selected mapping in area of overlap.

- In real applications combinations of these types often occur and the boundaries might be blurred (see ibd. p.20):

CC BY NC SA

Outline

- Introduction
- Knowledge organization systems in economics
- Definition of interoperability and structural models for mapping
- Mapping process
- Empirical examples
- Results
- Conclusion and future outlook

CC BY NC SA

Mapping process

Previous work:

– Outdated mapping JEL > STW (descriptor level)
KoMoHe project context (2004-2007)
Mapping relations:
- equivalent relations (=)
- broader/narrower relations (>/<)
- associate relations (^)
- compound mappings (+)
- including a relevance rating (high, medium, low)

22	8	A10	>o	g	Zeitkonsistenz
22	8	A10	>o	g	Ökonomisches Prinzip
22	8	A10	>o	m	Wirtschaftsmodell
22	8	A10	>o	m	Wirtschaftswissenschaft
22	8	A11	<o	g	Wirtschaftsforschung
22	8	A11	>o	g	Wirtschaftswissenschaftler
22	8	A12	<	g	Wirtschaftswissenschaft
22	8	A13	^	m	Wirtschaftsethik
22	8	A14	=	h	Wirtschaftssoziologie
22	8	A19	<	g	Wirtschaftswissenschaft
22	8	A1	=	m	Wirtschaftstheorie
22	8	A20	<o	m	Pädagogik
22	8	A20	o	h	Wirtschaftspädagogik
22	8	A20	^o	m	Arbeitswissenschaft
22	8	A20	^o	m	Berufsbildung
22	8	A20	^o	m	Wirtschaftsstudium
22	8	A20	^o	m	Wirtschaftswissenschaft
22	8	A21	<	g	Wirtschaftsstudium
22	8	A22	<	g	Wirtschaftsstudium
22	8	A23	<	g	Wirtschaftsstudium
22	8	A29	<	g	Wirtschaftsstudium

– Outdated concordance
STW (classification system) > JEL
  – On the third level of JEL classes,
  – No specified mapping relations.

STW classification system	STW subject categories	JEL classes
V02-000 Microeconomics	V.02 V.02.05	B21 D00 …
V02-010 Houlsehold economics	V.02.01	D10 D11 …

Mapping process

What is new?

- …mapping on the level of the STW subject category system (instead of the STW classification level),
- …referring to a web-based interactive mapping platform,
- …using the SKOS vocabulary to build and to manage the mapping
  Note: this goes along with the assumption that both vocabularies could be mapped bilaterally,
- …referring to an iterative mapping process of a first and a second iteration and an approach of vocabulary enrichment of the mapping with additional keywords (JEL) and subject headings from STW together with equivalent concept relations from past vocabulary mappings .

cc BY NC SA

Outline

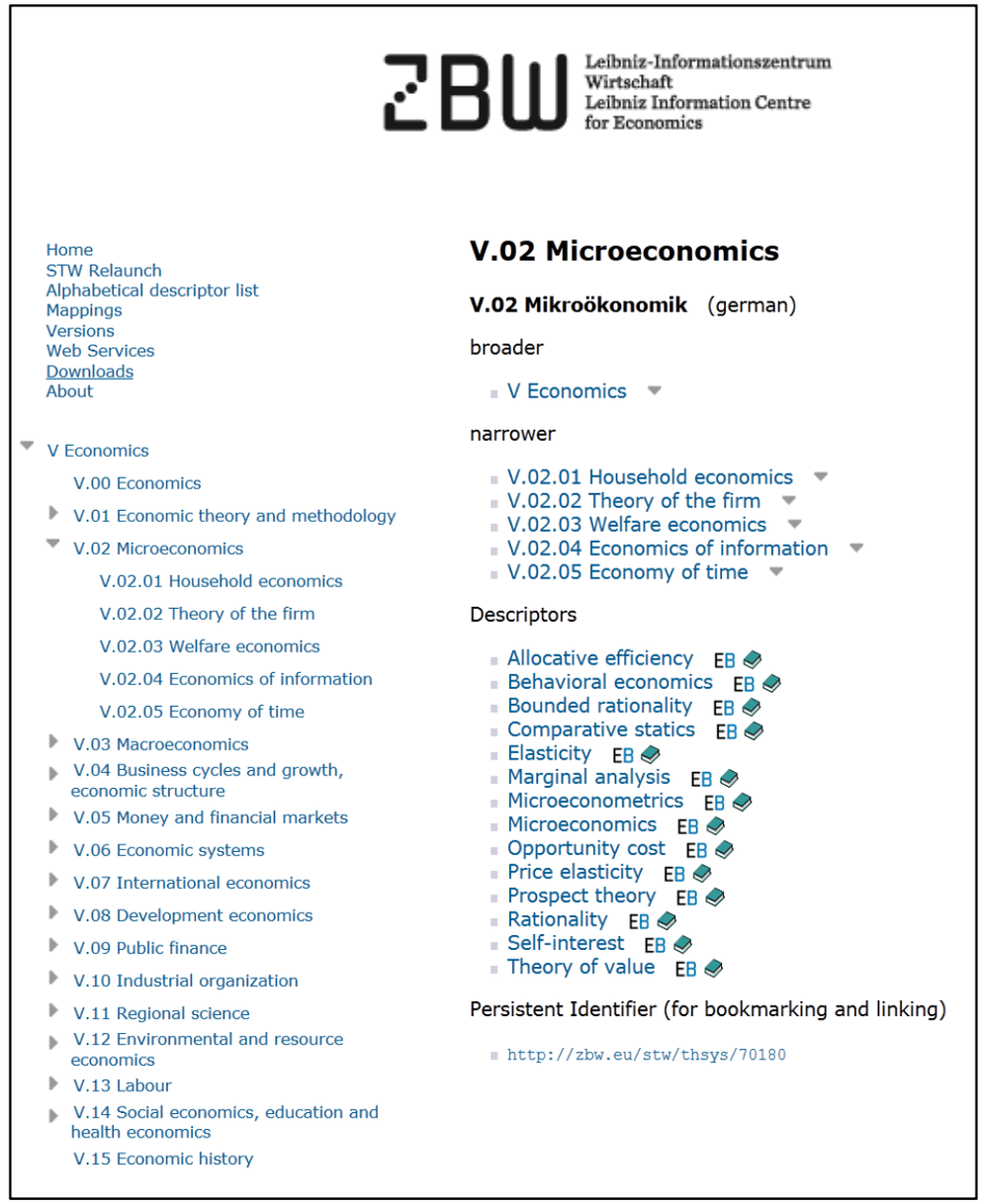
- Introduction
- Knowledge organization systems in economics
- Definition of interoperability and structural models for mapping
- Mapping process
- Empirical examples
- Results
- Conclusion and future outlook

CC BY NC SA

Empirical examples

Selection of STW subject categories:

- **STW subthesaurus V – Economics:**
  - V.02 – Microeconomics (1 subject category)
    V.02.01 – V.02.05 (5 s.c.)
  - V.15 – Economic history (1 s.c.)
    V.15 – (-)
- **STW subthesaurus B – Business economics**
  - B.07 – Marketing (1 s.c.)
    B.07.01 – B.07.06 (6 s.c.)
  - B.09 – Business information systems (1 s.c.)
    B.09.01 – B.09.03 (3 s.c.)



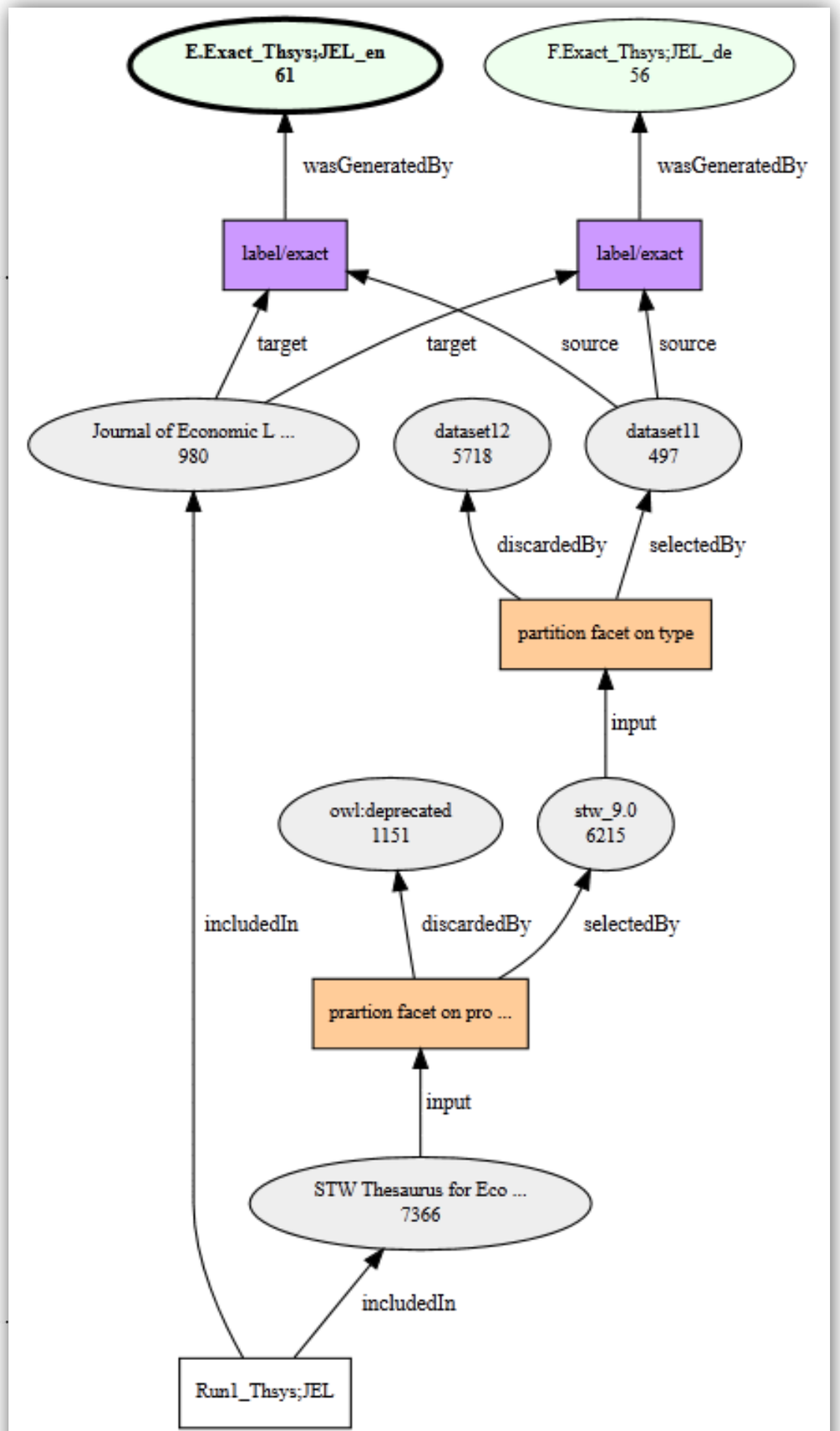
CC BY NC SA

Empirical examples

Mapping procedure:

- Use of the interactive alignment server **AMALGAME** (AMsterdam ALignment GenerAtion MEtatool)
  - Upload of the STW (v 9.0) in SKOS http://zbw.eu/stw/versions/latest/download/about.de.html
  - Upload of the JEL classification in SKOS http://zbw.eu/beta/external_identifiers/jel/about.en.html

➢ Exact language dependent string match of STW subject categories and JEL classes.



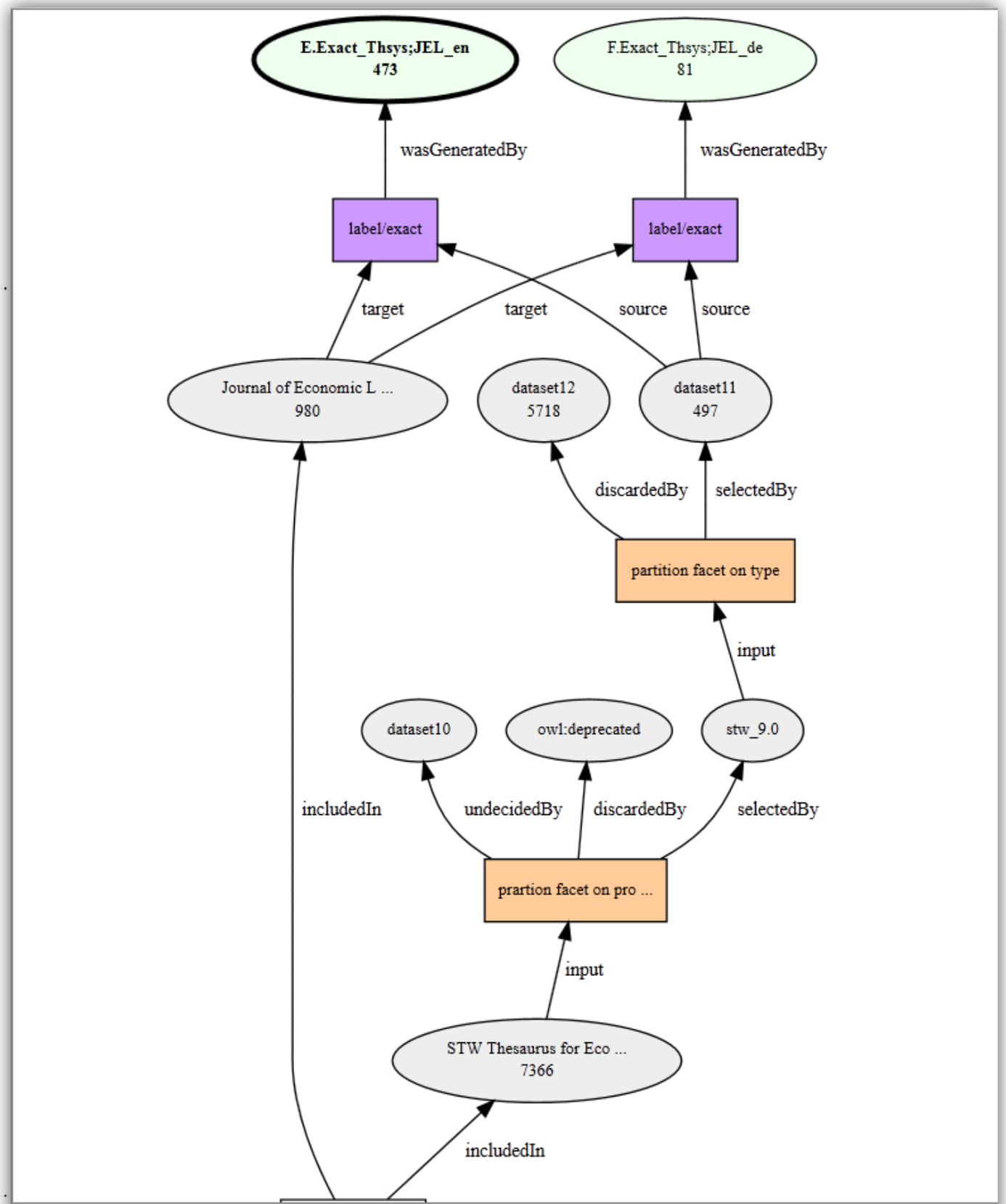
AMALGAME Mapping graph of the first run

Empirical examples

Second run:
(Same selection of subject categories.)

Enrichment of STW subject categories and JEL classes:

- STW subject categories enriched by:
  - STW descriptors + synonyms
  - Mapped (exactMatch) concepts from other vocabularies – descriptors + synonyms (GND, TheSoz, DBpedia, AGROVOC)
- JEL classes enriched by:
  - JEL keywords scraped from JEL guide
- German + English (if available)



AMALGAME mapping graph 2nd run.

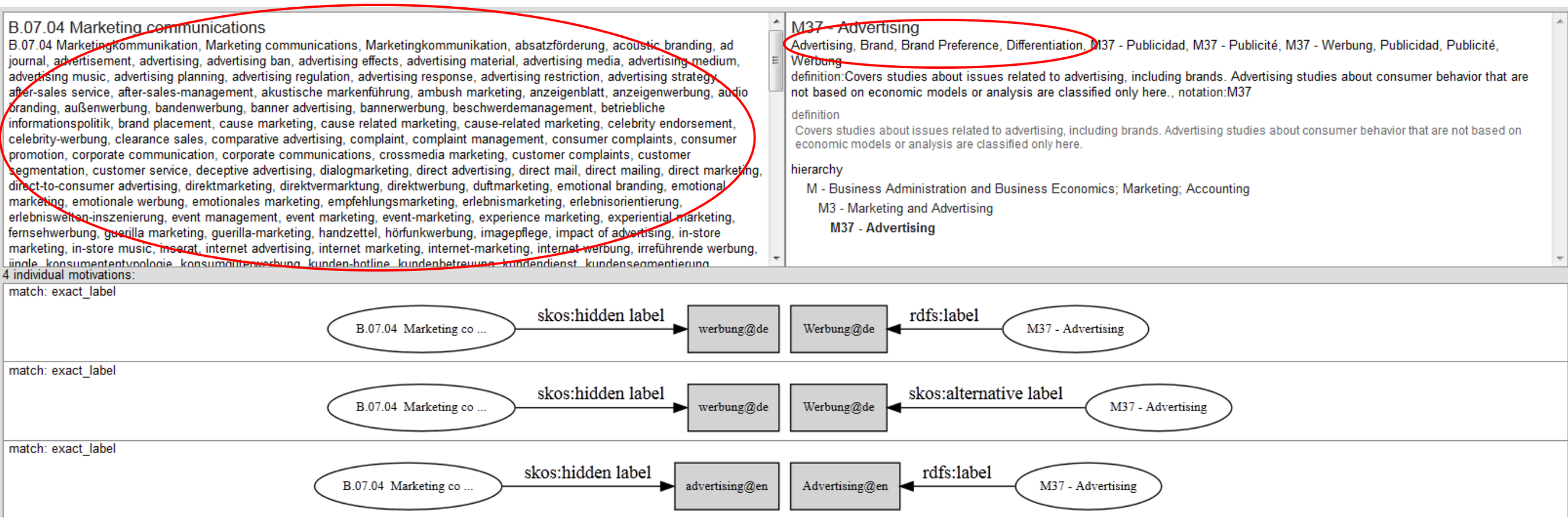
Empirical examples

STW subject categories enriched by:

- STW descriptors + synonyms
- Mapped (exactMatch) concepts from other vocabularies – descriptors + synonyms (GND, TheSoz, DBpedia, AGROVOC)

JEL classes enriched by:

- JEL keywords scraped from JEL guide https://www.aeaweb.org/jel/guide/jel.php



CC BY NC SA

Outline

- Introduction
- Knowledge organization systems in economics
- Definition of interoperability and structural models for mapping
- Mapping process
- Empirical examples
- Results
- Conclusion and future outlook

CC BY NC SA

		Intell. eval.	1st run (=)	(+)	Certain overlap ~	Wrong (-)	2nd run (=)	(+)	Certain overlap ~	Wrong (-)
V.02	Microeconomics	5					1 (13)	1	3	7
V.02.01	Household economics	5					3 (21)		6	11
V.02.02	Theory of the firm	4					2 (10)	1		7
V.02.03	Welfare economics	6					3 (10)		2	5
V.02.04	Economics of information	4					2 (4)			2
V.02.05	Economy of time	7					1 (9)			8
V.15	Economic history	74					9 (15)		4	2
B.07	Marketing	3	1 (5)		4		2 (20)	1	1	
B.07.01	Marketing management	6					(-)			
B.07.02	Product Management	2					1 (15)		5	9
B.07.03	Pricing strategy	1					- (11)		2	9
B.07.04	Marketing communications	2					1 (2)			1
B.07.05	Distribution	1					- (12)		2	9
B.07.06	Market research	3					- (2)		1	1
B.09	Business information systems	1					- (5)			5
B.09.01	Information system components	1					- (4)		1	3
B.09.02	IS development and management	1					1 (6)			5
B.09.03	Corporate information systems	1					- (4)			4

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

CC BY NC SA

Outline

- Introduction
- Knowledge organization systems in economics
- Definition of interoperability and structural models for mapping
- Mapping process
- Empirical examples
- Results
- Conclusion and future outlook

CC BY NC SA

Conclusion and future outlook

- String match can only generate mapping candidates it is blind to structural differences.
- The approach of vocabulary enrichment including JEL keywords, STW descriptors, synonyms, translations and equivalent terms and their synonyms from past vocabulary mappings led to a substantial increase of mapping candidates also included in the intellectual mapping.
  Note: A new use case for already established vocabulary alginments.
- Vocabulary enrichment has also led to new mapping candidates worth revisioning the current intellecutal mapping.

- Optional mapping procedure in the future:
  The STW as access vocabulary to JEL classes.

CC BY NC SA

Thank you for your attention!

Contact:

Dr. Andreas Oskar Kempf
a.kempf@zbw.eu
Joachim Neubert
j.neubert@zbw.eu
Manfred Faden
m.faden@zbw.eu

http://zbw.eu/stw
https://www.aeaweb.org/econlit/jelCodes.php
Inofficial multilingual/LOD version:
http://zbw.eu/beta/external_identifiers/jel)
http://semanticweb.cs.vu.nl/amalgame/

CC BY NC SA

References

BERTRAM, J. (2005) Einführung in die inhaltliche Erschließung. Ergon Verlag.

GÖDERT, W. (2010) Semantische Wissensrepräsentation und Interoperabilität. In: Information – Wissenschaft & Praxis, p. 5-28.

HUBRICH, J. (2008) CRISSCROSS: SWD-DDC-Mapping. In: Mitteilungen der VÖB 61, Nr. 3, p.50-58.

ISO 25964 Thesauri and interoperability with othr vocabularies. Part 1 (2011) +2 (2013).

MAYR, P./PETRAS, V. (2007) Building a terminology network for search: the KoMoHe project. Proceedings of the International Conference on Dublin Core and Metadata Applications.

SCHEVEN, E. (2013) ISO 25964 and GND, LIS`workshop 2013, Luxemburg http://digbib.ubka.uni-karlsruhe.de/volltexte/documents/2683976

CC BY NC SA