



Using BMEcat Catalogs as a Lever for Product Master Data on the Semantic Web

10th Extended Semantic Web Conference (ESWC 2013)
May 26-30, 2013, Montpellier, France

Alex Stolz, Bene Rodriguez-Castro, and Martin Hepp
{alex.stolz,bene.rodriguez,martin.hepp}@unibw.de



Problem

- **Lack of structured product master data on the Web (of Data)**
- Web Shop owners have difficulties in maintaining their product data
 - Combine product catalogs from different manufacturers
 - Enter products and feature data manually
 - Redundant definitions of same product data on the Web; often incomplete, inconsistent and outdated
- Meanwhile manufacturers maintain comprehensive catalogs of their items in PIM systems

Problem - Example

Manufacturer Product Features		Retailer Product Features		Coverage ¹
Samsung LED TV ES6300	89	15	amazon.de	28.09%
		39	notebooksbilliger.de	
		22	conrad-electronics.de	
		24	voelkner.de	
Siemens Kettle TW86103	25	10	amazon.de	23.64%
		22	redcoon.de	
		4	quickshopping.de	
		13	elektro-artikel-shop.de	
Suunto M5 Running Pack	33	12	amazon.de	49%
		3	sportscheck.com	
		1	otto.de	
		15	klepsoo.com	
		8	tictactime.de	

Our Idea

- Make product catalogs of manufacturers available as structured data on the Semantic Web
- Web shop owners can take advantage of high-quality product specifications provided by manufacturers
- Products and product datasheets can be linked based on *strong identifiers*, e.g. EAN, GTIN, MPN, brand name, ..., typically provided by both manufacturers and retailers

Our Idea - Overview

Manufacturer Web Site: Datasheet

<http://www.acme.com>

Product
page with
details

High-quality picture



weight: 250g
color: blue

EAN: 1234567890123
GTIN14: 12345678901234
MPN: ACME123
brand: ACME

Many Shop Sites (with incomplete product features)

<http://www.shop1.com> <http://www.shop2.de> <http://www.shop3.uk>

Shop 1:
Offer page

price: \$ 99.99

EAN: 1234567890123

Shop 2:
Offer page

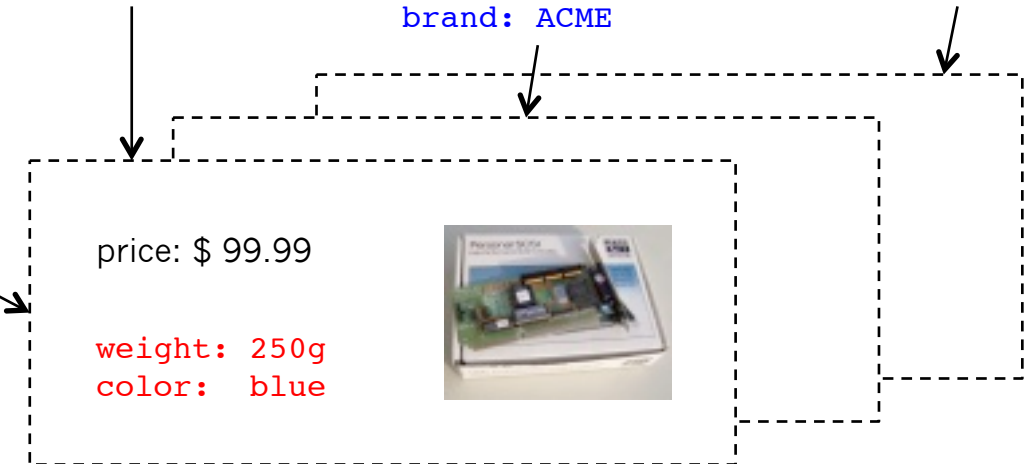
price: \$ 102.10

MPN: ACME123
brand: ACME

Shop 3:
Offer page

price: \$ 96.00

GTIN14: 12345678901234



Search Engine or Browser Plug-in

BMEcat Standard

- XML-based product catalog exchange format
- Well-known and widely used by German industry
- Popular: BMEcat 1.2 and 2005
- Supported by most PIM systems¹, as commonly used by manufacturers and large retailers to maintain their product data



¹ 82% of PIM software providers support BMEcat according to a market survey conducted in Germany

BMEcat 2005 Structure

HEADER

CATALOG
AGREEMENT
SUPPLIER
BUYER

T_NEW_CATALOG

CATALOG_GROUP_SYSTEM
CATALOG_STRUCTURE
PRODUCT
PRODUCT_DETAILS
PRODUCT_FEATURES
PRODUCT_ORDER_DETAILS
PRODUCT_PRICE_DETAILS
PRODUCT_TO_CATALOGGROUP_MAP

<PRODUCT>

<SUPPLIER_PID>SK26E800EU</SUPPLIER_PID>

<PRODUCT_DETAILS>

<DESCRIPTION_SHORT>SK26E800EU
speedMatic Compact-Geschirrspüler...</DESCRIPTION_SHORT>

<DESCRIPTION_LONG>...</DESCRIPTION_LONG>

<EAN>4242003474686</EAN>

<MANUFACTURER_PID>SK26E800EU</MANUFACTURER_PID>

<MANUFACTURER_NAME>Siemens</MANUFACTURER_NAME>

<MANUFACTURER_TYPE_DESCR>SK26E800EU</MANUFACTURER_TYPE_DESCR>

<SPECIAL_TREATMENT_CLASS type="NOT_RELEVANT">NONE</SPECIAL_TREATMENT_CLASS>

<KEYWORD>Geschirrspülmaschine</KEYWORD>

</PRODUCT_DETAILS>

<PRODUCT_FEATURES>

<REFERENCE_FEATURE_SYSTEM_NAME>ECLASS-6.1</REFERENCE_FEATURE_SYSTEM_NAME>

<REFERENCE_FEATURE_GROUP_ID>29150301</REFERENCE_FEATURE_GROUP_ID>

<FEATURE>

<FNAME>Marke</FNAME>

<FVALUE>Siemens</FVALUE>

</FEATURE>

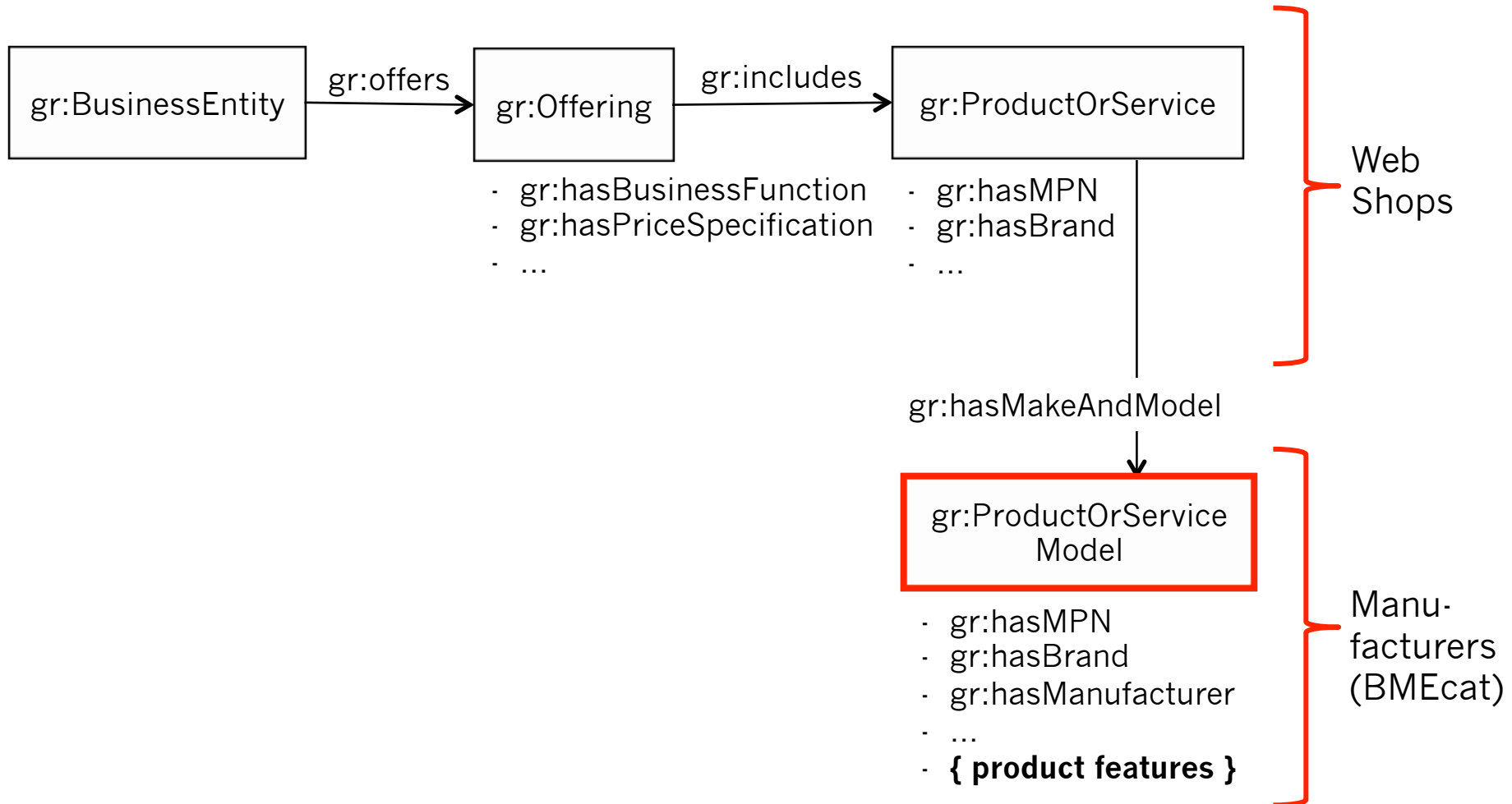
GoodRelations



- Web vocabulary for e-commerce
 - Industry-neutral
 - Valid across different stages of the value chain
 - Syntax-neutral

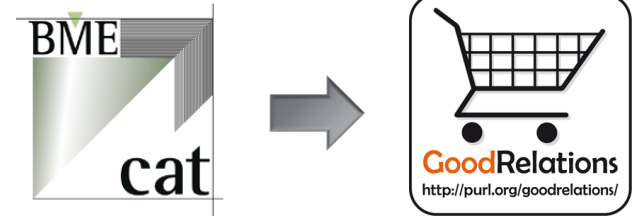
- Widely accepted and used
 - Schema.org integration (2012)
 - Google Rich Snippets (2010), Yahoo SearchMonkey (2008)
 - BestBuy, Sears.com, Arzneimittel.de
 - Rakuten.de, countless small Web shops based on standard shop software
 - Volkswagen UK, Renault UK
 - Extensions: Vehicle Sales Ontology (VSO), Tickets Ontology (TIO), Product Types Ontology (PTO)

GoodRelations Simplified Model



Conversion from BMEcat to GoodRelations

- Portable command-line application in Python
- Capabilities of converting
 - common product details (e.g. strong identifiers)
 - product features
 - product catalog group system
- Additional features
 - Multi-language support
 - Links to external feature and product classification standards
 - Distinction between datatype and object properties
 - Heuristics (e.g. generate quantitative value ranges out of “1.0-2.4”)



Mapping of Product Descriptions

BMEcat	GoodRelations	
PRODUCT	gr:Offering, gr:Individual/gr:SomeItems, gr:ProductOrServiceModel	
SUPPLIER_PID type={ <i>ean</i> , <i>gtin</i> }	gr:hasEAN_UCC-13, gr:hasGTIN-14	Strong ID
PRODUCT_DETAILS		
DESCRIPTION_SHORT lang={ <i>en</i> , <i>de</i> , ...}	gr:name with language <i>en</i> , <i>de</i> , ...	
DESCRIPTION_LONG lang={ <i>en</i> , <i>de</i> , ...}	gr:description with language <i>en</i> , <i>de</i> , ...	
INTERNATIONAL_PID type={ <i>ean</i> , <i>gtin</i> }	gr:hasEAN_UCC-13, gr:hasGTIN-14	Strong ID
MANUFACTURER_PID	gr:hasMPN	Strong ID
MANUFACTURER_NAME	gr:hasManufacturer → gr:BusinessEntity → gr:name	
PRODUCT_STATUS type={ <i>new</i> , <i>used</i> , ...}	gr:condition	

Basic, common product descriptions:

```
samsung:LEDTV_ES6300 a gr:ProductOrServiceModel;
  gr:name "Samsung LED TV ES6300"@en;
  gr:hasEAN_UCC-13 "1234567890123"^^xsd:string
  gr:hasMPN "ledtv_es6300"^^xsd:string .
```

Mapping of Product Features

BMEcat	GoodRelations	
PRODUCT_FEATURES		
REFERENCE_FEATURE_SYSTEM_NAME	referenced classification system identifier	Property 2
REFERENCE_FEATURE_GROUP_ID	rdf:type (class id of classification system)	Property 2
REFERENCE_FEATURE_GROUP_NAME	gr:category	
FEATURE		
FNAME	rdfs:label and property name in GR	Property 2
FDESCR	rdfs:comment	
FVALUE	gr:hasValueFloat	} Value
FUNIT	gr:hasUnitOfMeasurement	
FREF	feature id of referenced classification system, property name in GR context	Property 1

Product features that complement basic product descriptions:

```
samsung:LEDTV_ES6300 a gr:ProductOrServiceModel;
  samsung:compatible_3D "true"^^xsd:boolean;
  samsung:screen_size [ a gr:QuantitativeValueFloat;
    gr:hasValueFloat "101.6"^^xsd:float;
    gr:hasUnitOfMeasurement "CMT"^^xsd:string ] .
```

Mapping of Catalog Group Systems

BMEcat	GoodRelations
CATALOG_GROUP_SYSTEM	
CATALOG_STRUCTURE	owl:Class
GROUP_ID	class name of owl:Class
GROUP_NAME lang={ <i>en, de, ...</i> }	rdfs:label with language <i>en, de, ...</i>
GROUP_DESCRIPTION lang={ <i>en, de, ...</i> }	rdfs:comment with language <i>en, de, ...</i>
PARENT_ID	rdfs:subClassOf (class id of superclass)

Add type information to products/product models:

```
samsung:LED_TV rdfs:subClassOf samsung:TV .
```

```
samsung:LEDTV_ES6300 a gr:ProductOrServiceModel,  
    samsung:LED_TV .
```

Products are assigned categories using *PRODUCT_TO_CATALOGGROUP_MAP*.

Evaluation

Part 1:

Conversion of BMEcat catalogs of two manufacturers

1. Validation of the quality of the conversions using Semantic Web validators and the standard output of the converter

Part 2:

Collected product offer data of 2,500+ Web shops in SPARQL-capable RDF store

2. Complement the motivating example from the introduction by showing the lack of product master data in the crawl data
3. Show potential leverage of product master data on the Web

Conversion of BMEcat Catalogs

■ **Weidmüller Interface GmbH and Co. KG**

- Manufacturer of high-tech electronic components
- 32,585 product models, 47 properties on average
- Deployment available:
<http://catalog.weidmueller.com/semantic/sitemap.xml>

■ **BSH Bosch und Siemens Haushaltsgeräte GmbH**

- Supplier of white goods (e.g. vacuum cleaner, washing machine, ...)
- 1,376 product models, 29 properties on average

Validation of BMEcat to GR Conversions

Validation	BSH	Weidmüller
BMEcat2GoodRelations converter	warnings: (a) wrong values where numeric values were expected; (b) non-standard unit codes detected	warnings: (a) non-standard unit codes detected
RDF Validator	valid. warning: invalid lexical value for literal	valid
W3C RDF Validation	valid	valid
Pellet	valid. warning: malformed xsd:float detected	valid
GoodRelations Validator	step 32 failed: non-compliance of float literal with xsd:float	valid

- Most validation errors and warning messages caused by *data quality problems*:
 - wrong usage of unit codes, e.g. “l” instead of “LTR” (UN/CEFACT)
 - non-numeric feature values where floating point values are expected, e.g. “/”, “0.75/2.2”, “3*16”, ...

Evaluation

Part 1:

Conversion of BMEcat catalogs of two manufacturers

1. Validation of the quality of the conversions using Semantic Web validators and the standard output of the converter

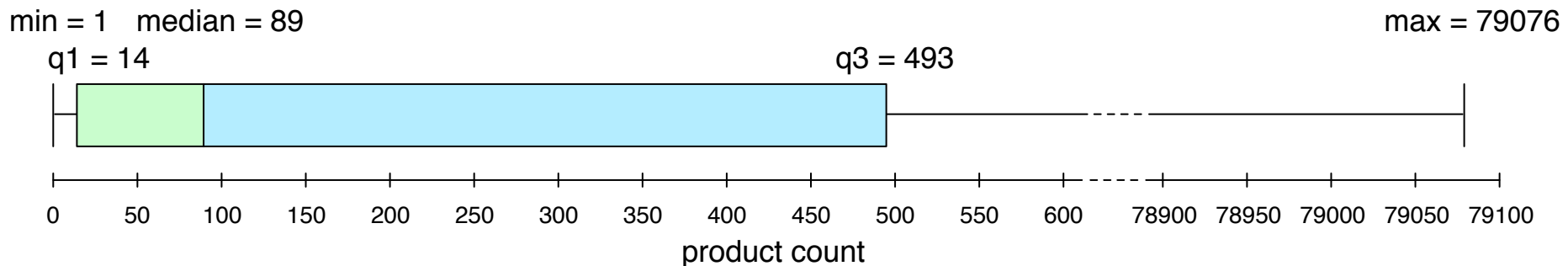
Part 2:

Collected product offer data of 2,500+ Web shops in SPARQL-capable RDF store

2. Complement the motivating example from the introduction by showing the lack of product master data in the crawl data
3. Show potential leverage of product master data on the Web

Collection of Web Product Offers

- Focused crawl over structured e-commerce data on the Web
 - 2,629 different-sized online shops exposing RDFa/Microdata
 - 2.7 million offered product instances
 - 11 product properties on average



Distribution of product count across Web shops

Results that Confirm Our Initial Example

Manufacturer Product Features		Retailer Product Features		Coverage ¹
Samsung LED TV ES6300	89	15	amazon.de	28.09%
		39	notebooksbilliger.de	
		22	conrad-electronics.de	
		24	voelkner.de	

⋮

BMEcat catalog data

Crawled Web shop data

BSH Product Features			Retailer Product Features		Coverage ¹⁵
TW86103	Wasserkocher	25	10	marketplace.b2b-discount.de	40%
(EAN: 4242003535615)					
Bodenstaubsauger Beutel VS06G2410 2400 W (EAN: 4242003356364)	30		10	www.ay-versand.de	30%
			9	www.megashop-express.de	
			8	fairplaysport.tradoria-shop.at	
Mikrowelle HF25M5L2 Edel- stahl (EAN: 4242003429303)	51	7		www.european-gate.com	13.73%

Motivation

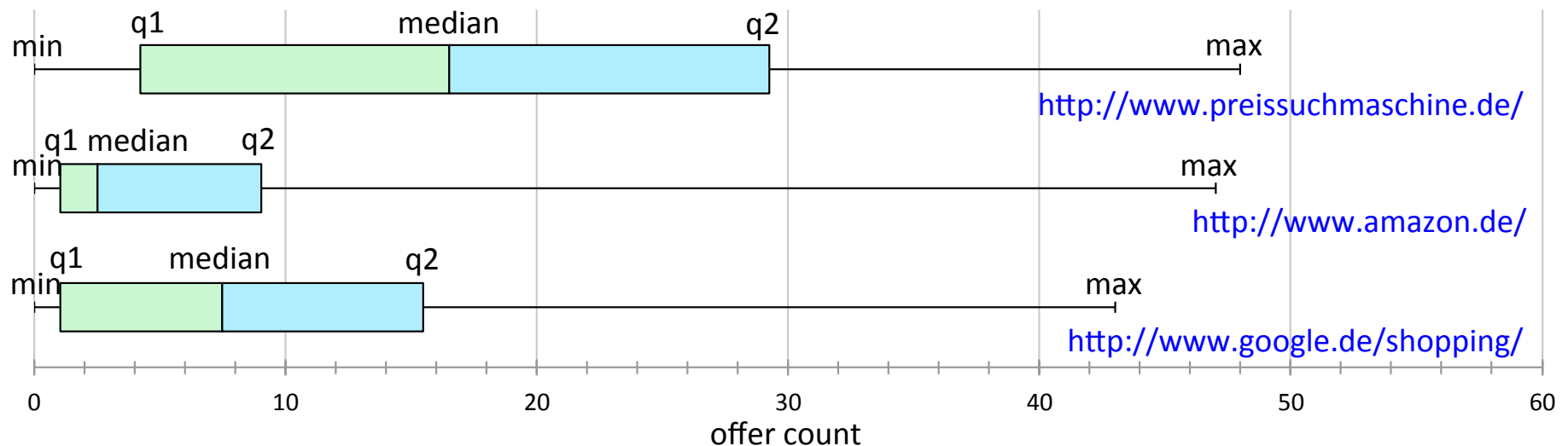
Our Findings

Leverage of Product Master Data on the Web

- Selected 90 random product EANs from BSH BMEcat catalog
{ population size = 1,376; 95% confidence level; 10% confidence interval }
- Looked up the number of product offers for each item in the sample, based on the European Article Number (EAN):
 - preissuchmaschine.de
 - Amazon.de
 - Google Shopping Germany

Results of the Potential Leverage of Master Data

- Surprisingly little results returned based on EAN
 - Max. number of product offers for a specific product EAN: 48
 - Most products on Amazon.de and Google Shopping Germany are offered less than 10 times



Distribution of shop offers per European Article Number (EAN)

Discussion of Results

- Possible explanations for the fairly small number of results:
 - Marketplace regulations try to limit competition among market participants (speculative)
 - Barrier of adding products to marketplaces for small shop owners
- Potential weaknesses in our experiment
 - Localized searches (.de-domain)
 - Shops rarely populate their products with EAN identifiers
 - Type of products in the sample (domain of white goods) are maybe not very attractive for being offered on the Web

Replacing Strong Identifiers and Product Type

- Repeated experiment with a more popular commodity product, searching by
 - EAN
 - product name
- “Canon PowerShot A2300 schwarz”

Marketplace	EAN hits	Product name hits
preissuchmaschine.de	45	45
Amazon.de	233	233
Google Shopping Germany	4	144

A combination of different types of strong identifiers could do the trick!

Conclusion

- We showed the potential leverage of enriching
 - product offers of Web shops
 - by structured product model data of manufacturers
 - relying on reliable product *strong identifiers* such as EAN, GTIN, MPN, brand name, ...
- High-quality product master data can reduce data cleansing effort
- Prerequisite for doing useful product searches and matchmaking

Thank You!

Alex Stolz

alex.stolz@unibw.de

Project page:

<http://code.google.com/p/bmecat2goodrelations/>



GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

Acknowledgments

- This work has been supported by:
 - German Federal Ministry of Education and Research (BMBF) by a grant under the KMU Innovativ program as part of the Intelligent Match project (FKZ 01IS10022B)
 - Eurostars program within the EU 7th Framework Program of the European Commission in the context of the Ontology-based Product Data Management (OPDM) project (FKZ 01QE1113D)
- Special thanks also goes to Mark Mattern, who provided a first mapping from BMEcat to GoodRelations