

Using webcrawler techniques for improved market surveillance – new possibilities for compliance and energy policy.

eceee Summerstudy June 2017

Peter Bennich and Linn Stengård,

The Swedish Energy Agency

Signe Friis Christensen,

The Danish Energy Agency

Teemu Hartikainen,

The Finnish Safety and Chemicals Agency Kasper Mogensen and Troels Fjordbak Larsen, Big2Great ApS



# What if... you could *watch* the market at *any instant?*



### Nordcrawl – a Nordic project to explore webcrawler as a tool for MV&E



• Financed by the Nordic Council of Ministers

tukes

Mannvirkja

- Nordic *energy agencies* joint project between autumn 2015

   summer 2017
- **Goal**: Develop a *software platform* for MV&E using web crawler data

### Content

- Background and core concepts
- Information crawled during pilot project
- Opportunities and challenges for monitoring, verification and enforcement (MV&E) including legal considerations



### Webcrawling in brief

#### Data harvest:

- Crawler that starts at URL and visits links down to a specific depth
- Can download multiple pages at the same time (parallel processes)
- Can download from normal pages (click on link page for next page)
- Downloads from non-sequential pages (scroll to bottom/ load new product to same page etc)
- Downloads variety of files (HTML, PDF, word, excel)



### Webcrawling in brief (cont)

	Produkt	Lägsta pris		Omdöme	•	Тур	Ljudnivå	Kapacitet	2
	Bosch SMU51M12SK (Vit) Diskmaskin	4 990:- ₹ 9 ≋	#√	★★★★★★★★★ 7 omdömen	1	Fristående	44 dB	13 kuvert	
	Bosch SMU50M95SK (Rostfri) Diskmaskin	<b>4 659:-</b> ♦ 13 ਸ਼ੱ	#√	★★★★★★★★ 2 omdömen	2	Fristående	44 dB	13 kuvert	
	Bosch SKS62E22EU (Vit) Diskmaskin	2 890:- ▼ 17 Ħ	#√	********* 2 omdömen	3	Bänkdiskmaskin	48 dB	6 kuvert	
	Siemens SN65M046EU Diskmaskin	<b>5 890:-</b> ₹ 12 ₹	#√		4	Integrerad	44 dB	13 kuvert	
	Matsui MDWTT15E (Vit) Diskmaskin	1990:- ₹ 1 ≋	#√	★★★★★★★★★★ 4 omdömen	<b>3</b>	Bänkdiskmaskin	49 dB	6 kuvert	
C C C	Siemens SN45M209SK (Vit) Diskmaskin	5790:- ₹ 6 ₹	#√	★★★★★★★★★★ 1 omdöme	6	Fristående	44 dB	13 kuvert	
E-CONTRACTOR	Electrolux ESF2200DW (Vit) Diskmaskin	3 290:- ₹ 11 ≍	#√	★★★★★★★★★★ 3 omdömen	7	Bänkdiskmaskin	56 dB	6 kuvert	
	Candy CDCF 6/E (Vit) Diskmaskin	<b>1790:-</b> ♦ 3 Ħ	#√	******** 2 omdömen	<b>3</b> ⇒	Bänkdiskmaskin	55 dB	6 kuvert	
	Siemens SX76P030EU Diskmaskin	<b>4 792:-</b> ≌ 14 ≋	#√		9	Integrerad	44 dB	13 kuvert	
	Bosch SMU50M15SK (Rostfri) Diskmaskin	<b>5990:-</b> ♦ 17 ছ	#√	★★★★★★★★★★ 1 omdöme		Fristående	44 dB	13 kuvert	

Typical website from Pricerunner and similar websites

# Raw information found in the web pages (product fiche)

Informationsblad enligt "KOMMISSIONENS DELEGERADE FÖRORDNING (EU) nr 1059/2010"

Varumärke:

Modellbeteckning:

Kapacitet i antal standardkuvert för standarddiskcykeln: 14

Energiklass: A+++

Energiförbrukning 237 kWh/annum, baserad på 280 standarddiskcykler vid kallvattenanslutning och förbrukning enligt energisparläge. Den faktiska energiförbrukningen beror på hur maskinen används.

Energiförbrukning i en standarddiskcykel: 0,83 kWh

Effektförbrukning i frånläge och viloläge: 0,50 W / 0,50 W

Vattenförbrukning 2660 l/annum, baserad på 280 standarddiskcykler. Den faktiska vattenförbrukningen beror på hur maskinen används.

Torkeffektsklass A på en skala från G (minst effektiv) till A (effektivast).

Standardprogrammet (eco 50) är den standarddiskcykel som informationen på etiketten och informationsbladet hänför sig till. Detta program är avsett för diskning av normalt smutsat gods och är det effektivaste programmet när det gäller kombinerad energi- och vattenförbrukning.

Programtiden för standarddiskcykeln: 195 min

Vilolägets varaktighet: 0 min

Utsläpp av luftburet akustiskt buller: 44 dB(A) re 1pW

Inbyggd



### Raw information found in the web pages (cont) Specifikationer

#### Förbrukning

•

### Web page with product information:

- 1) Downloads text from page
- 2) Saves for later
- 3) Extracts info from text

Energiklass	A++
Torkeffekt	A
Energiförbrukning	0,92 kWh
Vattenförbrukning	9,50 liter pr. drifttillfälle
Arlig vattenförbrukning	2660 liter
Arligt energiforbrug	262 kWh/år

#### Allmänna egenskapen

Färg/Material	Rostfritt stål
Standardkuvert	13
0 Ljudnivå	44 db(A)
Extra hög	Nej
Bestick bricka	Nej
Timer-funktion	Ja
AquaStop	Ja









### Structure

There are four parts:

- 1. Information on products, stored at various publicly available web sites.
- 2. The web crawler engine and a temporary *raw* data storage.
- 3. A product data base containing *processed* data.
- 4. The analysis module (a program), which access and display the data in various ways.



### Webcrawler data vs "GfK"-type data

- "GfK"-data:
  - Expensive
  - Quite high coverage of the market
  - Typically aggregated data, on a monthly or annual basis
  - Detailed data possible, but very expensive
  - Sales data
- Webcrawler data:
  - Cheap
  - Coverage has to be checked; can be equal to GfK-data
  - Data in real time, allows a very high time resolution
  - Can be aggregated or very detailed
  - Sales can be estimated, but requires special analysis

### Information crawled during pilot project

- The project have been crawling data from 2015 with a time resolution of one week
- Products on the Nordic market
- Variety of products regulated by ecodesign and energy labelling – see the table
- Preliminary results regarding
  - Vacuum cleaners
  - Washing machines
  - Refrigerators freezers
  - Lighting



Lot	Lot name	Appliance	Difficult	Reason	Attributes
1	813/2013 and 811/2013: Spa	Heat pump	Medium - Hard	little info on pages, and	Price, wattage, COP, type,
1	813/2013 and 811/2013: Spa	Space heaters	Medium - Hard	little info on pages, and	Price, wattage
2	814/2013 and 812/2013: Wat	Water heater	Medium	Little info on energy usage etc.	Price, wattage 230/400 V,
3	PC:s and servers: 617/2013	Desktop + Servers	Hard	Little info on energy usage etc.	Price, CPU, Ram ect.
3	PC:s and servers: 617/2013	Laptop	Hard	Little info on energy usage etc.	Price, CPU, Ram ect.
4	Imaging equipment, Voluntar	Imaging equipment	Medium	Little info on energy usage etc.	Price, type, wattage on/sta
5	Televisions: 642/2009 and 10	TV	Possible		Price, energy use, wattage
6	6 Standby and off-mode losses Standby		Depends on the a	Many different products, not all info on standby.	Price, wattage on/standby
7	Battery chargers and externa	Battery chargers and extern	Hard	Part of other products	
8-9	Tertiary Lighting: 245/2009 (L	See lot 18-19			
10	Residential ventilation: 1253/	Residential ventilation	Hard	New => not much info online	
10	Room air conditioning applia	Room air conditioning appl	Hard	See heat pumps	
11	Circulators in buildings: 641/2	Circulators in buildings	Hard	Few retailers, few info on pages	
11	Electric motors: 640/2009 an	Electric motors	Hard	Not sold to private online	
11	Ventilation fans: 327/2011 (L	Ventilation fans	Hard	Probably not relevant in SE	
11	Electric pumps , Lot 11	Electric pumps	Hard		
12	Commercial refrigerators and	Commercial refrigerators a	Hard	All commercial products are difficult	
13	Domestic refrigerators and fr	Refrigerator freezer	Possible		Price, energy use, energy c
13	Domestic refrigerators and fr	Chest freezer	Possible		Price, energy use, energy c
13	Domestic refrigerators and fr	Refrigerator	Possible		Price, energy use, energy c
13	Domestic refrigerators and fr	Upright freezer	Possible		Price, energy use, energy c
14	Domestic dishwashers: 1016/	Dish washer	Possible		Price, wash class, energy u
14	Domestic washing machines:	Clothes washer	Possible		Price, spin class, energy cla
15	Solid fuel boilers:2015/1187 a	Solid fuel boilers	Hard	Little and scattered info	
16	Laundry driers: 813/2013 and	Dryer	Possible		Price, energy class, noise, o
17	Vacuum cleaners: 666/2013 a	Vacuum cleaners	Possible		Price, wattage, energy clas
18	Simple set-top boxes: 107/20	Simple set top boxes	Possible	Little information on energy use	Price, wattage, wattage sta
18	Complex set-top boxes (Lot 1	Complex set-top boxes	Possible	Little information on energy use	Price, wattage, wattage sta
18	Directional lighting: 1194/201	Light source	Possible	We need to research futher into how to distinguish n	Price wattage, socket, volt
19	Domestic lighting (general lighting	Light source	Possible	We need to research futher into how to distinguish n	Price wattage, socket, volt
20	Local room heating products;	Local space heating produc	Medium		Price, wattage
21	Central heating products (oth	er than CHP), Lot 21	Hard		
22-23	Kitchen appliances: 65/2014 a	Oven	Possible		Price, volume, energy class
22-23	Kitchen appliances: 65/2014 a	Hobs	Possible		Price, Wattage, number of
22-23	Kitchen appliances: 65/2014 a	Domestic range hoods	Possible	May have some data gabs	Price, wattage, energy use,
24	Professional wet appliances a	and dryers, Lot 24	Hard	All commercial products are difficult	
25	Non-tertiary coffee machines	Coffee machines	Possible	Some may lack auto power off	
26	Networked standby losses, re	Networked standby	Depends on the a	Many different products, not all info on network star	Price, wattage
27	Uninterruptible power supplie	Study ongoing.	Hard		
28	Pumps for waste waters, Lot	Study ongoing.	Hard		
29	Large pumps and pumps for p	Study ongoing.	Hard		
30	Special motors and variable s	peed drives, Lot 30	Hard		
31	Compressors, Lot 31		Hard		
32	Window products,Lot 32		Medium - Hard	But it wil probably be easier in the future	Price, sizes, (some U-value
?	-	Microwave/mini oven	Possible		Price, effect, volume,
?	Medical imaging equipment		Hard		
?	Taps and shower heads, JRC		Medium - Hard		Shower heads: Price, sizes,
?		Tyres	Medium	Extremly number of combinations (42.000 on prisjak	Price, load index, Effecienc
ENTR Lot 1	Refrigerating and freezing equ	professional	Hard		
ENTR Lot 2	Distribution and power transf	professional	Hard		
ENTR Lot 3	Sound and imaging equipmen	Gaming consols	Hard	little real info on energy use	
ENTR Lot 4	Industrial ovens, ENTR Lot 4		Hard		
ENTR Lot 5	Machine tools		Hard		
ENTR Lot 6	Tertiary Air Conditioning:1253	3/2014 (ENTR Lot 6)	Hard		
ENTR Lot 7	Steam boilers: ENTR Lot 7	Study ongoing.	Hard		
ENTR Lot 8	Power cables, ENTR Lot 8	Study ongoing.	Hard		
ENTR Lot 9	Enterprise servers, ENTR Lot	Study ongoing.	Hard		

### How fast does the market *adapt* to new requirements? Ex: Vacuum cleaners

Missing Energy label 2016



Ecodesign and Energy label regulations in force 1<sup>st</sup> September 2014 (relatively recently)

### How fast does the market *move*? Vacuum cleaners (cont)



#### Energy label: The higher classes grow quickly

### How fast does the market *move*? Vacuum cleaners (cont)



Whereas the average power decreases Note: data only from "Prisjakt"; not so fast to update their information.

## Another example on how the market moves: Washing machines



The highest energy class grows also here. So labeling works... but the need for re-scaling is obvious!

### **Washing machines (cont)**



Washing machines average capacity *per week* 

Or does it? Checking the average capacity... Y = 7.2+0.00436x => increase *per year* = 0.23 kg

### Washing machines (cont)



The EEI drops as a function of capacity... The Pearson coefficient is = -0.57 rather than = 0



# Relation between *energy class and price:* Washing machines (cont)





Cheap – lower quartile



Expensive – upper quartile

### **Another example: Refrigerator-freezers (combi)**

Development in energy classes



### **Refrigerator-freezers (cont): Declared** *climate classes*



Discount in the energy class and declared annual energy use?



### Final example: Lighting



LED is growing on the market (A+, A++)

### Final example: Lighting



### Final example: Lighting



### So what do we learn?

- Possible to follow the market in real time, and over time
- Information on
  - Adaptation patterns
  - Speed with which the market evolves
  - Not only energy labels, but also other features
- In all, allow more in-depth analyses

![](_page_25_Picture_7.jpeg)

### Challenges

- Crawled websites are often dealer sites.
- Crawled websites do not include all dealers, nor do they include all brands and models of all products.
- Crawled data includes data on business-tobusiness products and products placed on the market before legal requirements came into effect.
- Only online data

![](_page_26_Picture_5.jpeg)

![](_page_27_Picture_0.jpeg)

- Copyright-protected information, such as webpages, websites or source code, shall not be copied.
- No personal data shall be copied or stored.
- Stored and used data shall be correct and up to date, as all such data may be subject to requests for public access.

### How can we use this for MV&E?

![](_page_28_Picture_1.jpeg)

- First of all, the data is only used as an *indication* when choosing the products – the Monitoring part
- The Verification and Enforcement are separate and more *formal* steps
  - Testing (in a lab)
  - Product information (provided in shops and at websites etc)
  - Dialogue with the retailer/manufacturer

![](_page_28_Picture_7.jpeg)

### How can we use this for MV&E (cont)?

- Much easier to work together on the Nordic (i.e. regional) level
- Allow for more flexible sampling strategies
  - Random sampling
  - Easier to identify products that may not be compliant
  - Groups of actors (e.g. old vs new companies)
  - Groups of products (recently regulated or not)
  - Or chose according to any parameter you like
- Allow for more efficient proactive MV&E
  - Inform actors before and during implementation; track the impact of these efforts

### **Future work**

- Continue to crawl data on a weekly basis on the Nordic market
- Continue to work on both the Swedish and the Nordic level with the aim to establish the crawler tool as a foundation for both MV&E and policy analysis
- Large interest from other jurisdictions from within but also outside EU - most notably AU, US and CAN, who have databases in place.
- Considering a H2020-application let us know if you are interested to join!

### **Thank you!**

#### Peter.bennich@energimyndigheten.se

![](_page_31_Picture_2.jpeg)