Statistics Poland

Online Based Enterprise Characteristics (OBEC) in Statistics Poland

Ewelina Niewiadomska
Statistics Poland

Web Intelligence Network Conference. From Web to Data, 4-5 February 2025, Gdańsk

1

Outline

- Community Survey on ICT usage

- Web data as information on enterprise characteristics

- Web data vs ICT survey 2023

- Challenges

General information about ICT survey

- Annual survey since 2005

- **Target population:** Enterprises with 10 or more employees, classified in sections C,D,E,F,G,H,I,J,L,M,N, group 95.1 of NACE Rev.2 (industry and service activities).

  Some activities are excluded i.e agriculture, mining, financial services

- **Scope:** Model Questionnaire - Questionnaire should be understand for non ICT specialists – efforts to avoid technical terms

- **Type of the survey:** Each country makes independent decision

- **High data users needs** – variety of topics, limit of the questions, exchange of the questions, short time series, breaks in series

Statistics Poland

ICT survey 2023

COMMUNITY SURVEY ON ICT USAGE AND E-COMMERCE IN ENTERPRISES

2023
General outline of the survey

Web presence

Use of a website		
A4.	**Does your enterprise have a website?** (Filter question)	Yes ☐ ☐ ->go to **A6**

A5.	**Does the website have any of the following?**	Yes	No
	a) Description of goods or services, price information	☐	☐
	b) Online ordering or reservation or booking, e.g. shopping cart	☐	☐
	c) Possibility for visitors to customise or design online goods or services	☐	☐
	d) Tracking or status of orders placed	☐	☐
	e) Personalised content on the website for regular/recurrent visitors	☐	☐
	f) A chat service for customer support (a chatbot, virtual agent or a person replying to customers)	☐	☐
	g) Advertisement of open job positions or online job application	☐	☐
	h) Content available in at least two languages Please, consider a multilingual website within a single domain (e.g. ".com") or multiple domains of your enterprise in different languages (e.g. ".es", ".uk").	☐	☐

Use of social media		
Enterprises **using** social media are considered those that have a user profile, an account or a user licence depending on the requirements and the type of the social media.		

A7.	**Does your enterprise use any of the following social media?** *(add national examples; replace existing examples if necessary)*	Yes	No
	a) Social networks (e.g. Facebook, LinkedIn, Xing, Viadeo, Yammer)	☐	☐
	b) Enterprise's blog or microblogs (e.g. Twitter)	☐	☐
	c) Multimedia content sharing websites or apps (e.g. YouTube, Flickr, SlideShare, Instagram, Pinterest, Snapchat)	☐	☐

A8.	**Does your enterprise use any of the above mentioned social media to:** - **optional**	Yes	No
	a) Develop the enterprise's image or market products (e.g. advertising or launching products)	☐	☐
	b) Obtain or respond to customer opinions, reviews, questions	☐	☐
	c) Involve customers in development or innovation of goods or services	☐	☐
	d) Collaborate with business partners (e.g. suppliers) or other organisations (e.g. public authorities, non-governmental organisations)	☐	☐
	e) Recruit employees	☐	☐
	f) Exchange views, opinions or knowledge within the enterprise	☐	☐

Statistics Poland

Web data – Enterprises having website

ICT sample 2023

20 039

Business Register
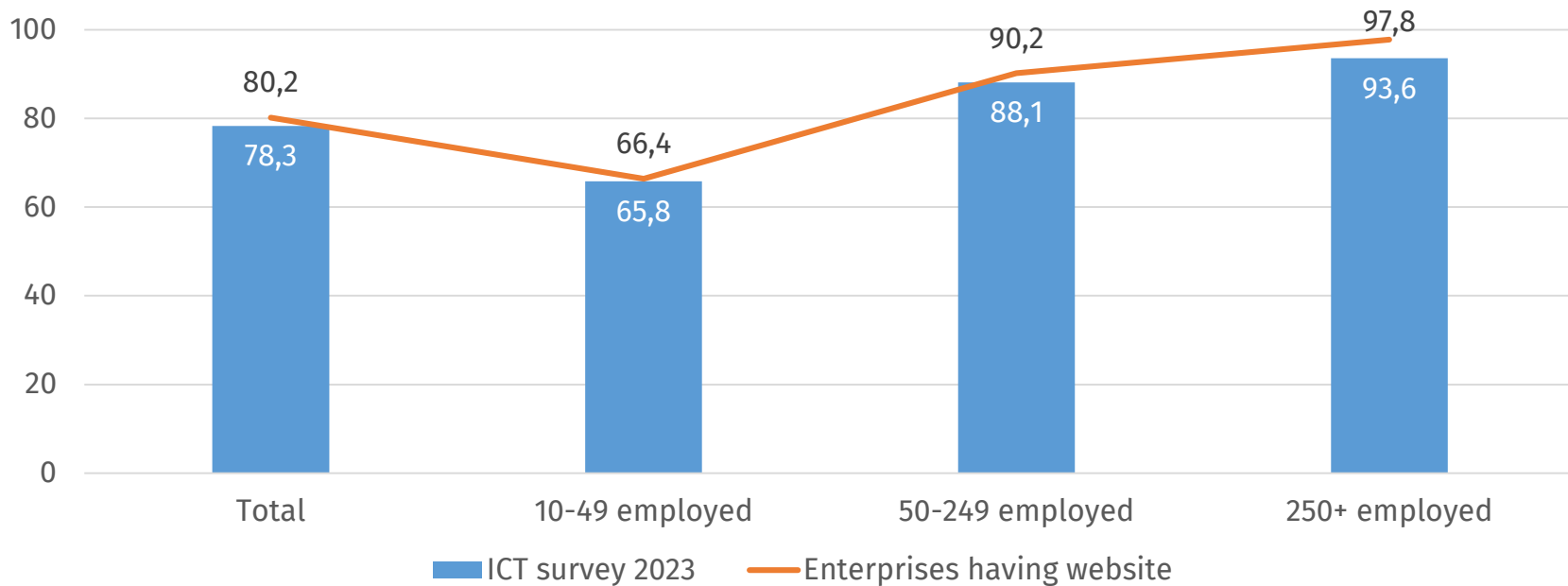
GUrlSearcher

Google Search API

URLs ICT 2022

robots.txt

Web presence		

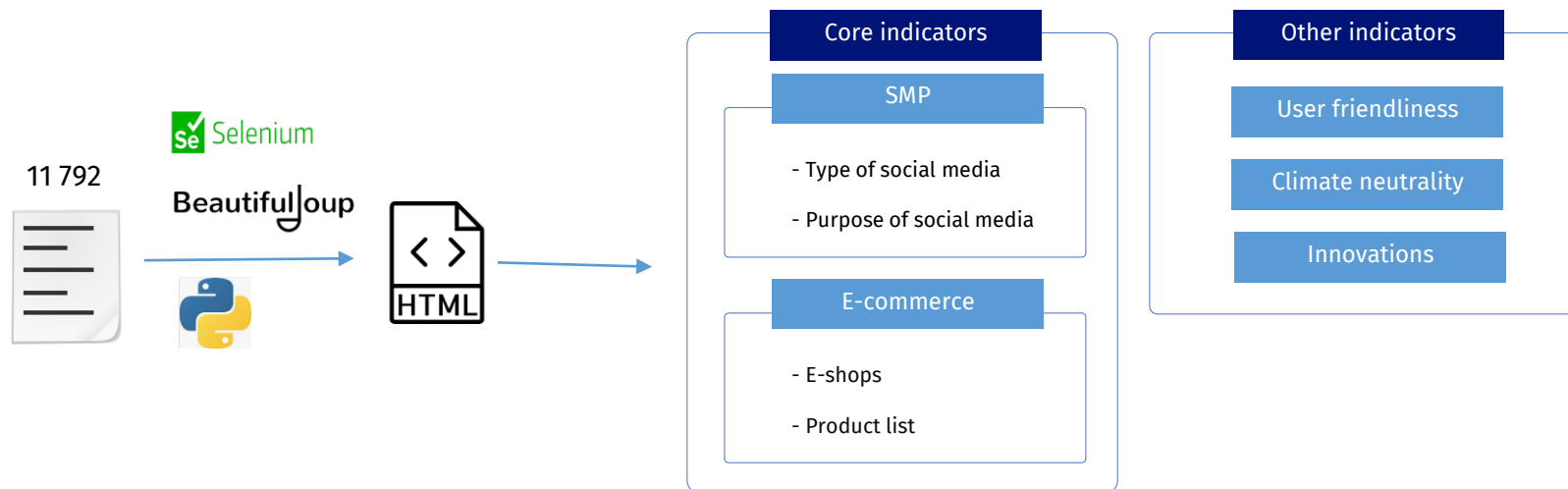|	Use of a website		
A4.	**Does your enterprise have a website?** (Filter question)	Yes ☐	No ☐ ->go to **A6**
A5.	**Does the website have any of the following?**	Yes	No
	a) Description of goods or services, price information	☐	☐
	b) Online ordering or reservation or booking, e.g. shopping cart	☐	☐
	c) Possibility for visitors to customise or design online goods or services	☐	☐
	d) Tracking or status of orders placed	☐	☐
	e) Personalised content on the website for regular/recurrent visitors	☐	☐
	f) A chat service for customer support (a chatbot, virtual agent or a person replying to customers)	☐	☐
	g) Advertisement of open job positions or online job application	☐	☐
	h) Content available in at least two languages Please, consider a multilingual website within a single domain (e.g. ".com") or multiple domains of your enterprise in different languages (e.g. ".es", ".uk").	☐	☐

Statistics Poland

ICT survey 2023

Enterprises having website

Web data – indicators

ICT sample 2023

20 039

11 792

Selenium

BeautifulSoup

HTML

Core indicators

SMP

- Type of social media

- Purpose of social media

E-commerce

- E-shops

- Product list

Other indicators

User friendliness

Climate neutrality

Innovations

Statistics Poland
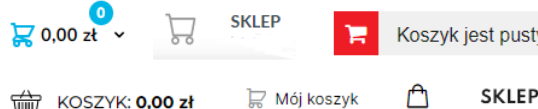
Web data - Enterprises having shopping cart

ICT survey

Does the website have online ordering or reservation or booking e.g. shopping cart?   Yes ☐   No ☐

Shopping cart

- Shopping cart
- Shopping bag
- Other shopping cart synonyms
- Icons

E-commerce Task Team

Defining a larger set of features

- IDs
- Classes
- WooCommerce
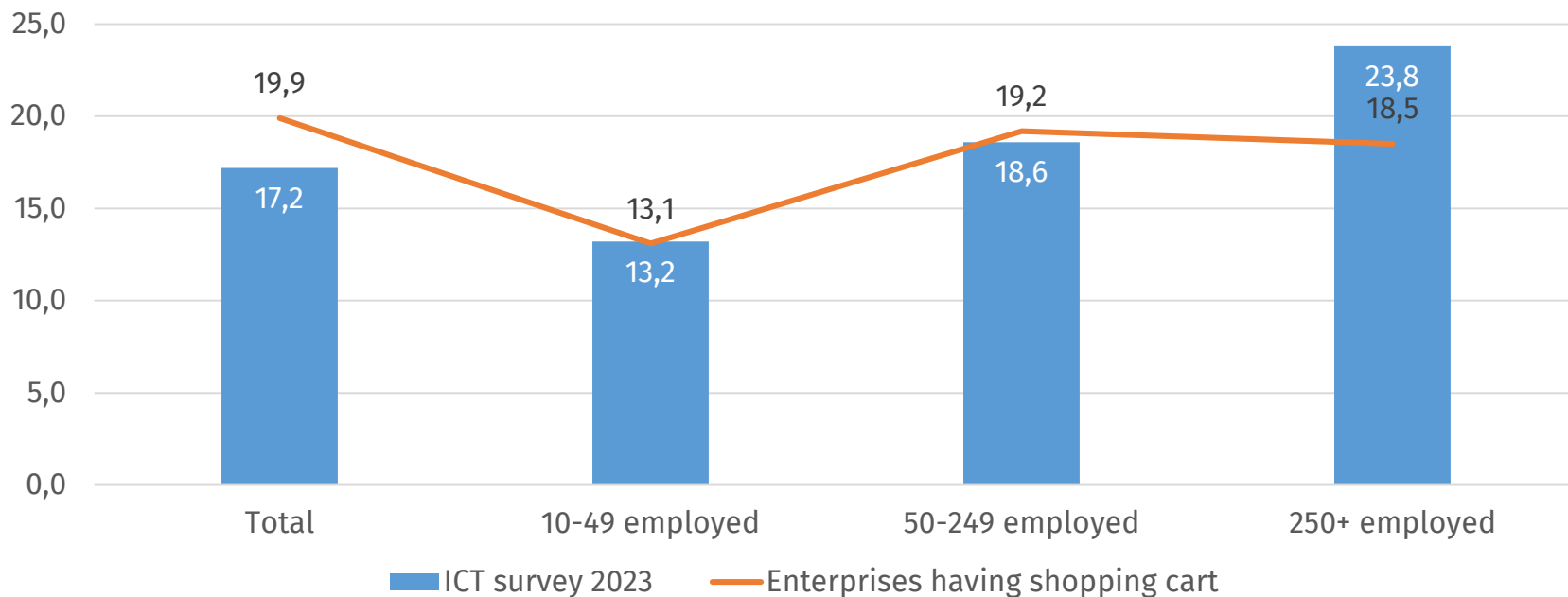- PrestaShop
- Magento
- PayPal
- Amazon Pay
......

#add_to_cart   #add-to-cart-buton
.add-to-cart   .cart_block
.woocommerce   link[rel*='woocommerce']
img[src*='/modules/prestablog/themes/']
script[data-requiremodule*='Magento_']
img[src*='paypal.com']
img[src*='amazonpay']

„Shopping cart" is the most important predictor

Statistics Poland

ICT survey 2023
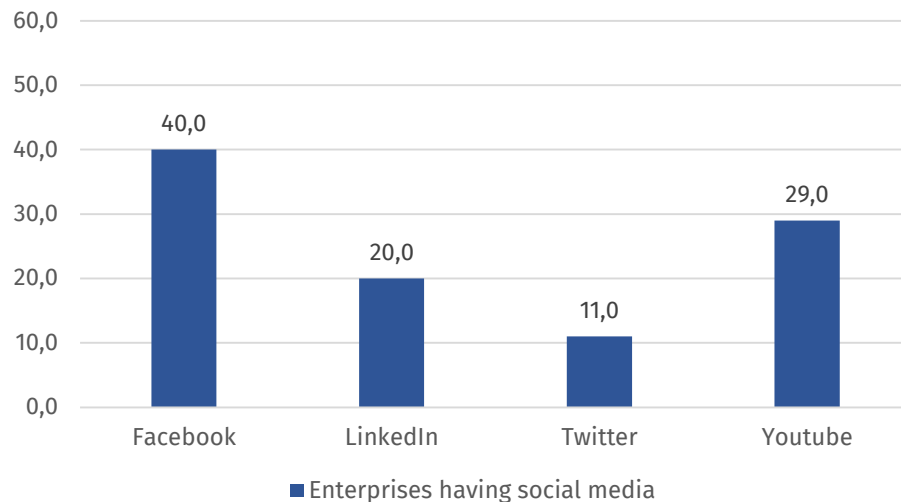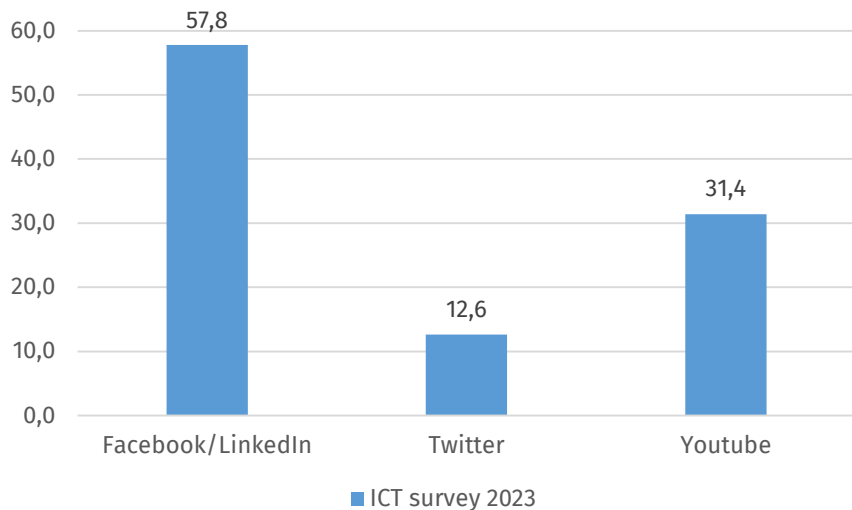
Enterprises having shopping cart

Social media presence

ICT survey

Does your enterprise use any of the following social media?   Yes ☐   No ☐

Statistics Poland

ICT survey 2023 validation

ICT Survey 2023

ID	Does your enterprise have a website?	a5b	Social networks (Facebook, LinkedIn)	Twitter	Youtube
00000000000001	1	2	2	2	2
00000000000002	1	2	1	2	1
00000000000003	1	2	2	2	2
00000000000004	1	2	1	2	2
00000000000005	1	2	1	2	1
00000000000006	1	2	2	2	2
00000000000007	1	1	1	2	2
00000000000008	1	2	1	2	2
00000000000009	1	2	1	2	2
00000000000010	1	2	1	2	1
00000000000011	1	1	1	1	1
00000000000012	1	2	2	2	2
......	.....	.....	.....	.....	.....
000000000000N	2	2	2	2	2

✖

➕

Web Data

URL	a5b	a7a	a7b	a7c
https://www.URL1	2	1	2	1
https://www.URL2	2	2	2	2
NO URL	N/A	N/A	N/A	N/A
https://www.URL4	2	1	1	1
https://www.URL5	2	2	2	2
https://www.URL6	2	2	2	2
https://www.URL7	2	1	1	1
https://www.URL8	2	1	2	2
https://www.URL9	1	1	1	1
https://www.URL10	2	1	1	1
https://www.URL11	1	1	2	2
https://www.URL12	2	2	2	2
.....	.....	.....	.....	.....
https://www.URLn	2	2	2	2

Statistics Poland

Challenges

- Problems with websites (outdated websites, missing websites, errors in URLs)

- Manual validations (the „annotation exercise") takes a lot of time and efforts

- Validation of survey results by checking social media account names

- Validation of e-commerce based on a few features, e.g., shopping cart and product list

Statistics Poland

Statistics Poland

Thank you!