

Using Web Data for energy statistics: Methodology and key lessons

Web Intelligence Network
Conference

From Web to Data

4-5 February 2025

GDANSK, POLAND



Mathieu SCHEFFER
Marianne DOUL
Sandrine HERBETH

Why Web Data?

- Growing need for timely, granular, and comprehensive statistics.
- Web data's potential to complement traditional datasets.

Project Overview:

- "Tapping New Data Sources" awarded by Eurostat.
- Focus: Integrating web-based and alternative data for energy statistics.

Objectives:

- Enhance timeliness and relevance of energy statistics.
- Develop innovative methodologies for web data extraction and validation.



Identification of data sources

- **Reliable, relevant, and accessible**
- **All energy products and energy prices**



71 institutions/associations' websites



69 national websites

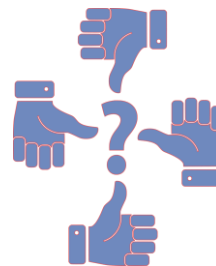
140 websites identified

**Additional
criteria**



Country coverage

Data quality



Time coverage

Energy products presented in the indicators, or prices



18 websites



ENTSO-E, ENTSO-G, GIE, the Energy Institute and Eurofuel

Tool developments – Platforms analysis and needs

entso-e
Transparency Platform

Central collection and publication of electricity generation, transportation and consumption data and information for the pan-European market.

Home Load Generation Transmission Balancing Outages Congestion Management System Operations Data Pre-5.1.15

Installed Capacity Per Production Unit

Installed generation capacity per unit [14.1.B]

Control Area: Bidding zone

Year: 2025

Area: Belgium (BE), BZN|BE

Production Type	Code	Name	Installed Capacity at the beginning of the year	Current Installed Capacity	Location	Voltage Connection Level	Commissioning Date
			[MW]	[MW]		[kV]	
BZN BE							
Nuclear	22WDOELX40000793	DOEL 4	1039	1039	Belgium	380	01.01.2014
Nuclear	22WTIHANG000242R	TIHANGE 3	1038	1038	Belgium	380	01.01.2014
Hydro Pumped Storage	22WCOOXII000070C	COO II T	690	690	Belgium	380	01.01.2014
Nuclear	22WTIHANG000239G	TIHANGE 1N	481	481	Belgium	380	01.01.2014
Nuclear	22WTIHANG000240V	TIHANGE 1S	481	481	Belgium	380	01.01.2014
Fossil Gas	22WHERDER0001288	HERDERSBRUG STEG	480	480	Belgium	150	01.01.2014
Hydro Pumped Storage	22WCOOXIX000067T	COO I T	474	474	Belgium	380	01.01.2014
Fossil Gas	22W201806271--D	EDF Luminus Seraing TGV	470	470	Belgium	220	14.10.2018
Fossil Gas	22WDROGEN0000863	DROGENBOS TGV	465	465	Belgium	150	01.01.2014
Fossil Gas	22WAMERCO000010Y	Amercoeur 1 R TGV	451	451	Belgium	150	01.01.2014

energy institute

Home » Resources and data downloads

Resources and data downloads

Use this section to download the PDF edition of the Statistical Review of World Energy, the Excel data – and more

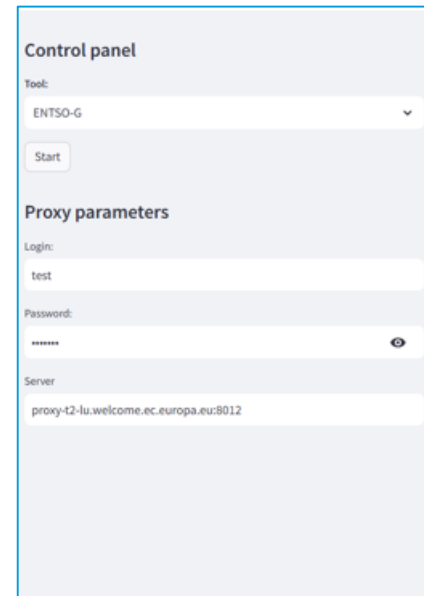
Overview documents

Key reports

- [Statistical Review of World Energy](#) PDF / 10MB ▼
- [Statistical Review of World Energy - Chinese Version](#) PDF / 10MB ▼
- [Statistical Review of World Energy Data](#) XSLX / 4MB ▼ Cite
- [2024 Country Transition Tracker Data](#) XSLX / 79.5KB ▼ Cite

Tool developments - PYTHON

- **Libraries used: Steamlit and BeautifulSoup**
- **Configurable tool:**
 - List of countries
 - List of indicators
 - Time ranges
 - API interactions
- **Extraction into CSV format**



The screenshot shows a 'Control panel' interface. It includes a 'Tool' dropdown menu set to 'ENTSO-G', a 'Start' button, and a 'Proxy parameters' section. The 'Proxy parameters' section contains fields for 'Login' (test), 'Password' (masked with asterisks), and 'Server' (proxy-t2-lu.welcome.ec.europa.eu:8012).



The screenshot shows a 'Data extraction' interface. It features a title 'Extract data from ENTSO-G transparency platform' and several input fields: 'Start date' (2023/12/08), 'End date' (2023/12/16), 'Indicator' (Physical Flow), 'Timezone' (CET), 'Country' (AT), and 'Direction' (entry). An 'Extract data' button is located at the bottom.

Tool developments – Each platform its own set of requirements



ENTSO-E and ENTSO-G

both required direct
database access via API.



The Energy Institute

provided an Excel file, so we
built a feature that allows
users to select only the
needed indicators.



GIE

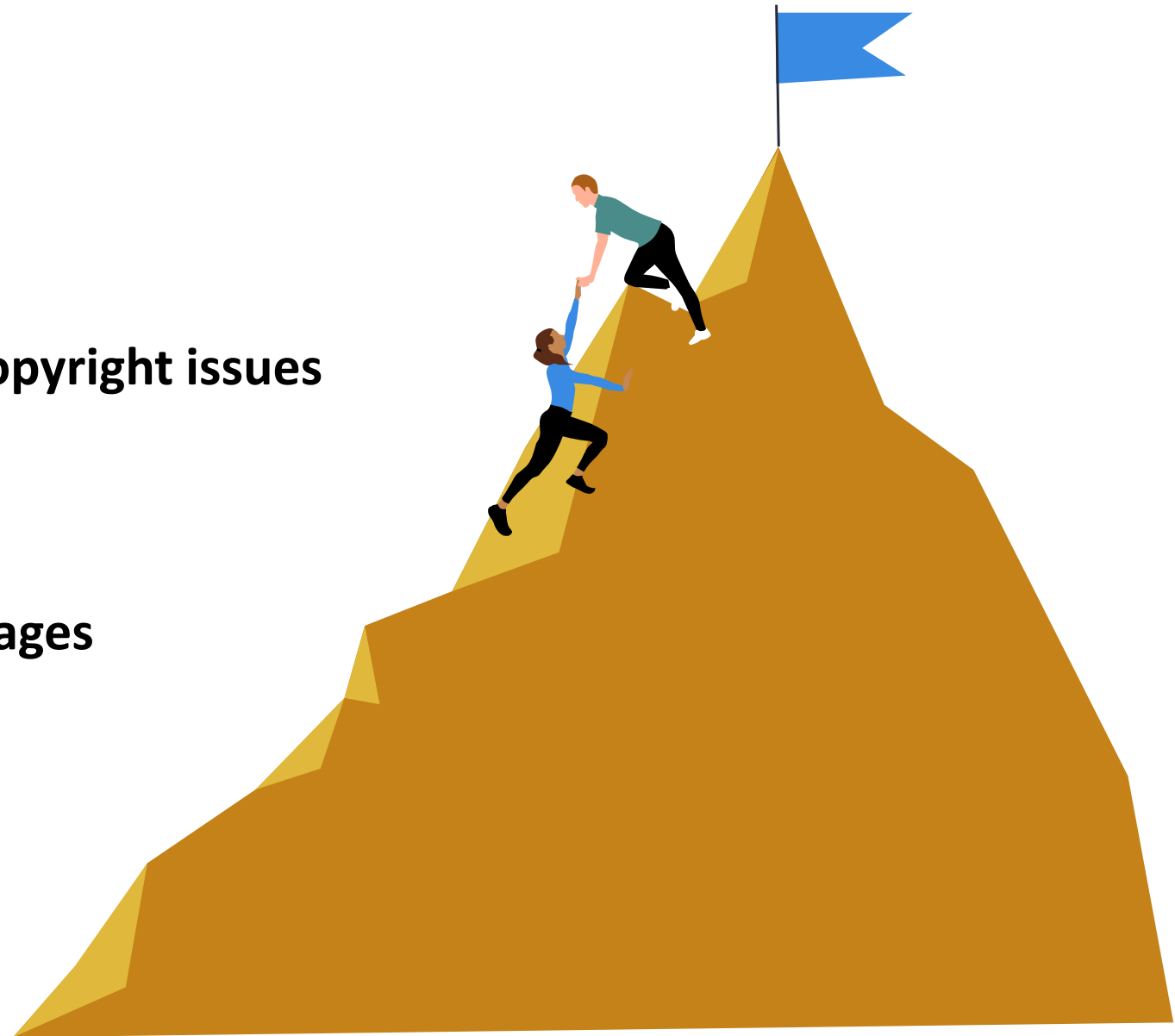
required API access, but the
selection process differed
slightly

Technical challenges

- **API limitations**
- **Proxy restrictions**
- **Confidentiality, access rights and copyright issues**
- **Adapting to different structures**

Specific case of Eurofuel

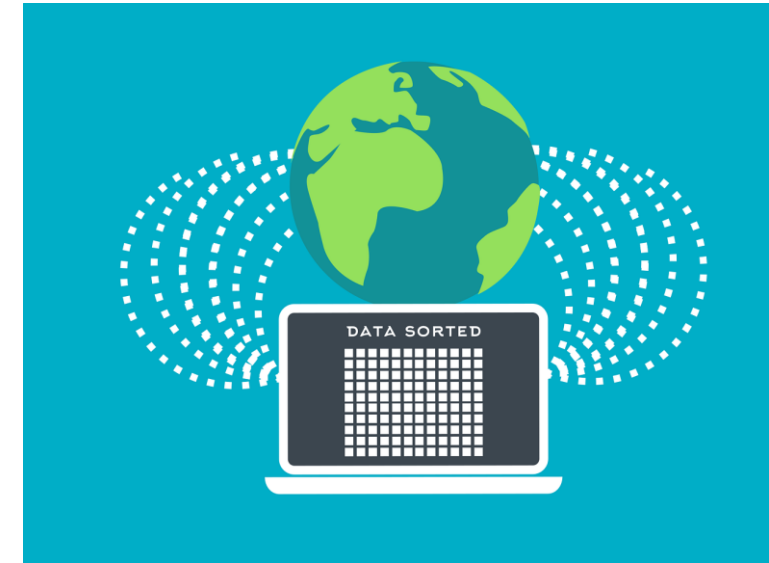
- **Extracting text-based data from images**
- **Proxy security created difficulties**



Results and scalability

- **Results:**

- Successfully validated and extracted data from key platforms.
- Developed reusable tools for scalable web data integration.



- **Scalability:**

- Modular design allows for the addition of new platforms and datasets.
- Potential for application in other statistical domains.

Lessons learned and conclusions

Lessons learned

- Technical adaptability is crucial
- Collaboration with data providers is essential
- Robust validation is necessary

Conclusions

- Potential to enhance official energy statistics
- Mainly used for comparison purposes
- May be used as proxy to improve timeliness
- Might be disseminated in the future?



Thank
You!

Using Web Data for energy statistics: Methodology and key lessons

Project Manager: Marianne.Doul@artemis.lu

Data Scientist: Mathieu.Scheffer@artemis.lu

Speaker and project manager (back-up):
Sandrine.Herbeth@artemis.lu

eurostat 

Eurostat project leader:

Michael.Goll@ec.europa.eu