ARTEMIS
Information Management

Since 1991

Using Web Data for energy statistics: Methodology and key lessons

Mathieu SCHEFFER
Marianne DOUL
Sandrine HERBETH

ARTEMIS
Information Management
Since 1991

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

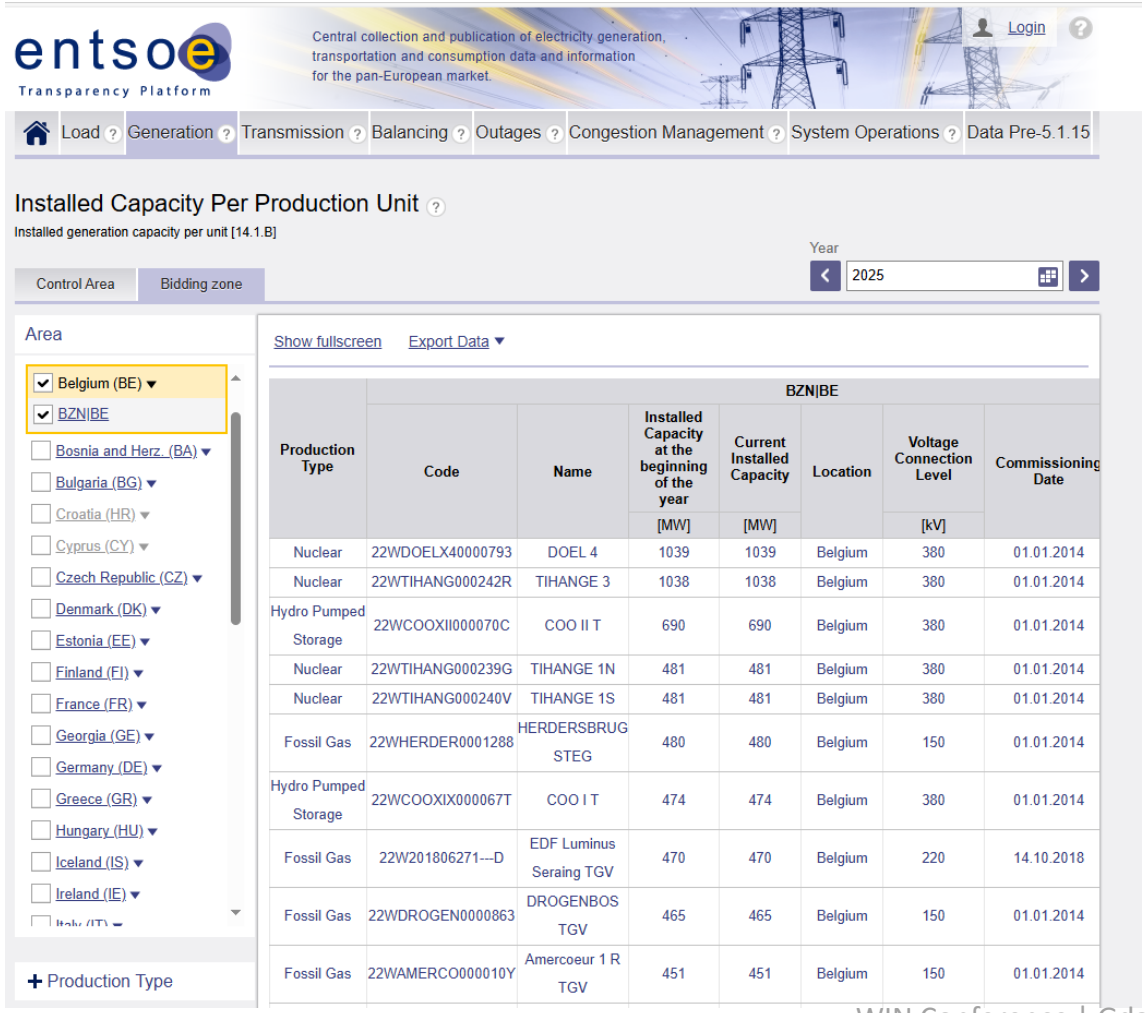
Time coverage

Energy products presented in the indicators, or prices

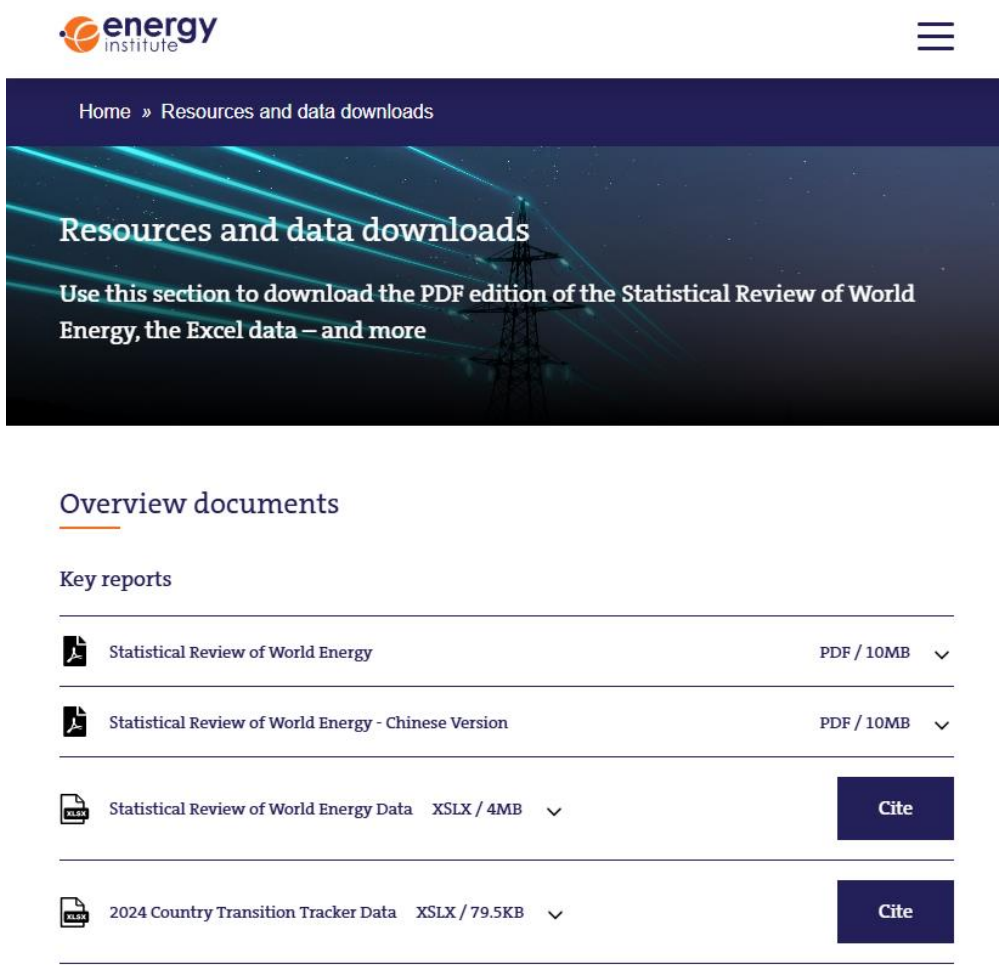
18 websites

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

Tool developments – Platforms analysis and needs

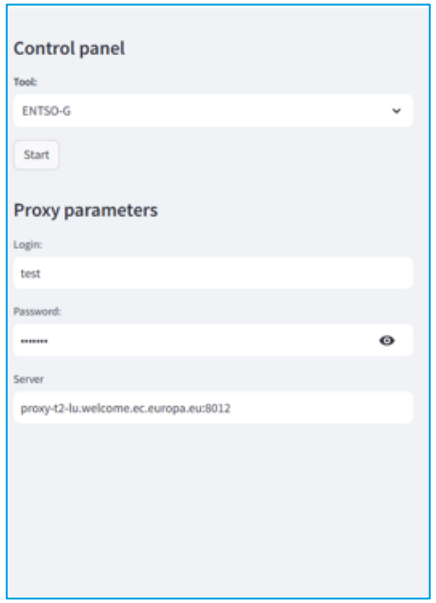


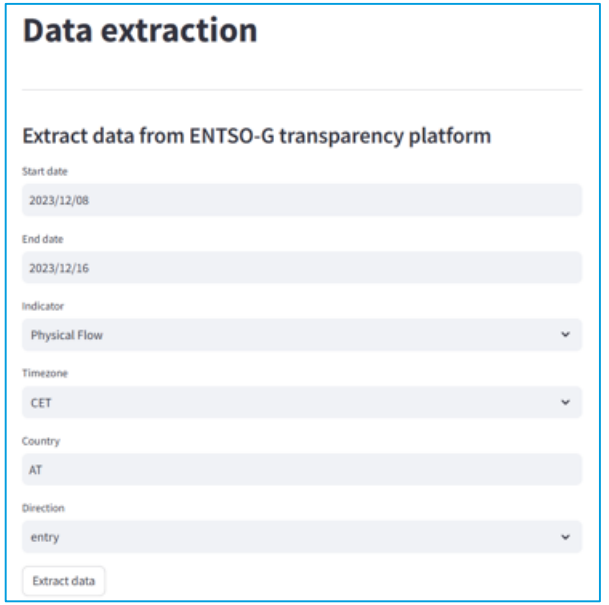
Production Type	BZN|BE						
	Code	Name	Installed Capacity at the beginning of the year	Current Installed Capacity	Location	Voltage Connection Level	Commissioning Date
			[MW]	[MW]		[kV]	
Nuclear	22WDOELX40000793	DOEL 4	1039	1039	Belgium	380	01.01.2014
Nuclear	22WTIHANG000242R	TIHANGE 3	1038	1038	Belgium	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



Tool developments - PYTHON

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





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

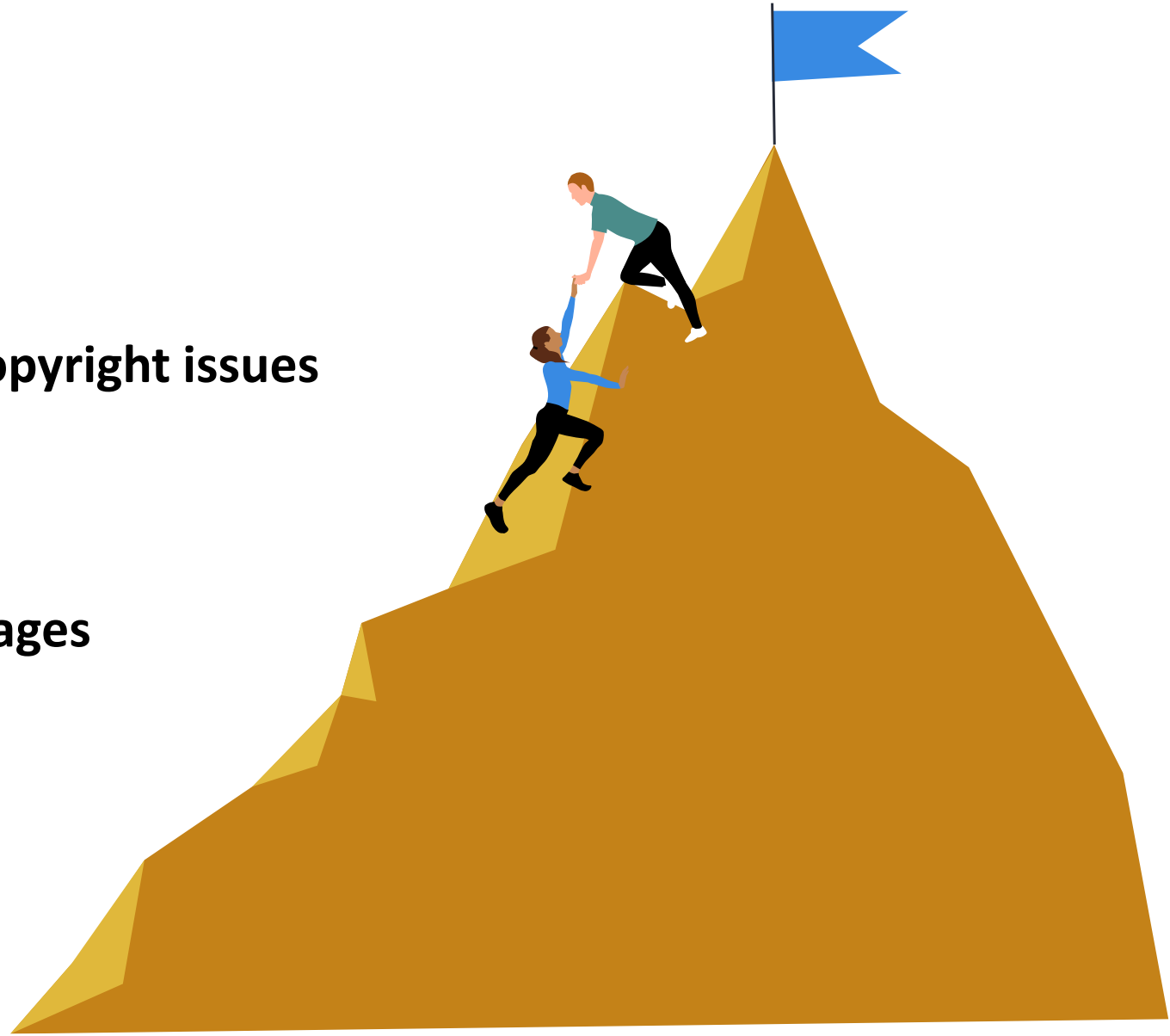
required API access, but the selection process differed slightly

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?

ARTEMIS
Information Management

Since 1991

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

WIN Conference | Gdansk, Poland | 5 February 2025