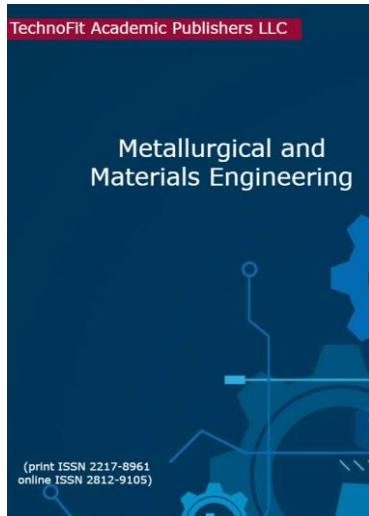# Open Journal Systems (OJS): Best Practices and Use Cases

## Scraping data from old website

Vaso Manojlović, University of Belgrade, Serbia

v.manojlovic@tmf.bg.ac.rs

# MANAGEMENT using PKP OJS platform

**Why OJS?**

• **Open Source Platform**: Free to use, active community, continuous updates.

• **Key Benefits**:
  - Streamlined editorial workflow.
  - Transparent processes for authors, reviewers, and readers.
  - Easy online submission and peer review.

• **Goal**: Improve visibility and credibility (DOAJ, WoS, SCOPUS).

Metallurgical and Materials Engineering (https://metall-mater-eng.com)
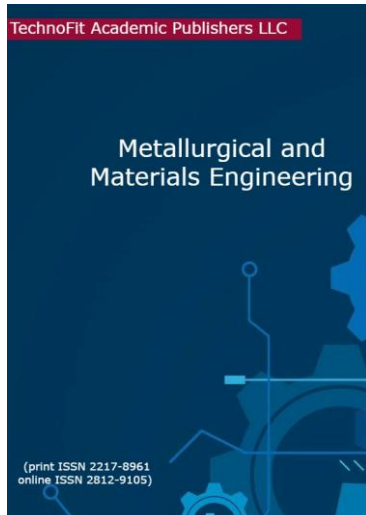
# MANAGEMENT using PKP OJS platform

**First Journal: Metallurgical and Materials Engineering**

- **Started using OJS v2 in 2016**:
  - Needed better editorial policies, transparency, and workflow.
  - Helped us get indexed in DOAJ after one year, later WoS (Extended list, 2017) and SCOPUS (2019).
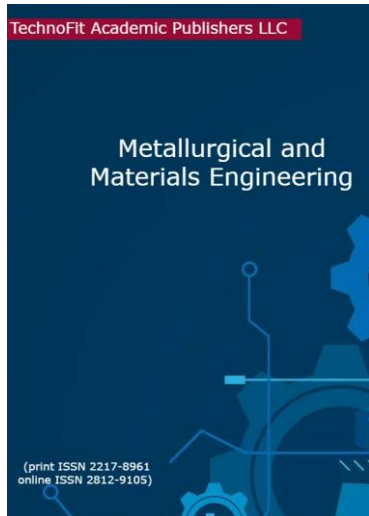- **Challenges**:
  - High volume of submissions (200+).
  - Limited editorial team (volunteers).
  - Decision to transfer ownership in 2023 -> new management lost DOAJ status due to lack of transparency

Metallurgical and Materials Engineering (https://metall-mater-eng.com) from 2016

# MANAGEMENT using PKP OJS platform

**Transition from OJS v2 to OJS v3**

•**Main Challenge**: Databases and plugins not always directly compatible.

•**Helpful Steps**:
- Thorough backup before upgrade.
- Test upgrade on a staging server.
- Manually migrate customized plugins or themes.

•**Advice**:
- Follow the PKP documentation ("Upgrade Guide").
- Expect potential downtime or manual fixes.

Metallurgical and Materials Engineering (https://metall-mater-eng.com) from 2016

# MANAGEMENT using PKP OJS platform

**DOI Implementation**

- **How**: Using the CrossRef (or other) DOI plugin in OJS.
- **What Worked**: Automated DOI assignment for new articles.
- **Issues**:
    - Ensuring accurate metadata (title, authors, references).
    - Maintaining a paid DOI account (CrossRef subscription fees).
- **Tip**: Always verify DOI links and metadata formatting to ensure correct citation.

The journal Metallurgical and Materials Data (https://metall-mater-data.com/) from 2023

Metallurgical & Materials Engineering Congress of South-East Europe
(https://book.mme-see.org/)  from 2025

# MANAGEMENT using PKP OJS platform

**Hosting Challenges**

•**In Academia**:
- Limited or no technical support from university IT.
- University servers may not allow required custom configurations.

•**Solution**:
- Rented a VPS (virtual private server) or cloud hosting for full control.
- Importance of **daily backups** (files + database).

•**Server Crash in 2023**:
- Data rescued by a colleague.
- Migrated to https://akademosbydhilos.com/ with regular support.

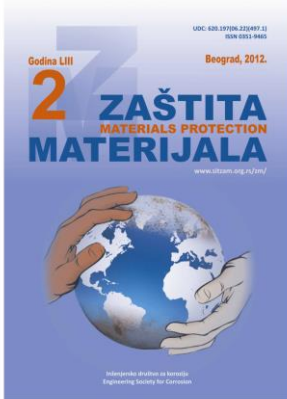The journal Metallurgical and Materials Data (https://metall-mater-data.com/) from 2023

Metallurgical & Materials Engineering Congress of South-East Europe (https://book.mme-see.org/) from 2025

# MANAGEMENT using PKP OJS platform

**Zaštita materijala**

•Main issue: Migrating historical data from an older website.

•**Native XML Plugin**: Import/export articles and issues.

•**When no XML is available**: Manual entry or scripting.

•Wrote a **Python script** on Google Colab to scrape old site data and generate XML, speeding up import.



The journal Zaštita materijala (https://www.zastita-materijala.org/)

# Scraping data from old website

## Scraping data from the website

```python
import requests
from bs4 import BeautifulSoup
import pandas as pd
```

```python
# Base URL of the website
base_url = "https://scindeks.ceon.rs/"

# Page URL containing the list of articles
page_url = "https://scindeks.ceon.rs/issue.aspx?issue=11946&lang=en"

# Function to scrape articles URLs
def scrape_articles(page_url):
    response = requests.get(page_url)
    soup = BeautifulSoup(response.content, 'html.parser')

    articles_data = []
    # Find all article links on the page
    for link in soup.select('div.naslov a'):
        article_url = base_url + link['href']
        articles_data.append(article_url)

    return articles_data
```

```python
articles.pop(10)
articles
```

```
['https://scindeks.ceon.rs/article.aspx?query=ISSID%26and%2611946&page=0&sort=8&stype=0&backurl=%2fissue.aspx%3fissue%3d11946%26lang%3den&lang=en',
 'https://scindeks.ceon.rs/article.aspx?query=ISSID%26and%2611946&page=1&sort=8&stype=0&backurl=%2fissue.aspx%3fissue%3d11946%26lang%3den&lang=en',
 'https://scindeks.ceon.rs/article.aspx?query=ISSID%26and%2611946&page=2&sort=8&stype=0&backurl=%2fissue.aspx%3fissue%3d11946%26lang%3den&lang=en',
 'https://scindeks.ceon.rs/article.aspx?query=ISSID%26and%2611946&page=3&sort=8&stype=0&backurl=%2fissue.aspx%3fissue%3d11946%26lang%3den&lang=en',
 'https://scindeks.ceon.rs/article.aspx?query=ISSID%26and%2611946&page=4&sort=8&stype=0&backurl=%2fissue.aspx%3fissue%3d11946%26lang%3den&lang=en',
 'https://scindeks.ceon.rs/article.aspx?query=ISSID%26and%2611946&page=5&sort=8&stype=0&backurl=%2fissue.aspx%3fissue%3d11946%26lang%3den&lang=en',
 'https://scindeks.ceon.rs/article.aspx?query=ISSID%26and%2611946&page=6&sort=8&stype=0&backurl=%2fissue.aspx%3fissue%3d11946%26lang%3den&lang=en',
 'https://scindeks.ceon.rs/article.aspx?query=ISSID%26and%2611946&page=7&sort=8&stype=0&backurl=%2fissue.aspx%3fissue%3d11946%26lang%3den&lang=en',
 'https://scindeks.ceon.rs/article.aspx?query=ISSID%26and%2611946&page=8&sort=8&stype=0&backurl=%2fissue.aspx%3fissue%3d11946%26lang%3den&lang=en',
 'https://scindeks.ceon.rs/article.aspx?query=ISSID%26and%2611946&page=9&sort=8&stype=0&backurl=%2fissue.aspx%3fissue%3d11946%26lang%3den&lang=en']
```

# Scraping data from old website

```python
def get_article_details(article_url):
    response = requests.get(article_url)
    soup = BeautifulSoup(response.content, 'html.parser')

    # Function to safely extract text
    def get_text(soup, selector):
        element = soup.select_one(selector)
        return element.get_text(strip=True) if element else None
```

```python
    return {
        'Title (EN)': title,
        'Title (SR)': title_sr,
        'Abstract (EN)': abstract_en,
        'Abstract (SR)': abstract_sr,
        'Keywords (EN)': keywords_en,
        'Keywords (SR)': keywords_sr,
        'Authors': authors,
        'Affiliations': affiliations,
        'DOI': doi_url,
        'File': file_link,
        'References': references,
        'Pages': pages
    }
```

# Scraping data from old website

```
[ ]  pd.set_option('display.max_columns', None)
     pd.set_option('display.max_rows', None)

     df_articles
```

| | Title (EN) | Title (SR) | Abstract (EN) | Abstract (SR) | Keywords (EN) | Keywords (SR) | Authors | Affiliations | DOI | File | References | Pages |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Surface engineering of polymers: Case study: P... | Površinski inženjering na polimerima - studija... | Surface engineering as a subset of materials s... | Površinski inženjering je deo nauke o materija... | surface engineering;physical vapor deposition;... | površinski inžinjering; fizičko isparavanje i ... | {'A': 'Trajkovska Petkoska Anka', 'B': 'Nasov ... | {'A': 'Faculty of Technology and Technical Sci... | https://doi.org/10.5937/ZasMat1401003T | https://scindeks-clanci.ceon.rs/data/pdf/0351-... | *** Development of injection and blow extrusio... | 3-10 |
| 1 | A contribution to integrated metal valorisatio... | Doprinos integrisanoj valorizaciji metala u hi... | Zinc belongs to the group of non-ferrous, wide... | Cink spada u grupu obojenih metala sa širokom ... | zinc hydrometallurgy; cobalt valorisation; cob... | hidrometalurgija cinka; valorizacija kobalta; ... | {'A': 'Stanojević D.', 'B': 'Filipović Petrovi... | {'A': 'University of East Sarajevo, Faculty of... | https://doi.org/10.5937/ZasMat1401011S | https://scindeks-clanci.ceon.rs/data/pdf/0351-... | *** Treatment of Residues and Precipitates of ... | 11-25 |
| 2 | Investigation of corrosion behavior of Al-allo... | Ispitivanje korozionog ponašanja Al-legura sis... | Corrosion inhibitors are substances which, add... | Inhibitori korozije su supstance koje dodate u... | corrosion;corrosion rate;corrosion inhibitors;... | korozija;brzina korozije;inhibitori korozije; ... | {'A': 'Grudić Veselinka', 'B': 'Zejnilović Ref... | {'A': 'Metalurško-tehnološki fakultet Podgoric... | https://doi.org/10.5937/ZasMat1401027V | https://scindeks-clanci.ceon.rs/data/pdf/0351-... | Davis, J.R. (1999) Corrosion of aluminum and a... | 27-32 |

# Scraping data from old website

## ˅ Creating XML file

Links:

```
[ ]   Ending publication ID: 515
      Ending internal ID: 531
      Ending file ID: 538
```

```
[ ]   import xml.etree.ElementTree as ET
      import pandas as pd

      # INPUT:
      locale_en = 'en'
      date_submitted = '2013-12-15' #date
      status = '3' #of paper
      submission_progress = ''
      starting_publication_id = 516 #Starting ID
      starting_internal_id = 532  # Starting ID for internal id
      starting_file_id = 539  # Starting ID for file_id
      stage = 'production'
      copyright_year = '2013' #year
```

```
[ ]   # Explode the Citations column so each citation is in its own row
      df_citations = df_articles.explode('References')

      # Export to Excel
      df_citations.to_excel('ZM_vol_55_1_REF_DOI.xlsx', index=False)
```

# MANAGEMENT using PKP OJS platform

**Practical Tips & Conclusion**

•**Keep OJS Updated**: Security patches, new features, plugin improvements.

•**Maintain Good Metadata**: Essential for indexing (DOAJ, SCOPUS, WoS).

•**Transparency & Editorial Ethics**: Critical for journal credibility.

•**Time Commitment**: OJS maintenance requires ongoing effort—testing, backups, and staff training.

# Questions?