

# New way to collect web data to ensure valid academic research

Submitted by: BlueSky Education

Wednesday, 15 June 2022

---

In developing a new framework, researchers from Rotterdam School of Management (RSM) Erasmus University Rotterdam, have unearthed new and underexploited 'fields of gold' associated with web data.

The RSM researcher along with colleagues from Tilburg University, INSEAD, and Oxford University, sought to demystify the use of web scraping and application programming interfaces (APIs) and thereby facilitate broader adoption of web data in academic research.

According to Dr. Johannes Boegershausen:

"Our framework covers the broad spectrum of validity concerns that arise along the three stages of the automatic collection of web data for academic use: selecting data sources, designing the data collection, and extracting the data. In discussing the methodological framework, we offer a stylized marketing example for illustration. We also provide recommendations for addressing challenges researchers encounter during the collection of web data via web scraping and APIs."

While marketing researchers increasingly employ web data, the particular and sometimes treacherous challenges in its collection have received limited attention. For example, how can researchers ensure that the datasets generated via web scraping and APIs are valid? This research team addresses highlights how addressing validity concerns requires the joint consideration of unique technical and legal/ethical challenges.

The article, published in the 'Journal of Marketing', further provides a systematic review of more than 300 articles using web data published in the top five marketing journals. Using this review, the researchers identify how web data has advanced marketing thought. Understanding the richness and versatility of web data is invaluable for scholars curious about integrating it into their research programs.

The full article and author contact information is available for free at:

<https://doi.org/10.1177/00222429221100750>

Interested in readers might also find the companion website to the article relevant, <https://web-scraping.org/>, and can consider participating in webinar about this research hosted by the American Marketing Association on 23 June at 18.00h CET. For more information about the webinar and signing up see <https://www.ama.org/events/webinar/jm-webinar-series-insights-for-managers/>

ENDS/

For more information, a copy of the research paper, or to speak with the researchers, contact Kate Mowbray at BlueSky PR on [Kate@bluesky-pr.com](mailto:Kate@bluesky-pr.com) or call +44 710022871