

Amsterdam Business Research Institute

Analyzing Digital Data in Business and Management Research

Course Manual

Academic year 2025 - 2026



Course title	Analyzing Digital Data in Business and Management Research
Coordinator(s)	dr. Joey van Angeren (main contact); dr. Hakan Ozalp
Lecturer(s)	dr. Joey van Angeren; dr. Hakan Ozalp
Study period	April 2026
ECTs	5 ECTS
Tuition	€1250 20% discount on early bird registration: €1000
Target groups	The Analyzing Digital Data in Business and Management Research course is open to research master students, PhD candidates, postdocs, and junior faculty engaging in research in business and management, broadly defined. This course is an advanced methods course that assumes prior knowledge of business and management or organization studies as well as a basic understanding of quantitative and/or quantitative research methods. Familiarity with programming in Python or R is also required, although full proficiency in those programming languages is not necessary though.
Course content	Many phenomena that are of interest to management and organization scholars are captured in the form of rich digital (trace) data. Detailed work and navigation processes are captured in information systems and activity logs, product characteristics are described in textual product descriptions, gig platforms capture a wealth of data about workers, and CEOs manage the impression of their firms in video-recorded press conferences. The availability of such rich digital data provides novel opportunities for theorization and analysis. To this purpose, scholars of management and organization have in recent years increasingly turned to methods originally developed in computer science, most notably forms of machine learning, to collect and work with digital data. A distinct feature of those methods is that they are applied in both quantitative and qualitative research, for purposes that range from data exploration, to theory development, and onto hypothesis testing. The course Analyzing Digital Data in Business and Management Research provides an introduction to a variety of methods that can be used to analyze digital data in management and organization research, be it for the purpose of quantitative or qualitative analyses. The focus of the course is on the end-to-end research process. The course introduces website scraping and application programming interfaces (APIs) as ways to collect digital data. It surveys and develops hands-on experience with the main research methods to analyze large-scale numerical, text, and audiovisual data, and explores how others have applied those methods in both quantitative and qualitative research. The publication process of papers based on those methods is also covered in the course.

Course goals After the successful completion of this course, participants will be able to:

- Understand the main methods for analyzing large-scale, text, and audiovisual data as well as how those methods can be applied in quantitative and/or qualitative research
- Reflect on the application of methods for the analysis of digital data in research in business and management or related disciplines
- Collect digital data using website scraping or application programming interfaces (APIs)
- Apply methods to analyze digital data in the context of a quantitative and/or qualitative research study
- Navigate the publication process for empirical papers that are based on digital data

Course design This course is centered around a three-day module on the VU campus in Amsterdam, with smaller online sessions before and after. Each course component is dedicated to a specific aspect of conducting research using digital data (e.g., data collection, data analysis, publication process). The three-day module that is taught on campus provides theoretical, reflective, and hands-on experience with specific methods for data analysis. All sessions consist of a mixture of interactive lectures, paper discussions, and hands-on coding sessions. Hence, it is expected that students come well prepared.

Workload The estimated time participants spend on study activities is:

Actively participating in interactive sessions	24 hours
Reading literature and class preparation	50 hours
Working on the individual research assignment	66 hours
Total	140 hours (5 ECTS)

Assessment Attending and actively participating in all sessions is mandatory.

The assessment for this course is based on the following partial grades:

- Class participation that includes paper discussion and a peer review (50%)
- Individual research paper (50%)

Course structure	Dates	Time	Location	Topic
	Tuesday 31 March 2026	10:00-13:00	Online	Introduction to digital data and programmatic data collection
	Tuesday 7 April 2026	10:00-13:00 14:00-17:00	NU-5A59 (VU campus)	Analyzing large-scale data
	Wednesday 8 April 2026	10:00-13:00 14:00-17:00	NU-5A59 (VU campus)	Analyzing textual data
	Thursday 9 April 2026	09:00-12:00 13:00-16:00	NU-5A65 (VU campus)	Analyzing audiovisual data
	Tuesday 14 April 2026	10:00-13:00	Online	Generative AI and the publication process
	TBC	TBC	Online ¹	Individual research project consultations
	Friday 29 May 2026	23:59	Online	Submission deadline individual research paper
Software prerequisites	Throughout the course, students will gain hands-on experience with the application of the various methods that are taught in this course. For this, students will use their own devices. We will provide support for both Python and R (students are free to choose). Installation instructions and further details on other prerequisites will follow in a separate companion document.			

¹ Meetings will be coordinated individually.

Literature

Readings to be updated.

Session 1: Introduction to digital data and programmatic data collection

Method readings:

Claussen, J., & Peukert, C. (2019). Obtaining Data from the Internet: A Guide to Data Crawling in Management Research. Working paper *available from SSRN*: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3403799.

von Krogh, G., Robertson, Q., & Gruber, M. (2023). Recognizing and Utilizing Novel Research Opportunities with Artificial Intelligence. *Academy of Management Journal*, 66(2), 363-373.

Optional readings:

Boegershausen, J., Datta, H., Borah, A., & Stephen, A. T. (2022). Field of Gold: Scraping Web Data for Marketing Insights. *Journal of Marketing*, 86(5), 1-20.

Haans, R. F. J., & Mertens, M. J. (2025). The Internet Never Forgets: A Four-Step Scraping Tutorial, Codebase, and Database for Longitudinal Organizational Website Data. *Forthcoming in Organizational Research Methods*.

Session 2: Analyzing large-scale data

Method readings:

Choudhury, P., Allen, R. T., & Endres, M. G. (2021). Machine learning for pattern discovery in management research. *Strategic Management Research*, 42(1), 30-57.

Shrestha, Y. R., He, V. F., Puranam, P., & von Krogh, G., (2021). Algorithm Supported Induction for Building Theory: How Can We Use Predictive Models to Theorize? *Organization Science*, 32(3), 856-880.

Discussion papers:

Dahlander, L., Beretta, M., Thomas, A., Kazemi, S., Fenger, M. H. J., & Frederiksen, L. (2023). Weeding out or picking winners in open innovation? Factors driving multi-stage crowd selection on LEGO ideas. *Research Policy*, 52(10), 104875.

Majzoubi, M., & Zhao, E. Y. (2023). Going beyond optimal distinctiveness: Strategic positioning for gaining an audience premium. *Strategic Management Journal*, 44(3), 737-777.

Tidhar, R., & Eisenhardt, K. M. (2020). Get rich or die trying... finding revenue model fit using machine learning and multiple cases. *Strategic Management Journal*, 41(7), 1245-1273.

Optional readings:

Berente, N., Seidel, S., & Safadi, H. (2019). Data-Driven Computationally Intensive Theory Development. *Information Systems Research*, 30(1), 50-64.

Rathje, J., Katiila, R., & Reineke, P. (2024). Making the most of AI and Machine learning in organizations and strategy research: Supervised machine learning, causal inference, and matching models. *Strategic Management Journal*, 45(10), 1926-1953.

Session 3: Analyzing textual data

Method readings:

Aceves, P., & Evans, J. A. (2024). Mobilizing Conceptual Spaces: How Word Embedding Models Can Inform Measurement and Theory Within Organization Science. *Organization Science*, 35(3), 788-814.

Hannigan, T. R., Haans, R. F. J., Vakili, K., Tchalian, H., Glaser, V. L., Wang, M. S., Kaplan, S., & Jennings, P. D. (2019). Topic Modeling in Management Research: Rendering New Theory from Textual Data. *Academy of Management Annals*, 13(2), 586-632.

Discussion papers:

Aranda, A. M., Vaara, E., Etchanchu, H., & Guyt, J. N. (2025). Discursive Struggles and Contested Stigma Extensions: Explaining the Gradual Stigmatization of the U.S. Tobacco Industry. *Forthcoming in Organization Science*.

Corritore, M., Goldberg, A., & Srivastava, S. B. (2020). Duality in Diversity: How Intrapersonal and Interpersonal Cultural Heterogeneity Relate to Firm Performance. *Administrative Science Quarterly*, 65(2), 359-394.

Lix, K., Goldberg, A., Srivastava S. B., & Valentine, M. A. (2022). Aligning Differences: Discursive Diversity and Team Performance. *Management Science*, 68(11), 8430-8448.

Optional readings:

- Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2022). Text preprocessing for text mining in organizational research: Review and recommendations. *Organizational Research Methods*, 25(1), 114-146.
- Miric, M., Jia, N., & Huang, K. G. (2023). Using supervised machine learning for large-scale classification in management research: The case for identifying artificial intelligence patents. *Strategic Management Journal*, 44(2), 491-519.
- Valtonen, L., Mäkinen, S. J., & Kirjavainen, J. (2024). Advancing reproducibility and accountability of unsupervised machine learning in text mining: Importance of transparency in reporting preprocessing and algorithm selection. *Organizational Research Methods*, 27(1), 88-113.

Session 4: Analyzing audiovisual data

Method readings:

- Luo, X., Jia, N., Ouyang, E., & Fang, Z. (2024). Introducing machine-learning-based data fusion methods for multimodal data: An Application of measuring trustworthiness of microenterprises. *Strategic Management Journal*, 45(8), 1597-1629.

Discussion papers:

- Banerjee, M., Cole, B. M., & Ingram, P. (2023). Distinctiveness for What and from Whom? Deep Learning-Based Product Distinctiveness, Social Structure, and Third-Party Certifications. *Academy of Management Journal*, 66(4), 1016-1041.
- Choudhury, P., Wang, D., Carlson, N. A., & Khanna, T. (2019). Machine learning approaches to facial and text analysis: Discovering CEO oral communication styles. *Strategic Management Journal*, 40(11), 1705-1732.
- Zhang, S., Lee, D., Singh, P. V., & Srinivasan, K. (2022). What Makes a Good Image? Airbnb Demand Analytics Leveraging Interpretable Image Features. *Management Science*, 68(8), 5644-5666.

Optional readings:

- Hildebrand, C., Efthymiou, F., Busquet, F., Hampton, W. H., Hoffman, D. L., & Novak, T. P. (2020). Voice analytics in business research: Conceptual foundations, acoustic feature extraction, and applications. *Journal of Business Research*, 121, 364-374.
- Lehmann-Willenbrock, N., & Hung, H. (2024). A multimodal social signal processing approach to team interactions. *Organizational Research Methods*, 27(3), 477-515.

Shrestha, Y. & He, V. F. (2022). Integrating Multimodal Data and Machine Learning for Entrepreneurship Research. *Working paper available from SSRN*: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4204889.

Session 5: Generative AI and the publication process

Discussion papers:

Bechky, B. A., & Davis, G. F. (2025). Resisting the Algorithmic Management of Science: Craft and Community After Generative AI. *Administrative Science Quarterly*, 70(1), 1-22.

Tranchoero, M., Brenninkmeijer, C.- F., Murugan, A., & Nagaraj, A. (2024). Theorizing with Large Language Models. *NBER Working Paper No. 33033*: <https://www.nber.org/papers/w33033>.