DEVELOPING A DATALAB FOR LOGISTIC DATA ANALYSIS AT THE ROTTERDAM UNIVERSITY OF APPLIED SCIENCES

K. M. Paardenkooper, Rotterdam University of Applied Sciences (CoE HRTech)

R.G. Hoogendoorn, Rotterdam University of Applied Sciences (CoE HRTech)

Abstract

This paper explores the role of data labs at universities of applied sciences and addresses the central research question: How can a data lab for logistics data analysis be effectively established as part of a larger ecosystem within and beyond a university of applied sciences? The study provides valuable insights that can guide similar initiatives in data lab development. Research indicates that data labs are essential across various domains and levels to facilitate interdisciplinary collaboration. These labs enable the collection, storage, and analysis of data using Machine Learning and AI models and form a part of learning communities. Among the 12 major universities of applied sciences in the Netherlands, four currently operate data labs. The Data Lab at the Amsterdam University of Applied Sciences is considered a best practice model. At the Rotterdam University of Applied Sciences (RUAS), there are four data labs, with two still in development—one of which focuses on logistics data analysis. To ensure consistency and adaptability, RUAS has established central guidelines that can be tailored to each data lab's needs. The study concludes that the Data Lab for Logistics Data Analysis should be developed by adapting the five-step plan established at RUAS. Critical components of this strategy include choosing appropriate tools such as Jupyter Notebook and data containerization, utilizing Research Drive for secure data storage, connecting to public platforms and GitHub for open collaboration, applying the FAIR principles to ensure data accessibility and reusability, and deploying EduBadges to facilitate learning and credentialing. By following these structured steps, universities can create robust, efficient, and collaborative data labs that support both research and education in logistics data analysis.

1. Introduction

Data analysis is becoming increasingly vital in logistics, just as it is in many other fields. The rapid growth of available data, generated by logistics systems, combined with advancements in technology, has made information exchange easier than ever. Effectively collecting, structuring, cleaning, and analysing this data allows for the use of Machine Learning and Artificial Intelligence (AI) to optimize local logistics processes and entire supply chains (Farahbod et al., 2024; Liu et al., 2023). These applications address key challenges in the logistics sector, such as inefficient collaboration, excessive emissions, and labour shortages. Universities of applied sciences play a crucial role in developing data-driven solutions and equipping students with essential digital skills (Battal, 2023). Data labs are at the heart of these efforts.

The term "data lab" can have multiple meanings, ranging from software tools for data analysis to datasets that can be examined using specialized tools. In the Netherlands, the government has launched 51 data lab initiatives to identify societal trends and challenges (Digitale overheid, 2025). Additionally, several commercial companies provide data lab services, including DataCamp, Datalab, and DGP Media as they see potential profit in helping companies with data collection and analysis (DataCamp, 2025; Datalab, 2025; DGP Media, 2025). Large corporations, such as Unilever, also operate their own data labs (Unilever, 2025).

This paper focuses on data labs within an educational and research context. At Rotterdam University of Applied Sciences (RUAS), data labs are defined as "both physical and digital environments where data and AI can be experimented with and used in a safe and responsible manner" (RUAS, 2025a). A data lab serves in its physical form, on the one hand as a space where data is ingested, stored, and analysed. On the other hand it is a place where students, researchers and people from the industry can work together on data science issues assisted by data science professionals. Furthermore, data labs form a part of learning communities and larger ecosystems.

At RUAS, the Artificial Intelligence and Logistics research group within the Centre of Expertise HRTech is developing a dedicated logistics data lab. This initiative is part of a broader effort to establish data labs across different faculties within RUAS. The program AI & Ethics, led by the research centre Creating 010, has developed an initial blueprint for these data labs. Their implementation is guided and funded by the university’s strategic agenda for digitalization (RUAS, 2023). The vision is to create an interconnected ecosystem of data labs within RUAS that also links to external data labs. However, realizing this vision presents several challenges, as each individual data lab must navigate key decisions regarding its purpose, expected data volume, hosting, data ownership, accessibility, continuity, scalability, and cybersecurity.

This paper situates logistics data labs within a broader scientific and practical context and explores their development. The central research question is: How can a data lab for logistics data analysis be effectively established as part of a larger ecosystem within and beyond a university of applied sciences? The paper contributes valuable insights into data lab development that may benefit other initiatives.

The remainder of this paper is structured as follows: Section 2 provides an overview of the state of the art in data labs based on existing literature. Section 3 examines data labs at various universities of applied sciences in the Netherlands. Section 4 delves into the development of data labs at RUAS, followed by Section 5 with a focus on the trajectory of the logistics data lab. Finally, Section 6 presents the conclusions and key takeaways.

2. Data labs in the literature

Seven publications discuss the setup and use of data labs for educational purposes. These studies address common themes, including the significance of data labs across various domains and user groups, the role of interdisciplinarity, the varying levels of users' technical expertise, architectural considerations such as applying Jupyter Notebook and data containerization, the use of GitHub repositories, and the importance of the the FAIR principles.

The literature highlights the necessity of data labs in business, research, and education. Berkaoui and Gahi (2023) propose a data lab for business applications, enabling companies to leverage machine learning algorithms to extract value from diverse, non-conventional data sources. Campbell et al. (2001) advocate for an educational and research-oriented data lab focused on signal processing. Hollaway et al. (2020) introduce a Data Science Lab for environmental data science, facilitating big data analysis, result storage, and dissemination. Lopez-Pernas et al. (2022) present a Data Virtual Lab designed to train personnel in data analysis through a hands-on approach at multiple institutions. Wallum et al. (2023) describe a data lab for early undergraduate chemistry education, using self-generated data. Additionally, Xiao et al. (2022) and Zhang et al. (2016) both describe data labs designed for academic research: the former introduces a unified, data-oriented DATALab for Natural Language Processing, while the latter presents a DataLab for managing big data workflows, version data management, and analytics (Berkaoui & Gahi, 2023; Campbell et al., 2001; Hollaway et al., 2020; Lopez-Pernas et al., 2022; Wallum et al., 2023; Xiao et al., 2022; Zhang et al., 2016).

Except for the dedicated academic labs and the one with a business focus, interdisciplinarity is a key theme. Campbell et al. (2001) highlight the need for collaboration between software and electrical engineering students, whereas Hollaway et al. (2020) stress the importance of cooperation among scientists from various environmental disciplines, stakeholders, and policymakers. User expertise levels also can vary significantly. Campbell et al. note differences in students' mathematical proficiency, while Hollaway et al. observe disparities in computer science knowledge. To address these challenges, user groups are categorized based on technical skills. Lopez-Pernas et al. (2022) distinguish between technically skilled and unskilled users, as data scientists and engineers collaborate with managers lacking programming knowledge.

These differences in the knowledge and experience with computer science influence data lab architecture, which makes accessibility is a critical consideration. Academic labs described by Campbell et al. (2001), Xiao et al. (2022), and Zhang et al. (2016) employ complex, self-programmed setups.

Conversely, Berkaoui et al. (2023), Hollaway et al. (2020), and Wallum et al. (2023) advocate for Jupyter Notebook, a web-based application compatible with programming languages Python and R, which accommodates users with varying programming experience. For inexperienced users, as in Wallum's study, pre-programmed applications can be deployed. Lopez-Pernas et al. (2022) find Jupyter Notebook too complex and recommend an alternative application.

In order to provide secure working spaces for users with different level of knowledge and experience, several studies support containerized data architectures. Berkaoui et al. (2023), Hollaway et al. (2020), and Zhang et al. (2016) propose data containers, isolated workspaces with specific software packages, allowing users to manipulate data copies without risking source data corruption. Additionally, containerization enhances scalability, with Docker, a commonly used sortware platform for testing and deploying applications.

To guarantee the reusability of the applications, especially or educational purposes, the use of GitHub repositories is emphasized by four of the seven papers (Hollaway et al., 2020; Lopez-Pernas et al., 2022; Wallum et al., 2023; Xiao et al., 2022). These repositories facilitate the storage and sharing of open-source applications, aligning with the principles of open science and the FAIR (Findable, Accessible, Interoperable, Reusable) framework, as highlighted by Hollaway et al. (2020) and Xiao et al. (2022).

3. Data labs at universities of applied sciences

From the 12 major universities of applied sciences in the Netherlands, five have established data labs: NHL Stenden, Saxion, Zeeland, Amsterdam, and Rotterdam universities of applied sciences. As the Amsterdam University of Applied Sciences has the most labs at the end of this section it is described more in detail as a best practice.

The mentioned universities of applied sciences have furnished their data labs according to their educational and research needs. One of the data labs of NHL Stenden in Leeuwarden supports the Artificial Intelligence for Image Data minor, focusing on image mining, processing, and pattern recognition using machine learning (NHL Stenden, 2025a). The second, the Maritime Hacking Lab, has a research focus, it is operated by the Maritime IT Security research group and is used for vulnerability analyses of maritime equipment (NHL Stenden, 2025b). Similarly, the DAta + AI lab at Saxion University of Applied Sciences, is designed for research, processing large quantities of data. It utilizes a data server and a high-performance computing system for machine and deep learning-based analyses. The Data + AI Lab is part of the broader Ambient Intelligence Lab, which also includes the IoT and XR Labs (Saxion, 2025). Zeeland University of Applied Sciences participates in the Data Science Lab, located within the Joint Research Center Zeeland (JRCZ). This lab specializes in data visualization and related applications (JRCZ, 2025).

The Amsterdam University of Applied Sciences (HvA) established its Data Lab five years ago as part of a network of 12 specialized labs that actively collaborate and share a common ICT infrastructure (HvA, 2025). This facility supports a wide range of research projects, from small-scale studies that can be processed on a laptop to complex, high-volume data analyses. When a project surpasses the lab's capacity, it transferred to the university's ICT service for further handling (Bons, 2025).

Designed to provide a secure and powerful ICT infrastructure, the Data Lab is used by lecturers, researchers, and students for conducting advanced quantitative analyses (HvA, 2025). It plays a crucial role in technology education, particularly benefiting students in fields such as mathematics, electrical engineering, automated systems, architecture, and logistics. Students receive training at various levels, from basic data handling and document management to mastering analytical tools and developing custom programming solutions (Bons, 2025).

The Data Lab is powered by 10 Linux servers running Ubuntu, each configured to support multiple users. The decision to use open-source technology ensures that data remains secure, local, and confidential. Jupyter Notebook is the primary coding environment, which is widely adopted for its versatility and ease of use (Brugge & Metseler, 2025).

The system administrators oversee the lab's operations, create user accounts and workspaces, and install necessary software packages. Work environments are customized based on the requirements of lecturers, ensuring that students and researchers have access to the tools they need for their projects. The lab can be accessed locally, where administrators provide hands-on support, or remotely via the university's VPN connection (Brugge & Metseler, 2025).

4. Data labs at the RUAS

Currently, Rotterdam University of Applied Sciences has four data labs at different stages of development: Datalab Rotterdam, HR Datalab Healthcare, HR Datalab EAS, and IGO Datahub. Additionally, data labs for the economic domain and logistics analysis are still under development. All these data labs are closely linked to educational programs (RUAS, 2025a).

Overview of RUAS Data Labs:

- Datalab Rotterdam is part of the Institute for Communication, Media, and Information Technology (CMI). It features high-performance servers suited for advanced data analysis, including large language models. Its primary focus is education, allowing students to experiment with complex AI applications in a safe and scalable environment.
- HR Datalab Healthcare supports research-intensive projects, facilitating professional development for lecturers, researchers, and students in AI and data-driven healthcare applications.
- HR Datalab EAS, affiliated with the Institute for Engineering and Applied Science, officially launched in 2025.
- IGO Digihub, an initiative from the Institute for the Built Environment, is still in its early development phase.

The data labs for the economic domain and for logistics analysis are still in development.

The AI & Ethics program at RUAS has developed an initial blueprint for the hardware, access protocols, system architecture, data storage, and training. The infrastructure includes specified workstations, servers, storage devices, and display screens, with detailed guidelines for both software and hardware configurations.

For data processing and analysis, the system primarily uses Jupyter Notebook, which facilitates data cleaning, transformation, simulation, statistical modelling, visualization, and machine learning. Essential software libraries are installed by IT professionals or researchers, with Docker Images used for additional installations. This container-based system allows manual integration of software components, enhancing flexibility and efficiency (AI & Ethiek, 2025).

To ensure secure data storage and sharing, RUAS utilizes Research Drive, with clearly defined responsibilities for data management. Various plug-ins support data analysis, and the labs are integrated with Anaconda, a widely used open-source platform for scientific computing. The system also connects with public cloud platforms such as Azure, AWS, SURF, and Jupyter Hub. To facilitate collaboration and dissemination of research outputs, GitHub repositories serve as a central hub for RUAS's Data Science Community of Practice (CoP). This platform enables lecturers, researchers, and students to collaborate on interdisciplinary projects, all within an open-source framework (RUAS, 2025b).

The Data lab can be acessed via a lent SSD (flash drive), which requires a deposit. These SSDs come pre-configured with the necessary operating system and connect via USB-C. Users can book time slots through a TEAMS channel, ensuring that multiple users can access the system simultaneously with a standardized configuration, reducing maintenance and enhancing scalability.

Before gaining access, users must complete training on hardware usage, ethical data handling, and AI best practices in accordance with FAIR principles. Additional training is offered in Natural Language Processing, Computer Vision, and Robotics, with participants earning Digital (Open) Badges—secure digital credentials that verify specific skills and competencies. These badges are stored on a digital platform, providing evidence of learning and achievement.

The development of the RUAS data labs is supported by the university's strategic agenda for digitalization (RUAS, 2023). While each data lab is embedded within its respective educational or research institute and follows its own development path, best practices are shared to facilitate broader implementation across the university.

5. Datalab for logistic data analysis

A data lab for logistics data analysis is currently under development at the Rotterdam University of Applied Sciences (RUAS) as part of the Artificial Intelligence and Logistics lectorate within CoE HRTech. This initiative serves two primary purposes: providing a data hub for high-impact projects and functioning as a research and education facility for AI and machine learning applications in logistics.

One of the lab's key roles is supporting major research initiatives such as PATH2ZERO and PEOPLE. PATH2ZERO (Paving the Way Towards Zero-Emission and Robust Inland Shipping) aims to revolutionize inland shipping by developing zero-emission solutions and sustainable business models. A central deliverable is a data hub that collects, stores, and analyses project-generated data while hosting a digital twin of inland shipping operations between Rotterdam and Duisburg (NWO, 2022).

Similarly, the PEOPLE project (Power to Energy Transition and Organizing Port Learning Experiences) focuses on the energy transition in the Port of Rotterdam, with an emphasis on workforce development. One of its main objectives is creating a simulation and data platform for educational use, including a digital twin of a wind turbine at RDM (Katapult, 2025).

Beyond supporting research projects, the data lab will serve as a dedicated space for AI and machine learning applications in logistics. It will be utilized by both researchers and students across related institutes, fostering innovation in automation and data-driven decision-making.

The new logistics data lab will be housed at RDM Innovation Dock, alongside the PEOPLE project's Hydrogen Lab. Designed to accommodate around 20 users on-site, it will also offer remote accessibility. Development will adhere to university guidelines and align with other data labs, necessitating a full-time system administrator for maintenance and support. Following and adapting the AI & Ethics (AI & Ethiek, 2025) guidelines, the lab's setup follows a structured five-step process:

1. Hardware Acquisition and Setup: Purchasing, installing, and testing equipment.
2. Cloud Integration: Establishing public cloud access through Microsoft Azure.
3. System Connectivity: Linking the lab to Jupyter-HUB via Surf.
4. Educational Integration: Deploying EduBadges to facilitate training and credentialing.
5. Certification and Cybersecurity: Obtaining certifications for Information Privacy and Ethical Emerging Technology while ensuring robust cybersecurity measures.

With this structured approach, the Data Lab for Logistics Data Analysis is poised to become a vital asset in advancing AI-driven logistics, sustainability, and education.

6. Conclusion

Data labs are becoming essential in universities of applied sciences, enabling interdisciplinary collaboration in collecting, storing, and analysing data using Machine Learning and AI models. This paper examines how a data lab for logistics data analysis can be effectively established within a broader academic and research ecosystem. Across various domains, data labs are playing a crucial role in advancing research and education. Key decisions in their development include choosing tools like Jupyter Notebook, data containerization, GitHub repositories, and following FAIR principles (Findable, Accessible, Interoperable, and Reusable). In the Netherlands, four out of the 12 major universities of applied sciences currently operate data labs. Among them, the University of Applied Sciences of Amsterdam stands out as a best-practice example. The university's choice to use an open-source LINUX system ensures privacy and security, avoiding potential data risks associated with proprietary software like Microsoft applications.

At Rotterdam University of Applied Sciences (RUAS), four data labs are operational, with two more under development—including the data lab for logistics data analysis. However, a limitation of this research is that it relied on university websites, meaning there could be additional data labs not easily found online. RUAS has developed a set of central guidelines for its data labs, which can be adapted to specific needs. The primary technological choices made by RUAS include Jupyter Notebook for coding and data analysis, data containerization using Docker Images, research Drive for secure data storage, integration with public platforms and GitHub, access through SSD drives, implementation of FAIR principles and the deployment of EduBadges for certification and education.

The main goal of the logistics data lab is to support data acquisition, storage, analysis, and digital twin development for major subsidized projects such as PATH2ZERO and PEOPLE. Additionally, it will serve as a hub for research and education in machine learning and AI applications. The logistics data lab at RUAS will be built following an adapted five-step approach, hardware acquisition – Procuring and setting up necessary infrastructure, cloud integration – ensuring secure and scalable cloud access, system connectivity – Establishing links with Jupyter-HUB and other tools, educational integration – deploying EduBadges for structured learning and certification & cybersecurity – ensuring compliance with privacy and security standards.

The answer to the central research question is: The data lab for logistics data analysis at RUAS will be developed by adapting the university's structured five-step plan. This includes using Jupyter Notebook, implementing data containerization, integrating research drive for data storage, connecting to public platforms and GitHub, adhering to FAIR principles, and deploying EduBadges. With these strategies in place, the lab will provide a secure, scalable, and collaborative environment for logistics research and education.

Bibliography

AI & Ethiek. (2025). *Generic data-hub set-up guide*.

Battal, M. (2023). Artificial Intelligence Integration of Large Logistics Companies. *Journal of International Trade, Logistics and Law*, *9*(2), 273-281. Artificial Intelligence Integration of Large Logistics Companies – DOAJ

Berkaoui, A., & Gahi, Y. (2023). A private cloud-based Datalab for scalable DSML pipelines. *Proceedings of the 6th International Conference on Networking, Intelligent Systems & Security*, 1. https://doi.org/10.1145/3607720.3607744

Bons, P. (2025, 27 January 2025). *Coordinator Data Lab* [Interview].

Brugge, N. v. d., & Metseler, D. (2025, 6 February 2025). *System Administrators Data Lab* [Interview].

Campbell, J., Murtagh, F., & Kokuer, M. (2001). DataLab-J: a signal and image processing laboratory for teaching and research. *IEEE Transactions on Education*, *44*(4). https://doi.org/10.1109/13.965780

DataCamp. (2025). *Getting insights from your data shouldn't be this hard…*. Retrieved 28 january 2025 from https://www.datacamp.com/datalab

Datalab. (2025). *Datalab*. Retrieved 28 January 2025 from https://www.datalab.nl/

DGP Media. (2025). *Bereik je doelgroep*. Retrieved 28 January 2025 from https://www.dpgmediagroup.com/nl-NL/adverteren/services/datalab?referrer=https%3A%2F%2Fwww.google.com%2F

Digitale overheid. (2025). *51 initiatieven voor Data lab*. Retrieved 28 January 2025 from https://www.digitaleoverheid.nl/initiatieftype/datalab/

Farahbod, K., Varzandeh, J., & Shayo, C. (2024). ARTIFICIAL INTELLIGENCE AND ITS PERSPECTIVE IMPACTS ON SUPPLY CHAINS AND LOGISTICS. *Journal of Business and Behavioral Sciences*, *36*(1), 113-124. JBBS_36.1_Spring_2024.pdf

Hollaway, M. J., Dean, G., Blair, G. S., Brown, M., Henrys, P. A., & Watkins, J. (2020). Tackling the Challenges of 21st-Century Open Science and Beyond: A Data Science Lab Approach. *Patterns*, *1*(7). https://doi.org/10.1016/j.patter.2020.100103

HvA. (2025). *Kijkje in de keuken van de datalab*. Retrieved 23 January from https://www.hva.nl/kc-techniek/gedeelde-content/nieuws/nieuwsberichten/2024/01/kijkje-in-de-keuken-bij-het-datalab.html

JRCZ. (2025). *Visualize Research Data*. Retrieved 23 January from https://jrcz.nl/en/data-science-lab.html

Katapult. (2025). *PEOPLE*. https://netwerk.wijzijnkatapult.nl/detail/1359/people/groeifonds/

Liu, Y., Tao, X., Li, X., Colombo, A. W., & Hu, S. (2023). Artificial Intelligence in Smart Logistics Cyber-Physical Systems: State-of-The-Arts and Potential Applications. *IEEE Transactions on Industrial Cyber-Physical Systems*, *1*. https://doi.org/10.1109/TICPS.2023.3283230

Lopez-Pernas, S., Munoz-Arcentales, A., Aparicio, C., Barra, E., Gordillo, A., Salvachua, J., Quemada, J., Santos, B. S., & Alford, G. (2022). Educational Data Virtual Lab: Connecting the Dots Between Data Visualization and Analysis. *IEEE computer graphics and applications*, *42*(5), 76-83. https://doi.org/10.1109/MCG.2022.3189557

NHL Stenden. (2025a). *I'M A.I. - Artificial Intelligence for Image Data*. Retrieved 23 January from https://www.nhlstenden.com/minoren/im-ai-artificial-intelligence-for-image-data?_gl=1*1p0mp8n*_up*MQ..&gclid=Cj0KCQiAy8K8BhCZARIsAKJ8sfQv-N1J1CHomUO1gI2ogAFUYskwzCY07rqP_GlSQDKC-TlIUPsJyL4aAsvyEALw_wcB

NHL Stenden. (2025b). *Lectoraat Maritime IT Security*. Retrieved 23 January from

NWO. (2022). *Onderzoek ondersteunt transitie naar emissievrije binnenvaart*. Retrieved 22 January from https://www.nwo.nl/nieuws/onderzoek-ondersteunt-transitie-naar-emissievrije-binnenvaart

RUAS. (2023). *Talent voor transitie Hoger beroepsonderwijs en praktijkgericht onderzoek voor de samenleving van morgen*. Retrieved 4 Feruary 2025 from https://hint.hr.nl/globalassets/hr-brede-themas/hogeschool/koers/strategische-agenda-2024-2028/strategische-agenda-2023-2028-nl.pdf

RUAS. (2025a). *Datalabs*. Retrieved 29 January 2025 from https://hint.hr.nl/nl/HR/Voorzieningen--Services/AVICT/datalabs/?ticket=ST-6218411-hfy7hZQU9FZQ6xasfsiGdiDlQeAbecC3Y1z-20

RUAS. (2025b). *Repositories*. Retrieved 27 January from https://github.com/orgs/HR-DATA-FABRIC/repositories

Saxion. (2025). *Ambient Intelligence lab*. Retrieved 23 January from https://www.saxion.nl/onderzoek/labs/ambient-intelligence-lab

Unilever. (2025). *DataLab*. Retrieved 28 January 2025 from https://www.unilever.com/our-company/innovation-and-rd-at-unilever/datalab/

Wallum, A., Liu, Z., Lee, J., Chatterjee, S., Tauzin, L., Barr, C. D., Browne, A., Landes, C. F., Nicely, A. L., & Gruebele, M. (2023). An Instrument Assembly and Data Science Lab for Early Undergraduate Education. *Journal of Chemical Education*, *100*(5), 1866-1876. https://doi.org/10.1021/acs.jchemed.2c01072

Xiao, Y., Fu, J., Yuan, W., Viswanathan, V., Liu, Z., Liu, Y., Neubig, G., & Liu, P. (2022). DataLab: A Platform for Data Analysis and Intervention. Dublin, Ireland.10.18653/v1/2022.acl-demo.18

Zhang, Y., Xu, F., Frise, E., Wu, S., Yu, B., & Xu, W. (2016, 16-16 May 2016). DataLab: A Version Data Management and Analytics System. 2016 IEEE/ACM 2nd International Workshop on Big Data Software Engineering (BIGDSE), DataLab: A Version Data Management and Analytics System