

DataScEd4CiEn: Integrating data science into STEAM education for civic engagement and social justice – A case from Greece

Dionysia Bakogianni¹ and Ioannis Malafekas²

¹Karlstad University, Sweden;

²National and Kapodistrian University of Athens, Greece

dionysia.bakogianni@kau.se

This paper explores the implementation of a data science activity within a STEAM scenario on food waste and sustainability, carried out in a Greek lower secondary school as part of the Erasmus+ project DataScEd4CiEn. The Greek case offers a unique lens due to the traditionally formal and discipline-bound structure of the national curriculum, where data literacy and interdisciplinary approaches are still emerging and not yet systematically embedded. Through an iterative design process, students engaged with real-world datasets to explore social issues and propose civic actions. The findings reveal both the emergence of key data science practices as well as the challenges involved in supporting open-ended inquiry and integrating data insights into civic messaging. The paper also highlights broader pedagogical tensions in teaching data science in schools and emphasizes the strong potential of STEAM scenarios to promote social engagement through data-driven inquiry.

INTRODUCTION

In today's digitized and data-rich world, data has become central to how individuals interact, make decisions, and understand their surroundings. From health tracking and algorithmic recommendations to civic technologies and social platforms, data pervades modern life (Makar, Fry, & English, 2023, Ridgway, 2022). Navigating this environment requires not only technical proficiency but also critical dispositions to question, interpret, and act upon data in socially responsible ways (Bargagliotti, et.al., 2020, Burril & Pfannkuch, 2023, Noll, et.al., 2022; Gould, 2017).

This broader vision of data literacy has redefined the role of data science education in schools. Beyond traditional statistics, data science activity involves engaging with complex, often messy data to pose questions, identify patterns, build models, and communicate findings (Fielding & Makar, 2022, Zucker, et.al. 2020, Ridgway, 2022). When rooted in authentic contexts, such activity can empower students as informed inquirers and civic participants (Makar et al., 2023, Kahne & Bowyer, 2017, Engel, 2017). STEAM approaches, which integrate science, technology, engineering, arts, and mathematics, offer fertile ground for embedding data science in real-world problem-solving. Within such interdisciplinary frameworks, students can explore issues of equity, sustainability, and democratic participation through data (Friedrich, et.al., 2024, Schreiter, et.al., 2024). The DataScEd4CiEn Erasmus+ project builds on this vision by supporting the integration of data science into school education via civic engagement scenarios. It emphasizes student-led inquiry, critical thinking, and social relevance while offering tools and resources for teachers across Europe.

The project spans five countries with diverse educational traditions regarding STEAM and data literacy. The case of Greece is particularly distinctive, as data science and statistics education are still in their early stages of development across school curriculum. This lack of systemic support makes it challenging for teachers to design meaningful data-driven learning experiences within socio-scientific contexts. This paper addresses this gap by examining the iterative transformation of a data science activity designed as part of a broader STEAM scenario by a team of twelve Greek secondary school teachers in the context of the DataScEd4CiEn project. While the scenario was collaboratively developed and implemented, our focus lies on the mathematics teacher, whose decisions and reflections allow us to trace how data science activity was shaped in practice. The study explores how the activity evolved through redesign and teacher reflection, facilitated by the project's research team. Our aim is to understand the potential of teacher-led innovation in national settings that lack familiarity with statistical literacy, reasoning, and exploratory data practices.

We address two research questions:

1. How are various dimensions and practices of data science activity embedded and supported in a STEAM scenario designed to foster civic engagement?

2. What pedagogical aspects of teaching data science are reflected in the transformation of this learning activity in the Greek context?

THEORETICAL FRAMEWORK

In recent years, there has been a growing call to rethink the role of data in education, not simply as a mathematical object, but as a medium through which students can explore authentic, complex, and socially relevant questions. As argued by Meletiou-Mavrotheris et al. (2025), engaging students with real data sources, rather than static textbook datasets, can support deeper inquiry and foster civic thinking. This vision aligns with the work of Makar et al. (2023), who frame data science education as a form of critical citizenship, emphasizing four interconnected lenses: political (systemic understanding), social (data's societal role), self (personal reflection), and praxis (informed action). Within such a framing, data science becomes not only technical, but also ethical and political, encouraging students to investigate real-world problems and communicate their findings meaningfully.

STEAM education provides a fertile ground for embedding this kind of data science activity. By integrating science, technology, engineering, arts, and mathematics, STEAM encourages transdisciplinary inquiry around real issues, such as sustainability or digital wellbeing (e.g. Makar, et.al., 2023, Liston, et.al., 2022). The Erasmus+ project DataScEd4CiEn builds on this potential, promoting student-led exploration of “worthy problems,” collaboration with real data, and teacher-designed scenarios that connect STEAM content with civic engagement. The project responds to pressing challenges, especially in countries like Greece, where data literacy remains underdeveloped across formal education.



Figure 1. The DataScEd4CiEn project's conceptual framework.

To conceptualize and analyze how data science unfolds in school settings, the DataSETUP framework (Kazak, Leavy & Podworny, 2025) offers a dual-dimensional perspective:

The first dimension describes the process of doing data science through four iterative phases: *Getting and Exploring Data*, *Formulating Questions*, *Modeling and Analyzing*, and *Communicating Results and Acting*. These are supported by nine core data science practices: data preparation and processing, big data technologies, developing statistical questions, data governance and ethics, exploratory data analysis, data visualization, computational thinking, modeling, collaboration and communication. These practices help students move beyond static data use to inquiry, interpretation, and action.

The second dimension, highlights considerations essential for teaching meaningfully data science.

These include:

- (1) Context, grounding activities in real-world scenarios that matter to students;
- (2) Key Ideas, drawing from core statistical concepts like variability, correlation, and modeling;
- (3) Assessment, emphasizing student reasoning and project-based evaluation; and
- (4) Tools, such as CODAP, which support interactive data exploration and communication.

Together, these two dimensions help educators design and analyze data science learning not just in terms of what students do with data, but also in terms of how instructional design supports that work.

In this study, the DataSETUP framework is used to investigate how a STEAM scenario designed in Greece evolved to integrate data science in meaningful and civic-oriented ways.

METHODOLOGY

This study adopts a qualitative case study approach (Yin, 2009), to investigate how pedagogical innovations unfold in situated contexts. The study is embedded within the broader DataScEd4CiEn project. The project employs a design-based research (DBR) methodology (Cobb et al., 2003) to iteratively develop professional development (PD) resources and classroom scenarios. The present analysis focuses on the Greek pilot, implemented during the 2024–2025 school year, a context where data science and data literacy are largely absent from formal education.

Twelve teachers from various disciplines participated in the PD and co-designed STEAM scenarios for implementation in a lower secondary school in Athens. The scenario centered on food waste and environmental justice, combining real-world data analysis with student-driven inquiry and civic reflection. The implementation involved 20 students (aged 12–14) who had previous experience in project development and showed a reasonably solid level of familiarity with digital tools and technologies. Students, concluded the project by designing a public awareness poster to be displayed in a local supermarket, aimed at informing and engaging their community. In this paper, we focus on the part of the implementation that involved the data science inquiry led by the mathematics teacher. This component consisted of two two-hour teaching sessions, within a broader STEAM scenario that spanned a total of ten didactical hours.

Data used and analyzed here included: (a) two video-recorded lessons (totaling four teaching hours), (b) three student worksheets, and (c) three audio-recorded meetings with the teacher and the project's researcher (see first author). The implementation of the learning activity unfolded across three distinct phases: the initial design of the worksheet, followed by a first and a second two-hour classroom implementation. The initial design was discussed in depth both among the participating teachers and in collaboration with the project researchers. Each implementation phase was followed by a reflective conversation between the mathematics teacher and the first author (serving as the project's researcher), allowing for iterative refinement of the activity.

Data analysis was guided by the DataSETUP framework (Kazak, Leavy, & Podworny, 2025) and proceeded in two phases. The first focused on how core data science processes and practices were enacted across the three implementation stages. This phase drew primarily on the student worksheets, supported by the classroom video recordings. It provided insight into how data science activity evolved across design and practice. The second phase examined pedagogical factors, such as context, assessment, and tool use, that shaped the design's evolution and its alignment with civic engagement goals. This phase was based on the audio-recorded teacher meetings and was guided by the four pedagogical dimensions of the framework, taking into account both general teaching considerations and features of the specific national setting.

RESULTS

In this section, we present findings organized according to the two dimensions of the DataSETUP framework: (1) data science processes and practices, and (2) pedagogical considerations that shaped the activity's design and implementation.

Dimension 1: Data Science Processes and Practices.

Below, we present how the four core data science processes—Getting and Exploring Data, Formulating Questions, Modeling and Analyzing, and Communicating Results and Acting—were enacted through the worksheets designed by the mathematics teacher. Drawing on the student materials and classroom observations, we discuss how specific data science practices emerged within each process and how their implementation evolved across the three phases of the scenario: initial design, first implementation, and second implementation.

Getting and Exploring Data

This phase was consistently activated across all three versions. Students worked with real-world datasets from sources such as Eurostat, FAO, and ReFED, and used digital tools (e.g., spreadsheets, CODAP) to view and compare data. However, the exploration was largely guided through structured worksheets. In

the first implementation, students engaged in basic summaries and visualizations; in the second, CODAP enabled more flexible filtering and navigation through multivariate data.

This phase was consistently present across all three versions of the activity. Students engaged with real-world datasets from sources such as Eurostat, FAO, and ReFED, using digital tools (e.g., spreadsheets and CODAP) to view and compare data. However, their exploration was primarily guided by structured worksheets. In the first implementation, students performed basic summaries and created simple visualizations. In the second, CODAP supported more flexible filtering and navigation of multivariate data.

Formulating Questions

This proved to be the most challenging phase. Although students were prompted to reflect on data patterns, the tasks remained statistically framed, with pre-defined questions embedded in the worksheet. Little space was provided for students to pose their own investigative or civic-oriented questions. This limited the openness and interpretive potential typically associated with data science inquiry.

Modeling and Analyzing

Student engagement with analysis practices strengthened in the second implementation, especially with the use of CODAP. Students explored correlations (e.g., food waste vs. CO₂ emissions), created scatterplots and boxplots, and engaged in rudimentary scenario modeling (e.g., impact of 20% waste reduction). Practices like exploratory data analysis, visualization, and computational thinking were visibly supported.

Communicating Results and Acting

The final product, a poster to be displayed in a local supermarket, provided a meaningful civic outlet. Students synthesized their insights and translated them into a public message, showing the potential of data to inform and engage communities. This phase supported communication and framed data science as socially consequential. However, despite having explored multiple and diverse data sources and identifying relevant insights, students did not explicitly incorporate data-based findings into the content of the poster. This highlights a gap between data exploration and public communication, suggesting that further scaffolding may be needed to help students draw on their analyses when engaging in civic messaging.

Overall, students engaged directly with practices such as data preparation, visualization, and interpretation, while engaging indirectly, mainly through structured questions and teacher-led discussion, with statistical questioning, modeling, and ethical considerations. The progression across versions reflects a movement toward deeper and more meaningful engagement with data science, though space for open exploration and student-generated questions remains limited. A summary of how the Data Science Practices were activated across worksheet versions is presented in Table 1 below.

Table 1. Engagement with Data Science Practices Across Worksheet Versions.

Data Science Practice	Initial worksheet	Worksheet 1st session	Worksheet 2nd session
<i>Data preparation and processing</i>	Basic (one source of data, minimal processing)	Moderate (multiple data sources, data source awareness)	High (filtering, new variable creation)
<i>Big data technologies</i>	Not appeared	Not appeared	Data explorations through COPAD
<i>Developing statistical questions</i>	Low (teacher-defined questions)	Moderate (semi structured comparisons)	Moderate (task-embedded prompts)
<i>Data governance and ethics</i>	Implicit (data source not critiqued)	Emerging (source credibility introduced)	Implicit (discussing data sources)
<i>Exploratory data analysis</i>	Moderate (guided summaries and guided comparisons)	Improved (cross-country analysis)	High (dynamic visual exploration, prediction-making)
<i>Data visualizations</i>	Basic graphs	Extended (multiple visualizations)	Advanced (scatterplots, boxplots)

<i>Computational thinking</i>	Not appeared	Emerging (manipulating tables, pattern exploration)	Basic (formula use, variables)
<i>Modeling</i>	Not appeared	Not explicit	Emerging (basic predictions)
<i>Collaboration and communication</i>	Limited (group-based worksheet completion)	Group discussions	Active (peer sharing, plenary discussions)

Dimension 2: Pedagogical Considerations for Teaching Data Science.

The second dimension of the DATASETUP framework emphasizes design elements that support meaningful teaching of data science, including the learning context, core statistical ideas, assessment strategies, and the tools used for exploration and communication. The analysis revealed a number of key insights regarding these dimensions in the context of the Greek implementation.

Context

Several contextual factors shaped the planning and implementation of the activity. The scenario was framed within a STEAM-based civic inquiry focused on food waste and its links to poverty and environmental sustainability. This problem context served as a meaningful entry point for the data science activity. Through their analysis of real datasets from Eurostat, FAO, and ReFED, students reached some striking conclusions, notably, that Greece ranks among the highest in Europe in food waste, and that household consumption patterns are one of the most significant drivers of the problem. These insights highlighted the individual responsibility in addressing a collective issue, aligning well with the civic goals of the scenario.

At the institutional level, several challenges emerged. Although the school had a well-equipped infrastructure and supportive leadership, both essential to the smooth delivery of the scenario, the strict scheduling structure and subject separation of the national curriculum made collaboration across disciplines difficult. Despite the teachers' willingness and experience with projects, they were rarely able to co-design or co-teach, and valuable insights generated in one subject or group were not fully integrated into the broader implementation. This was particularly visible in the final product: although students discovered impactful patterns and civic implications during data exploration, these findings were not widely discussed across the teaching team, nor explicitly represented in the final public awareness poster.

Student and teacher background also played a role. While students lacked formal statistical training, they showed readiness and accuracy in interpreting real-world data. However, teachers' limited exposure to data literacy and exploratory approaches influenced the openness and depth of the inquiry process. The activity revealed that a certain level of technological background was essential on both sides as well: teachers needed digital confidence and familiarity with data tools to design and implement an effective inquiry, while students also relied on basic technological skills to navigate datasets, use exploratory tools like CODAP, and interpret visualizations meaningfully.

Key Ideas in Curriculum

Central to the teacher's planning was introducing students to the variety of data types and sources, including the concept of big data. "I wanted them to see as much data as possible," she noted, though she preferred to provide structured sources rather than open the task fully. Discussions with the researcher often focused on distinguishing data science from statistics, and while the teacher emphasized data diversity, the processes and reasoning of data science were less developed in her planning.

Assessment

Assessment remained an open question, primarily framed through alignment with curriculum standards. The teacher expressed uncertainty about how such exploratory activities could be formally assessed or integrated into existing content: "Some parts fit, like measures of central tendency or maybe prediction, but supporting a full data science activity would require more radical curricular change."

Tools

Three types of tools were most highlighted:

1. Public data repositories, which helped students visualize organized datasets and understand real-world structures.
2. CODAP, described by the teacher as “*powerful but simple*”, enabled variable creation, parameter adjustments, and prediction modeling, even with younger learners.
3. Project materials provided by the PD program were also instrumental: “*I read everything you gave us, explored the sources, did my best... but I still lack familiarity. You can't learn this in just a few hours.*”

Overall, this second dimension helped uncover the pedagogical conditions and constraints that shaped the design, implementation, and potential of the data science activity. It revealed how institutional context, curricular limitations, and teacher readiness intersect with the challenge of supporting meaningful student engagement with data science in STEAM and civic

DISCUSSION AND CONCLUSION

This study addressed two core questions concerning the integration of data science activity within a STEAM scenario aimed at civic engagement. Through iterative analysis of a mathematics teacher’s evolving design, we observed how data science processes and practices became more meaningfully embedded across implementation phases, and how pedagogical decisions shaped both affordances and constraints.

The findings suggest that data science activity holds significant promise, not only for advancing students’ data literacy but also for cultivating their civic awareness and agency. By working with real data, formulating questions, analyzing patterns, and communicating findings, students begin to see data as a tool for social understanding and action. Notably, students engaged with enthusiasm and depth, even in school environments and curriculum settings where the ideas of data literacy are still at an embryonic or developing stage. Their response underscores the potential of data science activity to support meaningful learning and civic thinking, even when students have had limited prior exposure through formal education.

At the same time, the study revealed persistent challenges. Institutional barriers, curricular rigidity, and limited teacher preparation constrained the full realization of both data science and STEAM approaches. The professional development offered by the DataScEd4CiEn project seems to provide an essential entry point for teachers’ initial engagement with data science concepts and tools. However, ensuring the sustainability and long-term expansion of such initiatives requires further systemic support, through ongoing training, curricular integration, and collaborative infrastructure that fosters meaningful, interdisciplinary, and socially relevant data inquiry in schools.

ACKNOWLEDGEMENTS

This work is being funded by the EU, under the ERASMUS + Key Action 2 program (DataScEd4CiEn -Data Science Education in STEAM for Civic Engagement and Social Justice from the Early Years (Ref. #. 2023-1-CY01-KA220-SCH-000164724)). Any opinions, findings, and conclusions or recommendations presented in this paper are those of the authors and do not necessarily reflect those of the EU.

REFERENCES

- Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. (2020). *Pre-K-12 Guidelines for Assessment and Instruction in Statistics Education (GAISE) report II*. American Statistical Association and National Council of Teachers of Mathematics.
- Burrill, G., & Pfannkuch, M. (2023). Emerging trends in statistics education. *ZDM – Mathematics Education*, 56, 19–29. <https://doi.org/10.1007/s11858-023-01501-7>.
- Cobb, P., Confrey, J., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9–13. <https://doi.org/10.3102/0013189X032001009>.
- Engel J. (2017). Statistical literacy for active citizenship: A call for data science education. *Statistics Education Research Journal*, 16(1), 44–49. <https://doi.org/10.52041/serj.v16i1.213>.

- Friedrich, A., Schreiter, S., Vogel, M., Becker-Genschow, S., Brünken, R., Kuhn, J., Lehmann, J., & Malone, S. (2024). What shapes statistical and data literacy research in K-12 STEM education? A systematic review of metrics and instructional strategies. *International Journal of STEM Education*, 11(1). <https://doi.org/10.1186/s40594-024-00517-z>.
- Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal*, 16(1), 22-25. <https://doi.org/10.52041/serj.v16i1.209>.
- Kahne, J., & Bowyer, B. (2017). Educating for Democracy in a Partisan Age: Confronting the Challenges of Motivated Reasoning and Misinformation. *American Educational Research Journal*, 54(1), 3-34. <https://doi.org/10.3102/0002831216679817>.
- Kazak, S., Leavy, A., & Podworny, S. (2025). Framework for developing data science education modules for teacher education in STEAM subjects [Poster presentation]. Presented at the 14th Congress of the European Society for Research in Mathematics Education (CERME14), Bolzano, Italy.
- Liston, M., Morrin, A. M., Furlong, T., & Griffin, L. (2022). Integrating data science and the internet of things into science, technology, engineering, arts, and mathematics education through the use of new and emerging technologies. In *Frontiers in Education*, 7(757866). <https://doi.org/10.3389/feduc.2022.757866>.
- Makar, K., Fry, K., & English, L. (2023). Primary students' learning about citizenship through data science. *ZDM – Mathematics Education*, 55, 967–979. <https://doi.org/10.1007/s11858-022-01450-7>.
- Meletioui-Mavrotheris, M., Serradó Bayés, A., Leavy, A., Frischemeier, D., Bakogianni, D., & Papanastasiou, E. (2025). Navigating subjectivities and perspectives in data science education: A Toulminian analysis of two STEAM scenarios from the European project DataScEd4CiEn. To be presented at the 14th International Forum for Statistical Reasoning, Thinking and Literacy (SRTL-14), Penn State University, USA.
- Noll, J., Schnell, S., Gould, R., & Makar, K. (2022). New ways of interacting with data, context, and chance in statistical modeling processes. *Mathematical Thinking and Learning*, 24(4), 331-335. <https://doi.org/10.1080/10986065.2021.1922855>.
- Ridgway, J. (Ed.). (2022). *Statistics for empowerment and social engagement: Teaching Civic Statistics to develop informed citizens*. Springer. <https://doi.org/10.1007/978-3-031-20748-8>.
- Schreiter, S., Friedrich, A., Fuhr, H., Malone, S., Brünken, R., Kuhn, J., & Vogel, M. (2024). Teaching for statistical and data literacy in K-12 STEM education: a systematic review on teacher variables, teacher education, and impacts on classroom practice. *ZDM – Mathematics Education*, 56(1), 31-45. <https://doi.org/10.1007/s11858-023-01531-1>.
- Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). Sage Pu
- Zieffler, A., Garfield, J., delMas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 40–58. <https://doi.org/10.52041/serj.v7i2.469>.
- Zucker, A., Noyce, P., & McCullough, A. (2020). JUST SAY NO! Teaching students to resist scientific misinformation. *The Science Teacher*, 87(5), 24-29. https://doi.org/10.2505/4/tst20_087_05_24.