Authentic open data inquiry in schools

Celis Vargas, Alejandra; Magnussen, Rikke; Mulder, Ingrid; Larsen, Birger

DOI
10.1016/j.caeo.2025.100287

Publication date
2025

Document Version
Final published version

Published in
Computers and Education Open

Citation (APA)
Celis Vargas, A., Magnussen, R., Mulder, I., & Larsen, B. (2025). Authentic open data inquiry in schools. *Computers and Education Open*, *9*, Article 100287. https://doi.org/10.1016/j.caeo.2025.100287

Computers and Education Open 9 (2025) 100287

Contents lists available at ScienceDirect

Computers and Education Open

journal homepage: www.sciencedirect.com/journal/computers-and-education-open

Authentic open data inquiry in schools

Alejandra Celis Vargas [a,*], Rikke Magnussen [a], Ingrid Mulder [b], Birger Larsen [a]

[a] Department of Communication and Psychology, Aalborg University, A. C. Meyers Vænge 15, Copenhagen 2450, Denmark
[b] Faculty of Industrial Design Engineering, Delft University of Technology, Landbergstraat 15, Delft 2628 CE, the Netherlands

ARTICLE INFO

Keywords:
Open data
Open data education in schools
Open data competencies
Authentic learning
Game-based learning

ABSTRACT

Open data offers the potential to enhance citizen participation, transparency, and accountability in society. However, a lack of competencies that enable citizens to engage in open data ecosystems remains a barrier. Although authentic open data inquiry has been identified as a promising approach to develop open data learning designs in schools, its connection to open data competencies is not yet well understood. To advance the understanding of open data competencies and learning designs, this study focuses on the inductive analysis of two design-based research cycles and four interventions in Danish schools. A cohort of 7th to 9th grade pupils (n = 78) and their teachers (n = 4) engaged with The Open Data Newsroom, an open data learning design that situates pupils in the role of data journalists to solve an environmental mystery. Following a thematic analysis approach, we examined qualitative data from observations, surveys, and interviews to identify four categories that encompass pupils' practices for (1) navigating open data: find and assess relevant information and data to identify a problem; (2) developing authentic open data analysis: analyse and interpret data in connection to real-world problems and local contexts; (3) building authentic data arguments and stories: explain a problem with data from different sources and domains to lay audiences; and (4) creating open data representations: build tools to support inquiry and communication. We argue that these practices, grounded in data literacy and real-world problem solving, contribute to defining open data competencies in schools, and we present a model to illustrate this connection.

1. Introduction

Governments, research institutions, and organisations have increasingly made datasets openly available on topics such as public services, environmental quality, and city infrastructure to enhance transparency, public engagement, and citizen-driven innovation [1,2]. These are examples of *open data*, which is defined as digital data made available with the technical and legal characteristics necessary to be freely used, reused, and redistributed [3]. While one of the reasons to openly release data is to foster greater civic participation and social value, a gap exists between the availability of these resources and the ability of non-data expert citizens or lay audiences to leverage them and participate in open data ecosystems [4,5].

The use of open data often requires specialised abilities or tools. Open data users need to be familiar with e.g. dataset formats, statistics, text processing software, programming languages, and interfaces [6]. These abilities are associated with competencies in Data Literacy [7,8] or Data Science, depending on the proficiency level [9]. However, to foster citizens' engagement in open data ecosystems, they must be able to not only technically manage available open datasets but also understand the perspectives it can create in an open and data-driven society [5–7]. Open data ecosystems have been defined as dynamic systems of data cycles [10]. In these ecosystems, actors can adopt several roles such as users, data providers, or intermediaries in interdependent and dynamic ways [11]. The diversity of roles implies a variety of skills and competencies which have not been clearly defined.

The broad problem of a lack of competencies in the open data field has been tied to a call for open data educational initiatives that enable non-expert citizens or lay audiences to participate in open data ecosystems [5,12,13]. In particular, the integration of open data in school curricula has been suggested to strengthen the community of participants and beneficiaries of open data (Open Data [3]), and to support several groups of citizens and communities in adopting different roles in open data ecosystems [5,14,15].

The increasing research on the integration of open data in school education has exhibited a potential space to develop data and digital

* Corresponding author.
E-mail address: cacv@ikp.aau.dk (A. Celis Vargas).

https://doi.org/10.1016/j.caeo.2025.100287
Received 24 April 2025; Received in revised form 4 September 2025; Accepted 8 September 2025
Available online 9 September 2025

skills, and to foster civic participation in young pupils by connecting classrooms with the real world [16–18]. The main focus of current educational interventions in school has been on data exploration activities for learning school subjects such as geography or statistics [19]. Although data exploration activities in school have been shown to increase data awareness in pupils, this perspective limits the pupils as users of data rather than as participants in open data ecosystems. Participation in open data ecosystems might range widely from being aware of a social issue to creating a data product to analyse it, or from collecting data to explaining it to non-data experts [20,21]. Therefore, data awareness does not necessarily increase the opportunities of young pupils to participate in open data ecosystems as active citizens or experts of their local issues and context [14].

Although the literature often addresses data literacy in civic or professional open data domains [22,23], little is known about the competencies that enable young students to engage with open data and open data ecosystems. Open data competencies might therefore be a wide set of skills and abilities [24].

Previous research has explored open data learning designs in schools to foster open data competencies grounded in the domains of data literacy and real-world problem-solving [14,24]. Building on this work, an open data learning design called The Open Data Newsroom has been iteratively developed to foster authentic open data inquiry as a promising learning space to develop open data competencies [25]. The Open Data Newsroom is a game-oriented learning design that immerses pupils in a simulated data journalism practice for solving an environmental mystery with open data. Although this design has been collaboratively developed and tested with pupils and teachers, further understanding is needed of the competencies and skills enacted, and the ways in which pupils practice them.

This study addresses that gap by empirically identifying competencies as they emerge in classroom practice, and by clarifying their relevance for open data education in schools. The current work seeks to answer the research question of how young pupils practice open data competencies through their engagement with an open data learning design that fosters authentic open data inquiry in schools. Guided by a design-based research methodological framework, we approach this question through the inductive thematic analysis of qualitative data collected from pupils and teachers during four interventions conducted in upper secondary school (7th to 9th grade) using The Open Data Newsroom.

2. Background

Previous research has identified that current open data educational initiatives for non-data expert citizens or lay audiences often emphasize activities that involve practicing data inquiry in authentic contexts and collaborating with communities [14]. In our previous studies on the underlying concepts of open data competencies and learning designs, we have found these activities to be closely related to the domains of data literacy and real-world problem-solving [24,25]. In the school context, data literacy has been defined as the ability to understand, find, collect, interpret, visualise, and support arguments using quantitative and qualitative data to answer real problems [26,27]. Real-world problem-solving has been defined as a dynamic analytical process used to address complex or ill-defined issues typically encountered in real-world environments [28]. Studies in civic and critical data literacies may be seen as bridge between these domains by encompassing the ability to ask socially relevant questions of concern from data, interpret and use it as evidence for civic action, while also critically assessing data's sources, biases, and roles in everyday life [29,30].

Common for data literacy and real-world problem-solving is that they involve processes of inquiry [24]. Steps or phases of inquiry such as problem definition, collecting or acquiring data, building arguments, elaboration of explanations, and solutions have been identified in both approaches [31,32]. A clear distinction between them is however that real-world problem-solving is mainly concerned with doing inquiry to solve authentic problems and including authentic contexts [31], and data literacy is mainly concerned with doing inquiry with data [26]. In data literacy, the inquiry process is focused on the specifics that data offers and not necessarily on the context of data. In real-world problem-solving, the inquiry concerns both the process of problem solving and connecting to authentic contexts. Considering both in our conceptualisation of open data competencies and learning designs, connecting data and authentic context is central in the inquiry process. We explore how these connect together, firstly from a theoretical lens, and secondly in an open data learning design.

2.1. Doing authentic open data inquiry

Inquiry-based learning is grounded in the philosophical and educational theories of John Dewey, who emphasizes the active, and reflective nature of learning as a process of inquiry [33]. Dewey [34] elaborates an epistemology of practice directed at solving issues through interaction with objects in the world rather than receiving decontextualized, subject-specific facts from books and blackboard teaching alone [35]. According to Dewey [33], knowledge arises from the combination of reflection and action and implies transformative reflection over time. This knowledge is acquired through a process of inquiry understood as the transformation of an indeterminate situation into a new situation [34,3634].

The inquiry process starts with an indeterminate situation and a collection of facts about it. Through reflection, observation, and thinking about what is problematic in the indeterminate situation, firstly, suggestions or ideas are proposed, based not just on the facts of the case, but also on the previous knowledge of pupils. Secondly, the combination of facts and ideas leads to creating hypotheses or proposals for action. Following, a coordinated action aims to test the hypothesis and create new knowledge about the problematic situation. In other words, this process represents the connection between doubt and its elimination [37]. Biesta and Burbules [36] emphasize that inquiry is not a strictly linear sequence but a flexible process that adapts to each experience.

This process of inquiry has been applied in data science and data literacy to explore and make sense of data through questioning, analysis, interpretation, and reflection. Inquiry in this context includes practices such as asking meaningful questions about data, exploring patterns, trends, and anomalies, drawing inferences, or making decisions based on evidence, and iterating the process through new questions and further investigation. Considering notions of experiential learning, Wolff, Cavero Montaner, et al. [8] frame the data inquiry process around authentic problems to develop data literacy in schools. Experiential learning stresses that learning should be situated within a real-world context and concerning pupils' experience rather than relying on rote learning of a collection of facts [34].

The alignment of educational practices with the real world has been described as authenticity in education [38]. Authentic learning is a pedagogical approach that situates learning tasks in contexts of future application, enabling learners to transfer robust knowledge into practice [39]. The realistic tasks challenge learners to solve problems by thinking in the same ways as professionals working in real-world contexts [40,41]. Shaffer and Resnick [42] define authenticity in educational interventions through four interdependent dimensions: (1) real-world authenticity, where learning refers to contexts outside school; (2) personal authenticity, where learning is personally meaningful; (3) disciplinary authenticity, where learning reflects the practices of a discipline; and (4) assessment authenticity, where assessment aligns with the learning process. Nachtigall et al. [43] elaborate on this definition to formulate four intentions of authenticity: (1) to create personally meaningful learning activities, (2) to emulate the work of professionals of a certain discipline, (3) to connect learners with a community of practitioners, and (4) to reflect experiences from real/daily life.

Authenticity is then achieved in the relation of the learning activity with the real world, a professional practice, and personally to the learners [42,4342].

Authentic inquiry is therefore rooted in real-world contexts, personally meaningful to learners, and aimed at producing knowledge or solutions with relevance beyond the classroom. In this perspective open data not just contributes with real facts, but facilitates a reflection on the data context which should be close to pupils' interests and daily lives.

Grounded in notions of authentic and experiential learning, authentic open data inquiry emerges theoretically from the combination of the data inquiry process of solving problems with data and the authentic inquiry process of addressing complex real-world issues. It encompasses processes where learners engage in the investigation of real-world problems using actual, meaningful open datasets, guided by self-generated or socially relevant questions to draw insights, make decisions, or take informed action.

2.2. Learning designs for authentic open data inquiry

In previous research, the iterative development of an open data learning design contextually explored game-based learning and authentic principles to create opportunities for authentic open data inquiry in schools [25,44]. In this design, game-based learning is an approach to situate abstract knowledge within engaging, problem-oriented activities, while authenticity to ensure that such activities are connected to real-world contexts, professional practices, and pupils' personal experiences. Together, these pedagogical approaches theoretically underpin learning activities where pupils are supported in engaging with the complexity of open data.

Educational games contribute to making knowledge that is abstract and common for experts more concrete and embedded in a situation [45, 46]. Games have been used in new approaches to authentic science education [45] to create environments where pupils think like professionals and solve problems as professionals do [40,41,47]. Games support the educational goal of producing worthwhile effects that transfer from one context to another [48].

In this study, the simulation of an authentic activity is central to enabling open data inquiry. An authentic activity has been defined as the ordinary activity of the practitioners [49]. In open data learning designs, it includes professional inquiry processes by identifying problems, collecting and analysing data, and communicating findings. In this way learners collaboratively solve ill-structured real-world problems through self-directed inquiry and investigation. The creation of authentic activities in the classroom can be supported through various types of game-oriented learning designs, such as storylines and role-play [50].

In an open data learning design, storylines are approached to embed open datasets and complex societal problems into a narrative that provides coherence and purpose for pupils' inquiry. On the other hand, role-play is used to further strengthen authenticity by assigning pupils professional roles with specific responsibilities for open data processes and ecosystems such as data exploration or collection, interpretation, and communication to lay audiences. To be authentic, this approach requires the integration of elements of an authentic open data practice such as open datasets, data exploration tools, and hybrid elements (physical and digital).

A game-oriented open data learning design therefore aims to enhance three kinds of authenticity: professional, personal, and real-world authenticity. Professional authenticity to support pupils engaging in data inquiry as professionals; real-world authenticity to support pupils linking the learning activity and the real world; and personal authenticity to support pupils working with problems that are personally meaningful and connected to their daily life.

3. The Open Data Newsroom: a game-oriented open data learning design

The Open Data Newsroom is a game-oriented learning design intended to simulate the practice of data journalism, positioning pupils as data journalists within a newsroom context under the guidance of an editor. Through the process of conducting authentic open data inquiry, pupils engage in addressing locally relevant environmental issues with open data [44].

The learning design builds a narrative where the chief editor of the so-called "Data Journalists Hackers", a fictitious international network, invites professional data journalists in a local newspaper to analyse several mysterious events that have recently affected local teenagers, schools, and the surrounding environment. Four cases are presented to several teams of data journalists (pupils) including open data, press articles, scientific papers, and social media posts are physically provided to contextualise the data. The data journalists need to unravel these mysterious events with open data and present the truth behind to the public in a Press Conference.

A chief editor guides the work of the pupils as data journalists in the newsroom by providing feedback and challenging the data behind pupils' conclusions to ensure the achievement of the interests of the newspaper and the citizens. This editor's role is proposed to be played by the teacher or an external facilitator. Simulating real interdisciplinary data journalist teams, pupils are divided into groups of two or three people, trying to ensure that a variety of skills and interests are included in each team.

Data journalists work with open data from different sources such as a public repository, data generated by so-called "Data Journalists Hackers", scientific reports and the press. An open data repository built of open data visualisations is provided to pupils in a digital online platform that allows exploration through interactive visualisations [51]. Through the platform, pupils work with both real water quality open datasets [52] and generated data about the game mysteries, also called synthetic data.

The learning activity is structured to mirror an authentic data journalism practice, unfolding across six interconnected phases. Initially, pupils work as data journalists work on four isolated cases, investigating each separately. However, as their clues emerge, the Data Journalists Hackers network and the chief editor in the newsroom suggest that the cases may be connected, prompting pupils to pursue a larger investigation. They ultimately uncover a broader case of water and environmental damage caused by a company. Throughout the activity, pupils document their claims in an open newspaper repository to enhance transparency. Table 1 outlines the six phases and their sequence, providing an overview of the activity's progression. A more detailed description of the full design and its iterative development is presented in Celis Vargas et al. [25].

The game-oriented learning design deliberately deploys inquiry processes through its game mechanics. The initial case exploration requires pupils to identify and frame a problem, the open data repository provides authentic material for evidence gathering and analysis, and the editorial meeting introduces the collaborative experimentation of hypotheses. The press conference represents the communicative phase of inquiry, where findings are presented to an audience.

In this way, the open learning design models an inquiry cycle that is both authentic and driven by open data, providing a structured context for pupils to practice authentic open data inquiry. The following sections focus on how pupils enact these practices in schools and how they contribute to the definition of open data competencies.

4. Methodology

The study adopts a design-based research methodological framework to support the iterative development of educational approaches grounded in both theory and practice [53–55]. This approach has been chosen

Table 1
Phases and sequence of The Open Data Newsroom activity based on Celis Vargas et al. [25].

Game phase		Description
1. Introducing the narrative and the activity		The chief editor presents the game's narrative, goal, and dynamic. The teacher creates teams, and they choose a case.
2. Getting and understanding data	Defining storylines	Data journalists receive physical envelopes with news and social media posts about the mysterious events to build their case board.
	Finding data insights	Data journalists analyse open data through data visualisations on a platform. This contains an open data repository of water quality data and case data.
3. Editorial meeting and open repository		All teams working on the same case join and meet the chief editor. Create an open online repository to avoid publishing unsupported claims.
4. Plot twist - One big case		The "Data Journalists Hackers" network in a video call presents the hypothesis of a bigger case based on the ongoing investigations and introduces self-collected data.
5. Preparing the data story		Data journalists prepare for communicating their solutions to local citizens in a press conference.
6. Delivering data: Press conference		Each team of data journalists presents. The critical audience poses questions. The chief editor presents the solution to the mystery.

as it enables the simultaneous improvement of classroom practice and the generation of context-sensitive design principles and theoretical insights through the collaboration between researchers and practitioners in real-world settings [56]. It develops in iterative cycles consisting of problem definition, design, intervention, and analysis [57] allowing continuous refinement of the learning design based on real-world implementation. Design-based research cycles have various phases that include context and domain research, defining design hypotheses, developing theory-based prototypes, and analysis with the dual focus on developing new practices and theory [58].

This methodological framework guided the development of The Open Data Newsroom through three cycles and five interventions in school [25]. The exploratory first cycle provided contextual insights, which informed the design of the learning space in the second cycle and supported the refinement of competency enactment in the third cycle. The present study focuses on the analysis of the qualitative data collected from the last two cycles which were conducted with The Open Data Newsroom to understand what learning and learning practices take place in school and analyse them in the theoretical context of authentic open data inquiry towards defining open data competencies.

4.1. Design-based research iterative cycles and participants

Two design-based research cycles and four interventions were conducted with the participation of 78 pupils (aged 14–16) and four teachers in Danish schools. The first cycle was aimed at defining learning design elements, while the second cycle was aimed at delving into the development of competencies. Each cycle included two interventions in Danish schools; each intervention consisted of playing The Open Data Newsroom, a game-oriented learning design, in the classroom during regular school time with a group of pupils and one teacher.

During the first intervention (int 1), we tested the learning design using physical elements with pupils in the 7th grade. During the second intervention (int 2), we incorporated a digital platform with pupils in 8th and 9th grade [44]. Both interventions were conducted with 17 pupils and one teacher. The third intervention (int 3) was conducted with 21 pupils in the 7th grade and one teacher, and the fourth intervention (int 4) included 23 pupils in the 8th grade and one teacher. Interventions were facilitated either by a member of the research team or the teacher. Each student had a laptop and access to the internet. Table 2 summarises the information about the four interventions conducted in Danish schools during the two cycles included in the current study.

4.2. Methods: data sources

We used a mixed-methods approach to observe and analyse interaction and collaboration processes that emerged in schools with The Open Data Newsroom [59]. During the interventions, the participation of pupils and teachers was documented for analysis using audio and video recordings, as well as photography. After each intervention, qualitative research methods were applied to investigate what kinds of learning and learning practices took place. Table 3 summarises the qualitative data collection methods and participants involved.

Observations were conducted to document pupils' practices, interactions, and use of the materials during the whole intervention in school. These observations were developed through audio and video recording of pupils' dialogues and interactions to understand their reflections and knowledge supported by the learning design [60]. Audio and video recordings were made both in the classroom and within each pupil's group.

A qualitative survey was applied to pupils immediately after the intervention to capture their perception. It consisted of a brief online form with two open-ended questions: "What did you like about the activity? Why?" and "What was the most challenging? Why?".

Semi-structured interviews were conducted with pupils and teachers to further understand their experience, as well as the potential outcomes and challenges of the project [61]. A focus group interview was held with all the participating pupils after the intervention to dive into their experience, and ideas. It was developed as a conversation for about 30 min following a guide that addressed three topics: the game experience, the learning experience and ideas for development. Open-ended questions such as "How did you feel as Data Journalists solving the mystery? What was different from everyday teaching? And how could the game be more interesting or fun for you?" guided the conversation.

Interviews with teachers were held after each intervention for approximately 45 min to gain insights for both improvement and validation of the learning design. These in-depth interviews consisted of two parts, the first part focused on the game and the intervention, and followed the same topics discussed with the pupils: the game experience, the learning experience, and ideas for development. The second part addressed broader questions about their experience teaching data literacy and real-world problem-solving, as the underlying domains of the learning design. This part included examples from their experience and a discussion about how these domains were practiced by pupils in the learning design and intervention. Following a guide, open-ended questions included, "What do you think were the main competencies practised by the pupils during the game? How do you think the game contributes to building data literacy and real-world problem-solving

Table 2
Interventions in Danish Schools during the two cycles included in the current study.

Intervention (Int)	Date	Grade	Participants	Duration
Int 1	January 2024	7th grade	17 pupils 1 teacher	3 h
Int 2	March 2024	8th and 9th grade	17 pupils 1 teacher	3 h
Int 3	May 2024	7th grade	21 pupils 1 teacher	3.5 h
Int 4	June 2024	8th and 9th grade	23 pupils 1 teacher	4 h

Table 3
Qualitative data collection methods during each school intervention.

Method	Participants	Duration	When
Observations	Pupils participating in the intervention	2.5 to 3.5 h (varied by session)	During the intervention
Qualitative survey	Pupils participating in the intervention	10–15 min	After the intervention
Focus group interview	Pupils participating in the intervention	20–30 min	After the intervention
Semi-structured interview	Teachers participating in the intervention	45- 60 min	After the intervention

competencies in pupils? And how can the game do it better?".

Following data protection regulations, sessions with pupils and teachers were recorded, anonymised, and transcribed. Furthermore, sessions with pupils were conducted in their primary language (either English or Danish), and translations to English were carried out for analysis purposes.

4.3. Analysis of data

The analysis of the qualitative data followed a *Thematic Analysis* approach to identify, analyse, and interpret patterns of meaning (themes) within textual data [62]. This approach enabled us to discover themes and patterns about the kinds of learning and learning practices that took place in school with 'The Open Data Newsroom' [63]. We conducted an inductive analysis to explore and uncover the practices that emerged among students [64,65]. Therefore, the analytical process was data-driven, themes were derived directly from the data collected rather than from pre-existing coding frames or analytical pre-conceptions. The resulting themes are discussed in relation to authentic open data inquiry and open data competencies to ensure they are coherent, distinctive, and conceptually consistent [62].

The collected qualitative data were coded and analysed for the identification of different themes [62]. Our process was highly iterative, with the direct participation of two researchers in the coding process and the participation of another researcher in iterative discussions to refine codes and ensure consistency. We analysed the data from the four interventions in separate packages and rounds to ensure that the results of each step accurately reflected the perspectives of pupils and teachers, and to minimise potential biases. Firstly, we considered the collection method. Data from qualitative surveys and interviews, and data from audio and video observations were analysed at different stages, considering that both had different purposes. Data from surveys and interviews captured the perception of pupils and teachers, while data from observations provided insight into the actual behaviour of pupils. Secondly, data collected from pupils and teachers were analysed separately considering that pupils' responses tended to describe their own practice, whereas teachers' responses often offered a more analytical perspective on pupils' behaviour.

According to Attride-Stirling [66], the coding process involved three main steps: first, the identification of initial codes derived from participant's textual responses; second, their classification on bigger categories according to similarities on their topics to summarise abstract principles; and third, the identification of global themes encapsulating a main idea present in different previous categories. In our process, we refer to these final themes as analytical categories, as they were derived inductively from the data, and subsequently interpreted and situated in relation to relevant literature and theoretical frameworks on authentic open data inquiry.

Initial codes were made from textual data obtained through surveys and interviews to encapsulate insights about what pupils learn with The Open Data Newsroom and how they learn it. From pupils' data, initial codes such as *'deciding what data was relevant'*, and *'figure out a mystery'* were generated (n = 123). From the teachers' data, initial codes including *'being critical of sources'* and *'to be critical about the data'* were identified (n = 99). Next, we identified categories from the initial codes. In this step, categories emerged from pupils' data (n = 29) and categories from teachers' data (n = 22). It was important to understand how these two perspectives related to one another; therefore, we combined them into categories according to their shared topic (n = 17). For example, an identified category called *'Finding out the problem'* included pupils' initial codes such as *'we got the perfect amount of data to find out the problem'*, and teachers' initial codes such as *'they need to have some data before they can understand that it is a problem'*. This example illustrates how a category was made to include two complementary viewpoints.

In the following step of the analysis, we examine the audio and video observation data to identify how pupils' practices aligned with the categories previously developed from surveys and interviews, and we coded examples of the categories. For example, we previously identified a category about *'make connections and see correlations'*, and we examined in pupils' dialogs and interactions, how they practice it during their engagement with the learning design. We then coded pupils' words such as *'we need to find out locations, dates, and times'*.

Finally, we created analytical categories that connected the various sources of data (surveys, interviews, and observations) to reflect on what pupils and teachers said and what pupils practiced. Analytical categories that unfold pupils' practices of authentic open data inquiry were identified in the end (n = 4). Fig. 1 shows an overview of the whole analytical process.

5. Results

The inductive thematic analysis of qualitative data from observations, surveys, focus groups, and semi-structured interviews allowed the identification of four analytical categories. Each analytical category encompasses processes and practices conducted by pupils while engaging with The Open Data Newsroom, which are: *(1) Navigating open data* focuses on finding and assessing relevant information and data allowing the identification of a problem; *(2) Developing authentic open data analysis* elaborates on the analysis and interpretation of data in connection to real-world problems and local environments; *(3) Building authentic data arguments and stories* entangles the process of explaining a problem with data from different sources and domains to lay audiences; and *(4) Creating open data representations* elaborates on tools to support

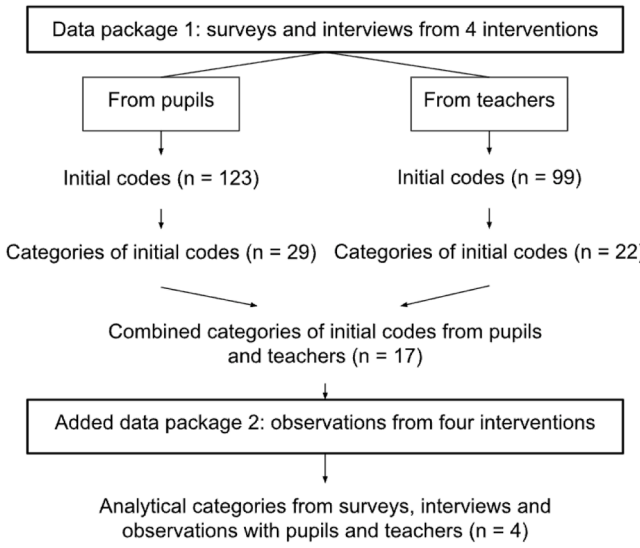

Fig. 1. Overview of the inductive thematic analysis process of qualitative data collected through surveys, interviews and observation during four interventions in Danish schools with The Open Data Newsroom learning design.

inquiry and communication.

Table 4 outlines the four analytical categories and their corresponding processes and practices.

Pupils' practices were highly iterative; therefore, these categories are better seen as overlapping rather than sequential. The following sections present findings on the open data learning practices that we have identified during the interventions with '*The Open Data Newsroom*' in schools.

5.1. Navigating open data

The practice of what pupils defined as 'navigating' different sources to make sense of data and information in relation to a problem was central in pupils' interaction with the learning design. Navigating open data was described by pupils as an exploration and sense-making practice. This practice involved two types of processes: 'deciding on the relevant data', and 'making connections and seeing correlations'. Deciding on the relevant data relates to finding a problem and assessing the reliability of sources and involves connecting different data and domains. Making connections and seeing correlations refers to linking pieces of relevant data and information from the sources. Both processes are interconnected and overlap.

In the open data learning design, pupils worked on solving different cases. They received articles, reports from authorities, and different sets of open data visualisations about the mysterious events occurring in their city. Pupils described their initial investigation processes as *"we had to find the data"* (Student_survey_Int 4), and *"we had to decide what data was relevant and what wasn't"* (Student_survey_Int 2). It is therefore central to understand what the concept of 'relevant data' means to pupils and teachers. For pupils, the relevant data was tightly connected to finding or explaining the problem. While deciding on the relevant data, pupils find a problem behind the data and make connections between the different data and materials provided in the learning design. The following example from a group work discussion occurs while they navigate the data and the mystery. In this example, the group of pupils worked with a case of sick teenagers. The pupils discussed and decided on the relevant data while navigating open data visualisations about water quality in different locations, synthetic data about the reported sick teenagers' cases, and the material provided to contextualise the cases such as articles and stories:

"Student 1: It sounds like we need to find a bit more about locations.

Table 4
Analytical categories that emerged in the thematic network analysis of qualitative data collected from observations, surveys, focus groups, and interviews with pupils and teachers.

Analytical category	Processes and practices

(1) Navigating open data	• Deciding on the relevant data
	  ○ Questioning the reliability of sources
	  ○ Finding the problem
	• Making connections and seeing correlations
	  ○ Finding a common ground in the group

(2) Developing authentic open data analysis	• Linking to authentic contexts outside school • Using local knowledge and experience from pupils' daily lives

(3) Building authentic data arguments and stories	• Bringing out open data arguments
	  ○ Creating hypotheses
	  ○ Making accurate conclusions
	• Creating authentic data stories to communicate
	  ○ Creating solutions
	  ○ Synthesising

(4) Creating open data representations	• Making an overview of the process and solution • Using data visualisations to explain the problem

Student 2: Yeah. Yeah. Dates and times
Student 1: I'm trying to make a list of dates on all these stories in a few more places.
Student 3: It is more likely to come from the water (…) This one just says 11 and 30 Student 4: They were younger when they got it
Student 3: How old were they in 2012? And how old were the others in 2012?
Student 1: If we look at the graph, we see a clear increase in cases of babies in the centre of the city.
Student 3: If a virus spread, it could have spread to other sources and places." (Students_Int 2)

In this dialogue, pupils are discussing and trying to connect several types of data while deciding on the different relevant parts. They worked on aligning the articles, stories, and datasets on water quality and sick teenagers' cases. They found that the location of cases, water quality, and age of sick teenagers are related, therefore they extracted pieces of relevant data such as dates, case numbers, and locations to understand and solve the mystery. During the interview, a teacher explained the process of navigating the open data as a critical thinking practice connected to extracting what is important:

> *"They [students] had to pinpoint what is necessary, being able to analyse, "we have this long text…What is the data in this text that we actually need and take with us?" (…) I think it is a big part of critical thinking, being able to consider different sides. What is relevant? And what is not relevant? Being able to read through articles or data and being able to pinpoint the important parts of it. "* (Teacher_Int 4).

The reliability of the data source was highlighted by the teachers as another criterion for deciding whether the data is relevant or not. One teacher mentioned:

> *"They [students] had to make judgement calls as to whether the sources were reliable or not (…) they had to ask, how the data was accumulated there? to what extent they [students] can rely on it and to what extent they have to fact-check and to double-check"* (Teacher_Int 2).

This practice described by the teacher could be seen as a process of assessment or validation of a data source by questioning how data was collected and double-checking its veracity. During the interventions, pupils reflected on the reliability of sources in different ways. They questioned the veracity and usefulness of the sources to contextualise the cases such as social media posts, press articles, and scientific reports. For example, a group of pupils discarded the social media posts as valid sources since they did not contain data. Other pupils double-checked the veracity of the information received about the cases in the game material by searching on the internet for the information that they received. In this process of navigation of open data, it was essential to create a common ground or a consensus within the group. During the survey, a student said *"It [the learning design] made people collaborate and find common ground about theories"* (Int 2_Students survey).

Navigating open data thus encompasses abilities to decide which data is relevant, to question the reliability of sources, and to draw connections between several domains and data sources, while framing a problem. In the Data Literacy model of Seymoens et al. [67], *using data* includes a navigating practice which is described as finding your way through a collection of different types of data and ways they were processed, and being able to extract the message or what you need from them. In our findings, pupils conducted the navigation of sources and data by making connections to define the problem and the relevant data.

Wolff, Gooch, et al. [32] use a data inquiry approach to categorise Data Literacy skills in connection to five stages of an inquiry process, and to real-world problem solving and ethics. Considering this model, our findings relate to the ability to ask questions based on data to identify problems or questions that can be solved with data. These findings could also be seen as the occurrence of a problem, and its specification [34], and identifying and defining authentic problems, which for Dede [31] is

an essential skill associated with problem solving. Considering that working with open data often implies working with fragmented data, crossing domains and disciplines, these abilities are highly relevant. When pupils navigate open data, they must critically assess its reliability, determine its relevance, and draw connections across datasets. Criticality in open data practices is essential not just to question the validity of sources, verify information through cross-referencing, and evaluate how data can be applied to specific problems, but also to dive into how the data was collected and its reflection of the context.

5.2. Developing authentic open data analysis

During the data analysis, pupils went beyond the learning design narrative to make connections to authentic problems outside the game context. Pupils performed authentic open data analysis in two ways: 'linking to authentic contexts outside school', and 'using local knowledge and experience from their daily lives'. Linking to authentic contexts outside school addresses the connections between the relevant data and problems in the game, and real problems outside the classroom. Using their daily life knowledge and experience refers to how pupils used their knowledge about the real places where the game cases were developed during the data analysis. These practices were central for pupils to formulate hypotheses and contextualise the open data for solving real-world problems.

The problems addressed by pupils in the learning design helped them to make a connection to the real world. As one student said in the survey *"I like that it was a problem that could have been true and took place in real life"* (Student survey_Int 3). First, they formulated ideas based on their knowledge. Second, they conducted searches on the internet to look for more data and information about them. The following example from the audio and video observation shows a group of pupils and the editor having a conversation about their case. Pupils worked with a case related to the mysterious appearances of sick fish in supermarkets and open water places. During the discussion, the pupils presented a hypothesis based on real problems to explain why the fish were found sick:

> *"Student 1: We don't think that these fish have been caught in those places*
> *Editor: Why do you think that this has happened?*
> *Student 2: We are not really certain. I have a hypothesis that some streams have been flooded and then the fish have gone.*
> *Student 3: It is always good to be careful about the flow of the river"* (Students_Int 4)

In this example, pupils built a hypothesis in relation to floods to explain the state of the fish reported in the case. The information and data provided in the learning design were limited to describing the sick fish, the locations where they were found, and the quality of water in those places. Pupils used their knowledge about floods and water streams from real problems outside the learning design narrative. The following dialogue between a group of pupils and the editor shows another example of how pupils used experiences from the real world during the data analysis. In this second example, the pupils worked with recent mysterious appearances of sick teenagers in different locations:

> *"Student 1: Can we explain why it is happening in 2024?*
> *Student 2: It could be a COVID mutation*
> *Student 3: In 2023, people started getting out*
> *Editor: Can you support that with the data?*
> *Student 2: Basically, what we need is more data to say it*
> *Editor: But you could always say that. Instead, think on specific what data you need*
> *Student 4: No, but looking at the data there have been cases before 2023. The reason should be a new disease*
> *Student 3: We have to check our ideas again"* (Students_Int 2).

To explain why teenagers are appearing sick, in this example, pupils discussed different ideas both connected to data and authentic experiences. From the data, they established that a phenomenon occurring in 2023 could have been relevant, and from their real experience, they made a connection to COVID. Pupils also reflected on the need to acquire more data to argue their hypotheses, which exemplifies an authentic open data practice.

Another way pupils practiced authentic open data analysis was by using their local experience, as the game events were developed in real local places. In the learning design, local open data was used and customised for each intervention at 20 locations near the school. During the focus group interview after the intervention, a student mentioned that *"it was easier to find the different places, because the topic was based on places in Denmark"* (Students_focus group_Int 3). This student made it evident that knowing the places contributed to the analysis of the data. During the interventions, pupils discussed the places and interacted with the digital maps on the platform. The following example from a group of pupils during an intervention shows the use of local places during the analysis of the data and elaboration of hypotheses:

> *"Student 1: So, it is connected to bathing. Where are these bathing spots?*
> *Student 2: Yeah, we've seen that most cases are around the same place.*
> *Student 3: It's very important that we can give a location*
> *Student 1: We can put the target here right [Pointing at a map].*
> *Student 4: Right. Fisketorvet got numbers. Right. Islands Brygge too*
> *Student 1: So, I'd like you to get quite specific about them"* (Students_Int 2)

In this example, we can observe how pupils explored open data from familiar places such as *Fisketorvet* and *Islands Brygge,* which they knew as bathing spots. The local knowledge contributed to their analysis by bringing their experiences and creating a hypothesis connected to bathing spots. The local open data also helped pupils to position themselves in the role of game characters. The following quotes from a group work discussion and final presentation show how they positioned one of the team members as a character in the game:

> "*[Student name] you live there [pointing at one affected location in the map] (…) [Student name] doesn't know what he has, (…) So Green Tech has been responsible for the fact that these teenagers have become ill, [Student name] who might have been you"* (Students_Int 3).

For the teachers, another way pupils made links to the real world was to understand the role of open data in real-world problems. During an interview, a teacher said:

> *"In Denmark, right now we talk a lot about our ocean environment because there is a company that lets out chemicals. So that is a real-world problem.*
> I think the students could link the game to the real world. And that they need to have some data before we can solve *the problem (…) how can they see that our ocean is bad? They can see it because they have some data, so it's important if they try to solve it that they have some data they can use"* (Int 3_Teacher).

In this quote, the teacher emphasizes the link to real problems, in a way that pupils could also link the role of open data in real life. It could be seen as a transfer from the classroom to the real world. In contrast, the pupils' use of real knowledge to deepen their analysis in the learning design could be seen as a transfer from the real world to the classroom. While the teachers' quotes highlight a potential learning outcome, the pupils' practice shows that the link to real problems and their personal experience outside the school was an active part of their inquiry process. These connections relate to notions of Critical Data Literacy, which has been defined as acquiring a critical view to assess open data and understanding its role in everyday life [49,68]. Through making links to real and local experiences, pupils understand the role of data in real-world problems and become critical about their local areas.

Developing authentic open data analysis elaborates how pupils' ability to link open data to real problems fosters a deeper understanding of what is specific to doing data analysis with open data, and it emphasises

the importance of contextualizing open data to address real-world issues. In open data ecosystems, these abilities may enhance young citizens' use of open data to be critical about local and global issues and could motivate practices such as the collection of data. These findings can be seen as ways of reflection supporting transformation in the inquiry process [34]. The indeterminate situation that pupils face at the beginning of the activity, becomes clearer and more understandable when they apply knowledge that is authentic and personally meaningful [43]. According to Dewey's five components of the inquiry process, our findings about developing authentic open data analysis could be seen as occurrences of a solving suggestion or supposition, when pupils focused on creating hypotheses.

5.3. Building authentic data arguments and stories

This category encompasses several practices and processes that pupils conducted to create data-driven arguments and authentic stories. They transitioned from identifying relevant open data and problems to creating full arguments and authentic stories that explain the issues with open data to non-expert audiences. The processes of creating authentic data arguments and stories were highly iterative. Pupils conducted back-and-forth processes to generate and evaluate hypotheses, make accurate conclusions, elaborate solutions, and synthesise.

A central practice for pupils was coming to 'accurate conclusions' emphasising the use of open data. In the survey, one student highlighted that in the activity they focused on *"making sure that everything we said was connected to the data and finding the most accurate conclusion"* (Int 2_Students survey). For teachers, the central practice was finding arguments from open data to explain the problem. A teacher said, *"They [students] bring out arguments of why they think this [hypothesis] is or isn't extremely important. They have to keep backing it up with data"* (Int 4_Teacher). Pupils built arguments while progressively deepening their understanding of open data and problems. The more they became familiar with open data, the better they practiced finding arguments to explain the problem. The following example from a group of pupils working with the sick teenagers' case during the editorial meeting outlines their argumentation process:

> *"Something we realised very quickly was that it had something to do with the water quality because water quality got exponentially better from here to here [pointing at data visualisations], so that gave us a couple of questions. If the water is important and the quality gets better, then it's probably not bad water for people to swim in. It might be something else. And, in all the articles we read. It said only teenagers could get the disease and we then thought What do teenagers have that other people or grown-ups don't have? And we thought it might have something to do with puberty, which will lead us to our theory. So, our hypothesis is that in 2011, when the current teens were one to four years old, the water quality was at its worst and their immunity was worse as well. So, the disease has been incubated for around 11 to 15 years and during puberty"* (Students_Int 2)

During the final stage of the learning design, the pupils presented their findings in a simulated press conference to inform the citizens about the mysterious events affecting the city. Pupils worked on making stories to explain the problem to other people and the solution in a way that citizens could understand the open data behind. The following example from a group presentation shows how pupils used open data to create a story and explain their argumentation. The group of pupils worked with a case related to sick fish:

> *"Christians Grave is one of the places that has gotten the sickest fish. We looked at the water quality there and it is really bad. There is also something with the temperature, there has been over 20 degrees in the long run of the water, then an alga proliferated creating some poisonous blockages. And when it rained a lot, the fish ran out into the sea"* (Students_Int 3).

Pupils used open data such as the location, water quality, temperature, and fish cases to create a narrative explaining the problem and to argue their hypothesis. During the Press Conference, the pupils immersed themselves in their role of data journalists. A student said: *"It [data journalist's role] was useful when we were doing a press conference. It just helped me put together my words"* (Int 2_Focus group). This quote exemplifies how the creation of authentic stories was supported by professional authenticity in the learning design. For teachers, the pupils' process of creating authentic data arguments and stories is the result of an analytical process and a synthesising practice. One teacher said, *"They [students] did a sort of synthesis of it all, I suppose it's putting it into a press release"* (Int 2_Teacher).

Pupils practised open data inquiry with a focus on solving authentic problems. In Wolff et al. [8] inquiry approach to Data Literacy, the analytical phase when learners analyse and create explanations to a problem from data is essential to solving a problem. Creating data arguments and authentic stories are ways of elaboration of suggestions or reasoning according to Dewey's five components of the inquiry process [34]. *In building authentic data arguments and stories,* pupils transform open data into meaningful insights and narratives, which is essential to using open data for real-world problem-solving. Pupils identified key patterns, formulated data-driven arguments, and drew accurate conclusions. The process of creating hypotheses and synthesizing information fosters critical thinking and analytical reasoning, which are essential for authentic open data inquiry. Additionally, storytelling with open data allows pupils to communicate complex findings in an accessible way.

5.4. Creating open data representations

Open data representations had a central role in pupils' authentic practices of open data analysis, and problem-solving. While engaging with the authentic open data inquiry process previously described, pupils created and used different data representations as tools to support their process and facilitate critical thinking, decision making, and communication. Regarding the creation of open data representations, two main practices were conducted by the pupils: 'making overviews' and 'using data visualisation to explain the problem'.

Pupils worked with making overviews as tools for supporting their practices such as making connections, deciding on the relevant open data, and making arguments. Pupils created different representations to register their insights and main findings during the exploration of the problem and the open data. Using the provided materials such as pen, paper, and sticky notes, pupils created mind maps, visual notes, among other representations. Different versions of the overviews were created as the analytical process of pupils progressed. Fig. 2 shows, in three pictures, the process of a group of pupils creating an overview and its different functions during the activity.

In Fig. 2, the first picture (left) shows the pupils collecting and registering pieces of relevant open data from the digital and physical materials. They wrote dates, locations, and actors involved in the case. The second picture (centre) shows how the overview was used for holding discussions inside the group. At this stage of the activity, pupils wrote ideas of the problematic situation behind the mysteries such as pollution and sickness, as well as their hypotheses. The third picture (right) shows the final version of the overview which was used for communicating to the public (pupils and teachers) during the press conference. Pupils divided this final version of the overview into three parts, first *"what we questioned"*, which reflects their process of deciding on the relevant open data, the second part contains their hypotheses, and the last part registers their solution and main data arguments. The different versions were developed as the case work progressed and thus became representations of pupils' knowledge of the case at a given moment during the learning design.

The following dialogue shows the discussion inside the group of pupils in Fig. 2, which took place during the creation of the overview.

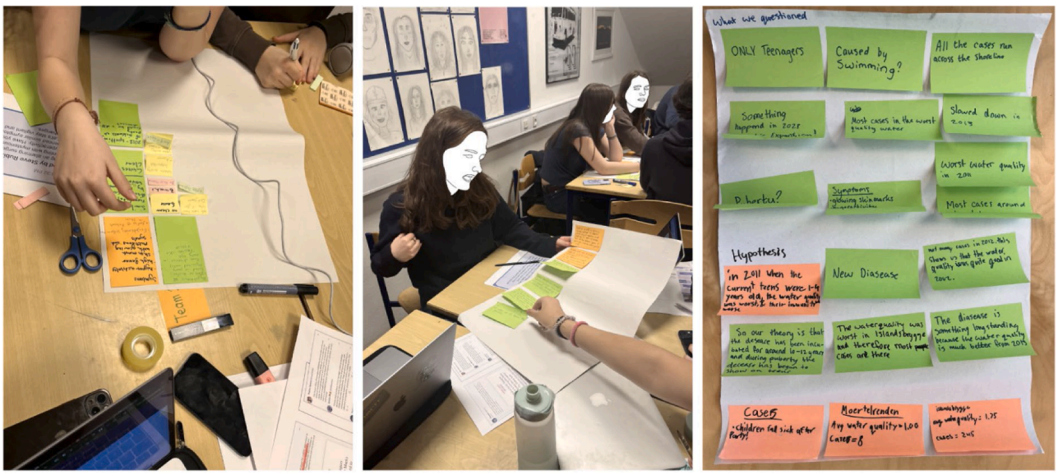

Fig. 2. Process of a group of pupils (anonymised) creating and using an overview about their case during the second intervention in school with The Open Data Newsroom learning design. The first picture (left) shows the pupils collecting and registering pieces of relevant data. The second picture (centre) shows how the overview was used for holding discussions. The third picture (right) shows the final version of the overview.

Pupils conducted their analysis while using different materials and sources given in the learning design, but also, they made Google searches by themselves:

> *"Student 1: I am reading [in Google] about an organism that is typically found in large bodies of water (…) It is an Amoeba in the water.*
> *Student 2: So, you should add this [pointing at the papers and sticky notes].*
> *Student 3: Maybe the rain could have something to do [with the mystery].*
> *Student 2: I think rain is a good one [pointing at the papers and sticky notes]. Might be something with an alga that only occurs in the summer season.*
> *Student 1: What can algae do to humans?*
> *Student 4: No, but it's only teenagers, the ones that are affected. It has to do with puberty. I was reading [game material], and it says puberty. It doesn't make logical sense.*
> *Student 1: We also have to talk about temperature.*
> *Student 2: I mean, these are our hypotheses. Can you write that [pointing at the overview]?" (Students_Int 2)*

In this example, pupils discuss open data and hypotheses at the same time. They use the overview as a tool to connect authentic open data and build arguments that drive their creation of hypotheses. This tool supports them in keeping a register of the inquiry process while they explore all the different sources, and to keep track of the relevant data or show the lack of it. Through this open data representation, pupils contextualised the open data, which is essential to generate knowledge on the game case. The functions of these tools were described by a teacher as:

> *"they [students] had a specific place to write down and put their insights on, so it helped them create an overview of what was the storytelling (…) they [students] had to put the pieces together and looking at the big picture (…) that's one of the things that they are currently learning in general that you have to check all of it and not just have a focus on one task but see the bigger picture"* (Int 4_Teacher).

The teacher linked the creation of an overview to collecting insights or pieces of storytelling, and to creating a big picture.

Other kinds of open data representation used by pupils were data visualisations. Pupils used the maps and graphs provided as open data visualisations as tools for building their authentic data argument and to support their communication. During a focus group, a student said: *"We use visualising from the graphs and others. So, we have something to point at instead of just talking"* (Student_focus group_Int 3). During the final stage of the learning design, the pupils presented their findings in a simulated press conference to inform the citizens about the mysterious events affecting the city. Pupils used data visualisations to support their stories and explain the problem to other people as data journalists would do. Fig. 3 shows a group of pupils during their press conference presentation; they are using a map that visualises open data to point at different locations that support the explanation of their hypothesis.

Data visualisation has long been an important component of data analysis, whether the intent is that of data exploration or as part of a model-building exercise [69]. Data visualisation allows users to explore and effectively communicate relevant information about voluminous data through graphic representations. The graphic mode includes visualisation of all kinds of information [69]. Data visualizations have increasingly been developed to allow non-data expert users to access insights about open data with personal relevance [70]. Effectively visualizing data is a key component of Data Literacy, ensuring that insights can be communicated clearly and persuasively.

In *creating open data representations*, pupils summarized and structured information and open data through data representations. Using open data visualizations such as maps and charts and creating open data representations to support their inquiry process enhances pupils' ability to explain their findings and solve problems. Pupils' open data representations like overviews or visuals for their presentations provided a glimpse of the knowledge that they had discovered about the data including contextual and domain information. According to Walny et al. [71], their practices can be seen as a transformation from numeracy to abstraction which in data visualisation supports the creation of speculative data hypotheses. In information visualisation, visual metaphors

Fig. 3. A group of pupils (anonymised) use an open data visualisation during their press conference presentation.

find uses in presenting information and in implicitly developing ways to manipulate these metaphors to make sense of the information [70].

In pupils' practice, the different open data representations created and used fulfilled these two functions, first as a visual representation useful for communication, and second, as malleable material that allowed reasoning and reflection in the inquiry process.

In the open data field, the visualisation has been stressed as crucial for enabling the interaction between non-data experts and data [16]. This competency is especially valuable in open data ecosystems when complex datasets or issues need to be accessible to diverse audiences.

6. Discussion

This study examines how pupils engage in authentic open data inquiry through an open data learning design, and how these practices relate to the development of open data competencies. While the Results section presents the specific practices and categories identified, this discussion emphasizes their broader implications. Results show that in authentic open data inquiry, pupils engage in navigating open data, developing authentic open data analysis, building authentic data arguments and stories, and creating open data representations. They practice identifying local problems, deciding on the relevant data to define and solve a problem, correlating arguments and data, making connections between fragmented data, domains, and contexts, analysing local open data visualisations, creating authentic hypotheses to explain a problem, delivering open data and data stories to the public, and creating tools to support the process and the presentation of the outcomes.

These practices are situated in the use of open data as a socially and contextually meaningful resource. They highlight how learners negotiate both technical and civic dimensions of open data use when engaging in authentic inquiry. These practices reveal how pupils not only interpret open data but also determine how it fits into a broader context to address real-world challenges [67,8]. By engaging in these analytical processes, pupils practice going beyond passive data consumption to active and informed decision-making. In this section, we discuss the theoretical and practical implications of these findings to understand how they contribute to the conceptualisation of open data competencies, the limitations of the study and further research directions.

6.1. Theoretical and practical implications

Our findings show how the inquiry process conducted by pupils with The Open Data Newsroom is simultaneously authentic and driven by open data. Pupils do authentic open data inquiry by making sense of open data in connection to authentic real-world problems which are personally relevant through navigating open data, developing authentic open data analysis, building authentic data arguments and stories, and creating open data representations.

In our open data learning design, constructing new knowledge results from a process that combines action and transformative reflection in time [36,34]. The four practices are enacted through an interplay of action and reflection enabled by open data. Pupils' action can be seen in the hands-on activities of exploring and navigating different sources and data, extracting clues, creating tools such as overviews, formulating solutions, and presenting them to the public. Pupils' reflection occurred in the decision making and criticality applied in different moments of the game-oriented learning design such as deciding on the relevant parts of the open data, making connections and correlations, and synthesising; also, in delivering the story and the data to an audience. Reflection was also supported by collaboration in group work and the dynamic interactions with the editor and other groups in the newsroom.

From a perspective on inquiry as a process of reflective problem-solving [34,33], this study understands authentic open data inquiry as a pedagogical approach that requires and simultaneously develops open data competencies. Authentic inquiry situates learning in real-world, personally meaningful contexts [34,43,42] and requires learners not only to engage with open datasets but also to practice abilities to critically navigate, analyse, argue and communicate problems and open data. Our analysis of school interventions with *The Open Data Newsroom* led to the identification of four practices and their associated skills that pupils developed through authentic open data inquiry:

(1) Navigating open data, pupils identify relevant open data and evaluated the reliability of sources; (2) Developing authentic open data analysis, they connect real-world problems to open data, deepening their understanding of data's potential to inform and solve issues; (3) Building authentic data arguments and stories, they interpret and communicate open data to explain local to lay audiences; and (4) Creation of open data representations, pupils produce tools to support their inquiry analysis, critical reflection and communication.

Although these practices resemble an inquiry process, they are not steps but are articulated practices of authentic open data inquiry that enable the practice of several skills and abilities connected to open data real practices [14]. In line with broader conceptualisations of data literacy [8,26], and real-world problem-solving [28,31] these practices become a pathway to meaningful participation in open data ecosystems as users, providers, and intermediaries [5,11,25].

In this way, authentic open data inquiry provides the context as the learning space in which pupils enact meaningful practices, while the four analytical categories (navigating open data, developing authentic open data analysis, building authentic data arguments and stories, and creating open data representations) are the observable manifestations of these practices. These practices, in turn, bridge directly to open data competencies as outcomes of engaging with the open data learning design. The competencies represent the knowledge, skills, and attitudes that enable learners to engage with open data ecosystems by solving real-world challenges using open data.

Based on our empirical findings and theoretical analysis, engaging in authentic open data inquiry supports the development of not only technical and analytical skills, but potentially higher-order competencies such as Ecosystems Thinking, Data Literacy, Problem-solving, Critical thinking and Civic Thinking. These competencies may contribute to learners' ability to engage in open data ecosystems and navigate real-world decision-making contexts. Fig. 4 provides an overview of the relationship between authentic open data inquiry, the four analytical categories, and the associated open data competencies. To clarify how these competencies are enacted in practice, we provide the following practical definitions to make explicit the skills and practices that underpin pupils' practices of authentic open data inquiry:

- Ecosystem Thinking: The ability to understand and navigate the relationships between different stakeholders, open data sources, societal contexts and domains, recognizing how local issues connect to broader scenarios.
- Data Literacy: The ability to locate, assess, interpret, analyse, and represent quantitative and qualitative data for evidence-based reasoning and informed decision-making.
- Problem-Solving: The capacity to define complex, real-world problems, generate hypotheses, explore solutions using data, and evaluate outcomes critically.
- Critical Thinking: The ability to evaluate the reliability, relevance, and limitations of open data sources and arguments, making reasoned judgments based on evidence.
- Civic Thinking: The disposition to use data responsibly to engage with societal issues, communicate findings to public audiences, and participate meaningfully in civic and community decision-making.

Our findings contribute to the ongoing discussion of how competencies support inclusive open data ecosystems. Critics to current open data systems have highlighted their focus on the data release process and data suppliers, rather than on the use of open data and the value created for citizens [1,2,72]. Literature has stressed the exclusion of several user

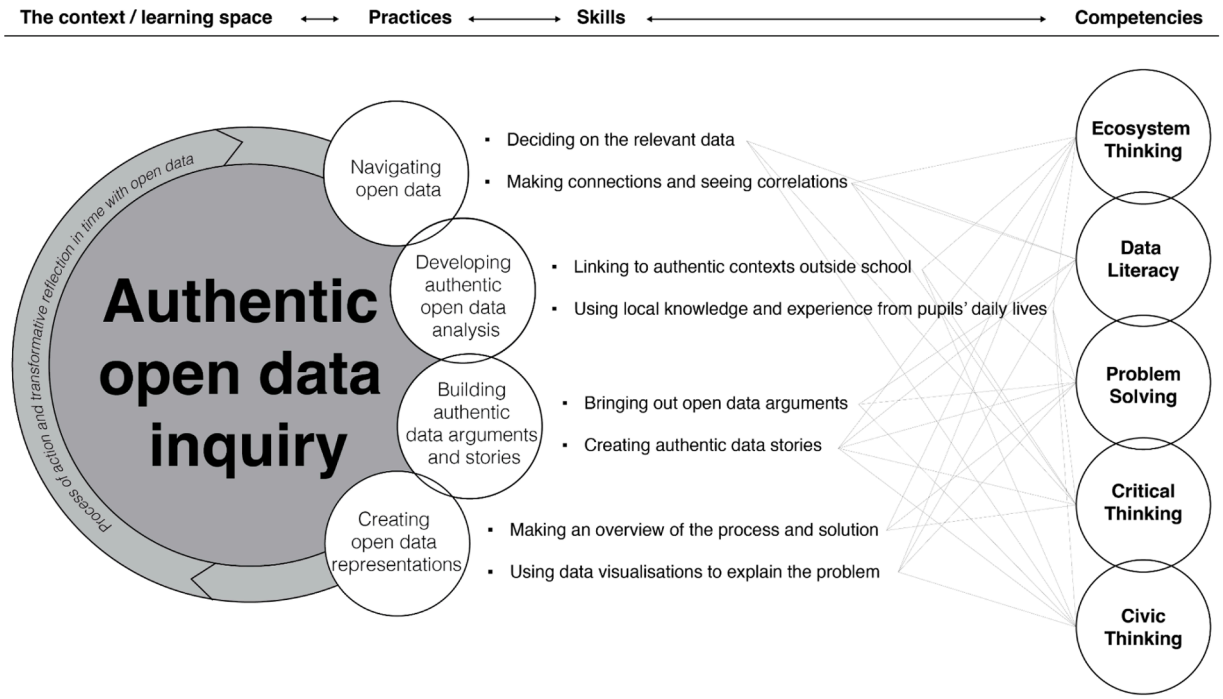

Fig. 4. Overview of the relationship between authentic open data inquiry, the four analytical categories, and underlying open data competencies.

groups such as young students due to a lack of competencies [5].

Our results show that in The Open Data Newsroom, pupils worked with abilities that can be connected to the different roles in open data ecosystems. Firstly, pupils use open and non-open data from different stakeholders to identify, understand, and explain problems. Secondly, they provide data from their analysis and are aware of paths to collect data which could be later provided as open data. Finally, pupils might be seen as intermediaries when creating hypotheses and solutions explaining problems with data to the public, or when local experts contribute their perspectives to the analysis of a problem with data.

This study contributes to advancing the field of open data education and creates a ground for further research and future initiatives towards user-driven and inclusive open data ecosystems. This study might support educators and open data practitioners in the design of initiatives aimed at increasing Open Data Literacy in young citizens.

6.2. Limitations of the study

Although this study provides important insights into how school pupils form knowledge and abilities to engage in open data ecosystems applying a design-based research methodological framework, in this section limitations are acknowledged.

First, although the design-based research methodological framework enables rich contextual insight, it might also limit the transferability of the findings [73]. The research was conducted within the Danish school system, which is characterised for strong curricular flexibility, high teacher autonomy, and openness to pedagogical innovation. Further research might focus on the transferability of findings to other educational contexts.

Second, the scope of this study is formal education in schools and does not cover extracurricular activities or informal learning environments. However, such environments could play a significant role in engaging young students in open data ecosystems and enhancing learning designs. Bridging schools, local governments and local communities could increase the authenticity of the learning activities and enhance the understanding of the impact of open data in real local contexts. Additionally, these environments could also facilitate the involvement of open data communities of practice in the learning activities.

Third, the scope of the study is on the learning practices of students rather than on the development of competencies.

6.3. Further research directions

Future research could explore the validation of these outcomes in different cultural and institutional contexts. Broader factors such as teachers' support, digital tool access, and professional development infrastructure might be critical for the implementation of our findings in diverse school systems. In addition, examining how authentic open data inquiry aligns with existing curricula and educational standards could provide insights into practical integration and policy considerations. Further studies might focus on the development of open data competencies by conducting pre-post-tests or performance assessments, in order to empirically measure the growth of pupils' skills and practices over time. Further investigation into how open data education can be meaningfully integrated across formal and informal learning environments could offer insight into governance mechanisms for its implementation. Finally, future studies are needed on designing age-appropriate, accessible, and engaging open data tools for young learners and teachers to engage with open data and reduce data abstraction and technical challenges. Lastly, further research might focus on the conceptualization of "Open Data Literacy" at the school level to advance the field and support curriculum development on a broader scale.

7. Conclusions

This study investigated how pupils engage in authentic open data inquiry in schools and how these practices relate to the development of open data competencies. Drawing on current research and our prior work, we identified four analytical categories that capture how pupils enact open data practices within a game-oriented learning design: navigating open data, developing authentic open data analysis, building authentic data arguments and stories, and creating open data representations.

Our findings show how an open data learning design provides

structured and meaningful opportunities for pupils to engage in authentic open data inquiry. Pupils interact with fragmented, heterogeneous, and ambiguous open data, make decisions about open data relevance, analyse and interpret real-world problems, construct evidence-based arguments, and communicate findings supported by open data to peers and the broader communities. This process cultivates competencies to foster young learners as active, informed contributors in open data ecosystems.

Authentic open data inquiry introduces complexity, unpredictability, and real-world relevance into the classroom, in contrast to traditional inquiry approaches that often rely on controlled or idealised problems. This complexity not only enhances open data learning but also mirrors the challenges of professional data practices, aiming at fostering transferable skills that extend beyond school.

In this way, authentic open data inquiry functions as the pedagogical context that transforms pupils' practices into competencies. It is not simply that pupils practice navigating, analysing, arguing, communicating and representing open data; rather, these practices acquire meaning and educational value because they are situated in authentic, open data–driven inquiry. Authentic inquiry thus provides the conditions under which open data practices underlying higher-level competencies. Through the theoretical analysis of our empirical findings, we have identified potential open data competencies such as Ecosystems Thinking, Data Literacy, Problem-solving, Critical thinking and Civic Thinking. Rather than presenting universal templates, the competencies identified here reflect situated practices that emerge when students engage in authentic open data inquiry.

The study contributes theoretically to the open data field by articulating how open data competencies are enacted through practices of authentic inquiry and practically by providing a framework for integrating open data learning into school education. These findings can guide educators and practitioners in designing learning activities that enhance open data learning while addressing challenges such as data complexity, teacher preparation, and pupils' motivation. The potential alignment with formal curricula has not been explicitly investigated in this study but remains an important consideration for implementation.

Limitations include the context-bound nature of the study, which was conducted within the Danish school system, and the focus on pupils' practices rather than on the validation of long-term learning outcomes. While contextual, the findings provide transferable insights for designing learning environments that prepare pupils to participate in open data ecosystems. Future research could examine the implementation of open data learning designs in diverse cultural and institutional contexts, empirically validate the development of open data competencies, explore integration across formal and informal learning environments, and design age-appropriate, accessible tools that facilitate the engagement of pupils and teachers with open data game scenarios. Further conceptual work on defining "Open Data Literacy" would advance the field and support broader curriculum development.

CRediT authorship contribution statement

Alejandra Celis Vargas: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Rikke Magnussen:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Ingrid Mulder:** Writing – review & editing, Conceptualization. **Birger Larsen:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 955569. The opinions expressed in this document reflect only the author's view and in no way reflect the European Commission's opinions. The European Commission is not responsible for any use that may be made of the information it contains.

Part of the computation done for the data analysis in this project was performed on the UCloud interactive HPC system, which is managed by the eScience Center at the University of Southern Denmark.

References

[1] Bachtiar A, Suhardi, Muhamad W. Literature review of open government data. In: Proceedings of the international conference on information technology systems and innovation (ICITSI); 2020. p. 329–34. https://doi.org/10.1109/ICITSI50517.2020.9264960.

[2] Hellberg AS, Hedström K. The story of the sixth myth of open data and open government. Transform Gov People Process Policy 2015;9(1):35–51. https://doi.org/10.1108/TG-04-2014-0013.

[3] Open Data Charter. ODC principles. ODC; 2015. https://opendatacharter.org/principles/.

[4] Janssen M, Charalabidis Y, Zuiderwijk A. Benefits, adoption barriers and myths of open data and open government. Inf Syst Manag 2012;29(4):258–68. https://doi.org/10.1080/10580530.2012.716740.

[5] van Loenen B, Zuiderwijk A, Vancauwenberghe G, Lopez-Pellicer FJ, Mulder I, Alexopoulos C, Magnussen R, Saddiqa M, Dulong de Rosnay M, Crompvoets J, Polini A, Re B, Casiano Flores C. Towards value-creating and sustainable open data ecosystems: a comparative case study and a research agenda. JeDEM - eJ eDemocr Open Gov 2021;13(2):1–27. https://doi.org/10.29379/jedem.v13i2.644.

[6] Ridgway J. Implications of the data revolution for statistics education. Int Stat Rev 2016;84(3):528–49. https://doi.org/10.1111/insr.12110.

[7] Audenhove LV, den Broeck WV, Mariën I. Data literacy and education: introduction and the challenges for our field. J Media Lit Educ 2020;12(3):1–5. https://doi.org/10.23860/JMLE-2020-12-3-1.

[8] Wolff A, Cavero Montaner JJ, Kortuem G. Urban data in the primary classroom: bringing data literacy to the UK curriculum. J Community Inform 2016;12(3):57—82. https://doi.org/10.15353/joci.v12i3.3278.

[9] Bhargava R, Deahl E, Letouzé E, Noonan A, Sangokoya D, Shoup N. Beyond data literacy: reinventing community engagement and empowerment in the age of data. Data-Pop Alliance White Paper Series. Data-Pop Alliance (Harvard HumanitarianInitiative, MIT Lab and Overseas Development Institute) and Internews; 2015.

[10] Pollock R. Building the (Open) data ecosystem. September 13. Open Data Foundation; 2011. http://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/.

[11] Harrison TM, Pardo TA, Cook M. Creating open government ecosystems: a research and development agenda. Future Internet 2012;4(4):900–28. https://doi.org/10.3390/fi4040900.

[12] Conradie P, Choenni S. On the barriers for local government releasing open data. Gov Inf Q 2014;31:S10–7. https://doi.org/10.1016/j.giq.2014.01.003.

[13] Kassen M. A promising phenomenon of open data: a case study of the Chicago open data project. Gov Inf Q 2013;30(4):508–13. https://doi.org/10.1016/j.giq.2013.05.012.

[14] Celis Vargas A, Magnussen R, Mulder I, Larsen B. Towards a framework for Open Data literacy in education: a systematic mapping review of open data skills and learning approaches. Interact Des Archit 2023;57:133–51. https://doi.org/10.55612/s-5002-057-008.

[15] Reggi L, Dawes SS. Creating open government data ecosystems: network relations among governments, user communities, NGOs and the media. Gov Inf Q 2022;39(2). https://doi.org/10.1016/j.giq.2022.101675. Scopus.

[16] Saddiqa M, Larsen B, Magnussen R, Rasmussen LL, Pedersen JM. Open data visualization in Danish schools: a case study. J WSCG 2019;2019(WSCG2019CS):17–26. https://doi.org/10.24132/CSRN.2019.2902.2.3.

[17] Saddiqa M, Magnussen R, Larsen B, Pedersen JM. Digital innovation in education: Perspectives, opportunities and challenges of educational open data and sensor data. CEUR Workshop Proc 2021;2991:74–83.

[18] Coughlan T. The use of open data as a material for learning. Edu Tech Res Develop 2020;68(1):383–411. https://doi.org/10.1007/s11423-019-09706-y.

[19] Pellegrino M, Antelmi A. At school of open data: a literature review. In: Proceedings of the 15th international conference on computer supported education; 2023. p. 172–83. https://doi.org/10.5220/0011747500003470.

[20] van Loenen B, Vancauwenberghe G, Crompvoets J, editors. Open data exposed. Open data exposed, 30. T.M.C. Asser Press; 2018. https://doi.org/10.1007/978-94-6265-261-3.

[21] Welle Donker F, van Loenen B. How to assess the success of the open data ecosystem? Int J Digit Earth 2017;10(3):284–306. https://doi.org/10.1080/17538947.2016.1224938.

[22] Capdarest-Arest N, Navarro CE. Promoting health data fluency skills by expanding data and informatics work in libraries: the role of a health library informaticist.

Med Ref Serv Q 2021;40(1):130–8. https://doi.org/10.1080/02763869.2021.1873648.
[23] Raffaghelli JE. Is data literacy a catalyst of social justice? A response from nine data literacy initiatives in higher education. Educ Sci 2020;10(9):1–20. https://doi.org/10.3390/educsci10090233.
[24] Celis Vargas A, Magnussen R, Mulder I, Larsen B. Learning designs for developing open data competencies in elementary school. In: Janssen M, Crompvoets J, Gil-Garcia JR, Lee H, Lindgren I, Nikiforova A, Viale Pereira G, editors. Electronic government. Electronic government, 14841. Springer Nature Switzerland; 2024. p. 354–70. https://doi.org/10.1007/978-3-031-70274-7_22.
[25] Celis Vargas A, Magnussen R, Larsen B, Mulder I. Open data learning designs in elementary school: defining the essential elements for developing open data competencies. Inf Polity 2025:15701255241297863. https://doi.org/10.1177/15701255241297863.
[26] Deahl ES. Better the data you know: developing youth data literacy in schools and informal learning environments. SSRN Electron J 2014. https://doi.org/10.2139/ssrn.2445621.
[27] Vahey P, Rafanan K, Patton C, Swan K, Van 'T Hooft M, Kratcoski A, Stanford T. A cross-disciplinary approach to teaching data literacy and proportionality. Educ Stud Math 2012;81(2):179–205. https://doi.org/10.1007/s10649-012-9392-z.
[28] Sarathy V. Real world problem-solving. Front Hum Neurosci 2018;12:261. https://doi.org/10.3389/fnhum.2018.00261.
[29] Atenas J, Havemann L, Timmermann C. Critical literacies for a datafied society: academic development and curriculum design in higher education. Res Learn Technol 2020;28:1–14. https://doi.org/10.25304/rlt.v28.2468.
[30] Wolff A, Knutas A, Palacin V, Seidelin C, Kun P, Mulder I. Civic data literacies for bottom-up data innovation. In: Proceedings of the 9th international conference on communities & technologies–transforming communities; 2019. p. 306–9. https://doi.org/10.1145/3328320.3328381.
[31] Dede C. Comparing frameworks for 21st century skills. 21st century skills: rethinking how students learn. Solution Tree Press; 2010. p. 51–76.
[32] Wolff A, Gooch D, Montaner JJC, Rashid U, Kortuem G. Creating an understanding of data literacy for a data-driven society. J Community Inform 2016;12(3):9–26. https://doi.org/10.15353/joci.v12i3.3275.
[33] Dewey J. Logic: the theory of inquiry. Henry Holt and Company; 1938.
[34] Dewey J. How we think: a restatement of the relation of reflective thinking to the educative process. Heath & Co Publishers; 1933.
[35] Rusmann A, Ejsing-Duun S. When design thinking goes to school: a literature review of design competences for the K-12 level. Int J Technol Des Educ 2022;32(4):2063–91. https://doi.org/10.1007/s10798-021-09692-4.
[36] Biesta GJJ, Burbules NC. Pragmatism and educational research. Rowman & Littlefield; 2003.
[37] Ejsing-Duun S, Skovbjerg HM. Design as a mode of inquiry in design pedagogy and design thinking. Int J Art Des Educ 2019;38(2):445–60. https://doi.org/10.1111/jade.12214.
[38] Fougt SS, Misfeldt M, Shaffer DW. Realistic authenticity. J Interact Learn Res 2019;30(4):477–504. *AACENumber*, https://www.learntechlib.org/primary/p/184664/.
[39] Herrington J, Reeves TC, Oliver R. Authentic learning environments (Eds.). In: Spector JM, Merrill MD, Elen J, Bishop MJ, editors. Handbook of research on educational communications and technology. Springer New York; 2014. p. 401–12. https://doi.org/10.1007/978-1-4614-3185-5_32.
[40] Clinton G, Rieber LP. The studio experience at the University of Georgia: an example of constructionist learning for adults. Educ Technol Res Dev 2010;58(6):755–80. https://doi.org/10.1007/s11423-010-9165-2.
[41] Oh, E. (2011). Collaborative group work in an online learning environment: a design research study [Dissertation].
[42] Shaffer DW, Resnick M. "Thick" authenticity: new media and authentic learning. J Interact Learn Res 1999;10(2):195–215.
[43] Nachtigall V, Shaffer DW, Rummel N. Stirring a secret sauce: a literature review on the conditions and effects of authentic learning. Educ Psychol Rev 2022;34(3):1479–516. https://doi.org/10.1007/s10648-022-09676-3.
[44] Celis Vargas A, Papageorgiou G, Magnussen R, Larsen B, Mulder I. The open data newsroom: a game approach for developing open data competencies in elementary school. In: Proceedings of the 18th European conference on games based learning; 2024. p. 197–206. https://doi.org/10.34190/ecgbl.18.1.2637. Vol. 18 No. 1 (2024).
[45] Gee JP. What video games have to teach us about learning and literacy. Palgrave Macmillan; 2004 (1. paperback ed).
[46] Greeno JG, Gresalfi MS. Opportunities to learn in practice and identity. In: Moss PA, Pullin DC, Gee JP, Haertel EH, Young LJ, editors. Assessment, equity, and opportunity to learn. 1st ed. Cambridge University Press; 2008. p. 170–99. https://doi.org/10.1017/CBO9780511802157.009.
[47] Shaffer DW, Squire KR, Halverson R, Gee JP. Video games and the future of learning. Phi Delta Kappan 2005;87(2):105–11. https://doi.org/10.1177/003172170508700205.
[48] Bransford JD, Brown AL, Cocking RR. How people learn: brain, mind, experience, and school: expanded edition. National Academies Press; 2000. p. 9853. https://doi.org/10.17226/9853.
[49] Brown JS, Collins A, Duguid P. Situated cognition and the culture of learning. Educ Res 1989;18(1):32–42. https://doi.org/10.3102/0013189X018001032.
[50] Hanghøj, T., Misfeldt, M., Fougt, S.S., & Bundsgaard, J. (2018). Unpacking the domains and practices of game-oriented learning. 10.13140/RG.2.2.25732.37764.
[51] The Open Data Newsroom. (2024). The open data newsroom. https://odeco-newsroom.myjuicebox.io/a/new_the_open_data_newsroom_copy/.
[52] European Environment Agency. (2023). Bathing water directive–status of bathing water, 2022 (Version 01.00) [Microsoft Excel (.xls, .xlsx)]. [object Object]. 10.2909/5D9A4D94-511A-486D-AFBB-4F01E5C73E23.
[53] Bakker A. Design research in education: a practical guide for early career researchers. Routledge; 2018. https://doi.org/10.4324/9780203701010.
[54] Brown A. Design experiments: theoretical and methodological challenges in creating complex interventions in classroom settings. J Learn Sci 1992;2(2):141–78. https://doi.org/10.1207/s15327809jls0202_2.
[55] Cobb P, Confrey J, diSessa A, Lehrer R, Schauble L. Design experiments in educational research. Educ Res 2003;32(1):9–13. https://doi.org/10.3102/0013189X032001009.
[56] Wang F, Hannafin MJ. Design-based research and technology-enhanced learning environments. Educ Technol Res Dev 2005;53(4):5–23. https://doi.org/10.1007/BF02504682.
[57] Kali, Y., & Hoadley, C. (2020). Design-based research methods in CSCL: calibrating our epistemologies and ontologies. 10.13140/RG.2.2.31503.20642.
[58] Ejersbo LR, Engelhardt R, Frølunde L, Hanghøj T, Magnussen R, Misfeldt M. Balancing product design and theoretical insight. The handbook of design research methods in education: innovations in science, technology, engineering and mathematics learning and teaching. 1st ed. Lawrence Erlbaum Associates; 2008. p. 149–63.
[59] Creswell JW, Creswell JD. Research design: qualitative, quantitative, and mixed methods approaches. 3rd ed. Sage; 2018.
[60] Onwuegbuzie A, Leech N. Linking research questions to mixed methods data analysis procedures 1. Qual Rep 2015. https://doi.org/10.46743/2160-3715/2006.1663.
[61] Kvale, S. (1996). InterViews: an introduction to qualitative research interviewing (p. 326). Thousand Oaks.
[62] Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77–101. https://doi.org/10.1191/1478088706qp063oa.
[63] Bingham AJ. From data management to actionable findings: a five-phase process of qualitative data analysis. Int J Qual Methods 2023;22:16094069231183620. https://doi.org/10.1177/16094069231183620.
[64] Hayes N. Doing psychological research. 2nd ed. Open University Press; 2021.
[65] Patton MQ. Qualitative research & evaluation methods. Sage; 2010. 3. ed., [Nachdr.].
[66] Attride-Stirling J. Thematic networks: an analytic tool for qualitative research. Qual Res 2001;1(3):385–405. https://doi.org/10.1177/146879410100100307.
[67] Seymoens T, Van Audenhove L, Van Den Broeck W, Mariën I. Data literacy on the road: setting up a large-scale data literacy initiative in the DataBuzz project. J Media Lit Educ 2020;12(3):102–19. https://doi.org/10.23860/JMLE-2020-12-3-9.
[68] Atenas J, Havemann L, Priego E. Open Data as Open Educational Resources: Towards Transversal Skills and Global Citizenship. Open Praxis 2015;7(4):377–89. https://doi.org/10.5944/openaxis.7.4.233.
[69] Prodromou T, editor. Data visualization and statistical literacy for open and big data. IGI Global; 2017. https://doi.org/10.4018/978-1-5225-2512-7.
[70] Vande Moere A, Purchase H. On the role of design in information visualization. Inf Vis 2011;10(4):356–71. https://doi.org/10.1177/1473871611415996.
[71] Walny J, Huron S, Carpendale S. An Exploratory Study of Data Sketching for Visual Representation. Comp Grap Forum 2015;34(3):231–40. https://doi.org/10.1111/cgf.12635.
[72] Veljković N, Bogdanović-Dinić S, Stoimenov L. Benchmarking open government: an open data perspective. Gov Inf Q 2014;31(2):278–90. https://doi.org/10.1016/j.giq.2013.10.011.
[73] Bakker A, Van Eerde D. An introduction to design-based research with an example from statistics education. In: Bikner-Ahsbahs A, Knipping C, Presmeg N, editors. Approaches to qualitative research in mathematics education. Springer Netherlands; 2015. p. 429–66. https://doi.org/10.1007/978-94-017-9181-6_16.