International Journal of Engineering Education Vol. 38, No. 3, pp. 756–773, 2022
Printed in Great Britain

0949-149X/91 $3.00+0.00

Development and Evaluation of an Approach for Integrating Data Science Concepts into High School STEM Curriculum*

VYTAUTAS ŠTUIKYS, RENATA BURBAITĖ, GIEDRIUS ZIBERKAS and RAMŪNAS KUBILIŪNAS
Software Engineering Department, Kaunas University of Technology, Studentu 50, Kaunas, Lithuania. E-mail: vytautas.stuikys@ktu.lt, renata.burbaite@ktu.lt, giedrius.ziberkas@ktu.lt, ramunas.kubiliunas@ktu.lt

Currently STEM education evolves rapidly towards a higher integration either among separate constituents (Science, Technology, Engineering and Mathematics) or by adding the new aspects, such as social. This paper aims at extending the science dimension in STEM-driven Computer Science (CS) education by integrating (through modelling) Data Science (DS) concepts into the high school curriculum. Three types of models and modelling (conceptual, feature-based, and physical modelling) incorporated into a coherent methodology form the background of the proposed approach. Models and modelling, as well processes with data, are essential attributes of engineering education too. The core result of this paper is a novel three-layered framework outlining a series of modelling processes to support integration along with the assessment model. The latter includes the Revised Bloom's taxonomy combined with computational and scientific thinking skills. The use of this approach in the real educational setting and provided experiments show that discovered models ensure a seamless integration of the DS component into STEM-driven CS education. This approach contributes to the increased students' motivation to learn due to the interesting real-world task and active learning of engineering aspects through constructing and testing own experimental system and data processing. The approach also enforces the learner's interdisciplinary knowledge by computational, scientific, and designing skills so important for engineering activities.

Keywords: STEM; computer science; data science; modelling and integration

1. Introduction

The Next Generation Science Standards (NGSS) direct educators towards finding meaningful and engaging ways to teach science content through incorporating engineering practices as well as computing and computational thinking [1]. More specifically, science and engineering are an integral part of STEM (Science, Technology, Engineering and Mathematics) education, one of the most striking educational movements in recent years worldwide [2]. The STEM movement has emerged as an interdisciplinary approach involving teaching science, technology, engineering, and mathematics under one roof. Its basic concept is to provide teaching and learning through real-world tasks solving with the focus on obtaining the interdisciplinary knowledge and skill through inquiry-based pedagogy [3]. In the digital age, the integrated interdisciplinary knowledge is of the highest value. It is possible to get these by studying STEM disciplines. Therefore, STEM educational research and practice evolve towards higher-level of efficiency and widening the scope of integration [4–8].

In this context, Computer Science (CS) education stands as one of the most important components for integrated STEM education. It is so, because CS is a cross-disciplinary field and brings computing knowledge practically needed for all. In addition, CS is a catalyst of providing computational thinking skills that, according to J. Wing, "will be a fundamental skill used by everyone in the world by the middle of this century" [9]. Therefore, since 2015 CS has become an integral part of STEM education in the USA educational system [10].

STEM components and many other fields outside STEM (if not all) rely on problem solving. No matter how problems are different, one common attribute combines all fields while we start dealing with problem solving. That attribute is data. In fact, any human activity relies on using data. What is happening now due to the ever-growing technological advancements is the continuous increase in the amount and accessibility to data. Therefore, we are gradually entering into the era of massive sets of data called "big data" now [11]. Before using, data should be first extracted, collected, stored, analysed, classified, and processed in a variety of ways to be useful for multiple applications. Practically those activities are independent upon applications, and they therefore are a matter of Data Science (DS). DS is an interdisciplinary subject that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from a variety of structural and unstructured data [12]. In addition, DS uses theories and techniques taken

* Accepted 6 December 2021.

from many fields such as mathematics, statistics, CS and others. Thus, the relationship among STEM, CS and DS is evident. Papers [13, 14] treat DS as a "fourth paradigm" of science (empirical, theoretical, computational and now data-driven) and assert, "Everything about science is changing because of the impact of information technology" and the data deluge.

In the context of the ever-growing role of data and its inherent relationships to science, engineering, and education, it is highly important to extend the integrative aspects of STEM education by introducing the DS concepts and practices into K-12 classrooms. To do that, in this paper, we discuss an approach and its implementation in the real educational setting. The proposed approach also includes a methodology based on conceptual modelling combined with feature-based and physical modelling. Note that modelling is the core in STEM, CS and science educational research and practice [15–17] as well in engineering education [7, 8]. The aim of this paper is to extend and enforce the previously researched STEM-driven CS education (presented in the book [15]) by adding the core concepts taken from DS. The contribution of this paper is two-fold: (i) the general framework for the seamless integration of DS concepts within processes of STEM-driven CS education and (ii) a set of modelling processes and model-driven components to implement the proposed framework in practice.

Next, we analyse the related work.

2. Related Work

Here, we focus on most essential aspects of integrated STEM and CS education in the context of K-12 (Stream A) and Data Science (DS) topics (Stream B).

Stream A. Nadelson & Seifert [18] define integrated STEM as the seamless amalgamation of content and concepts taken from different STEM disciplines. The integration appears in ways such that "knowledge and processes of the specific STEM disciplines are considered simultaneously without regard to the discipline, but rather in the context of a problem, project, or task". The report prepared by National Academy of Engineering and National Research Council (USA) [19] defines broadly the integrated STEM education in the context of K-12 through four general features or subcomponents (*Goals*, *Outcomes*, *Nature and Scope of Integration* and *Implementation*). These features, while considered in practice, are to be interdependent.

Hasanah [20] provides a literature review that focuses on four key definitions based on the selected studies: (a) STEM as discipline, (b) STEM as instruction, (c) STEM as field, and (d) STEM as a career. According to the author, STEM discipline is a fundamental part of STEM education because most of the initiatives in STEM education would be related to the disciplines. Kubat & Guray [21] assume that these four disciplines are to be taught as a holistic and an undistinguished collective, rather than teaching these four disciplines independently. Thibaut et al. [22] propose the framework containing five key principles: integration of STEM content, problem-centred learning, inquiry-based learning, design-based learning and cooperative learning. According to the authors, this framework has several benefits, such as its applicability in the classroom and the possibility to describe integrated STEM on multiple dimensions. In addition, this paper calls for further research to investigate the effects of integrated STEM on students' cognitive and affective learning outcomes.

Hobbs et al. [23] reveal four models of STEM implementation based on the discipline paradigm: (1) "Four STEM disciplines are taught separately"; (2) "Teaching all four but more emphasis on one or two"; (3) "Integration at least three disciplines"; (4) "The integration of all four subjects by a teacher".

Typically, acquiring of scientific knowledge goes through solving problems that are feasible, worthwhile, contextualized, meaningful, ethical, and sustainable [24]. In addition, solving of authentic problems transcends a single discipline. Hence, problem solving is an activity intrinsic to many (if not all) domains and, thus, it can serve as a general and common approach to teaching. Therefore, problem-solving must be seen as a multidisciplinary challenge along with the corresponding practices and processes, for example, in STEM education, in which different domains like science, technology, engineering, mathematics and computer science are involved simultaneously. To do that systematically, Priemer et al. [25] present a fine-grained, integrated, and interdisciplinary framework of problem solving for education in STEM and CS by cumulatively including ways of problem solving from all of these domains. This framework includes twelve activities represented as processes within the given flowchart to foster students' problem-solving competences. These competences (e.g., to identify problems, to review related information, to develop and evaluate options, to implement solutions and many others) are often seen as important twenty-first-century skills [26] along with computational and scientific thinking [9].

Yadav et al. [27] discuss the key computational thinking (CT) constructs, such as algorithms, abstraction, and automation in the context of provided educational reforms (Common Core and

Next Generation Science Standards). In addition, this paper provides specific means that would allow teachers to embed these ideas in their K-12 classrooms, including recommendations for instructional technologists and professional development experts for infusing CT into other subjects. According to this paper, CT ideas are key to moving students from merely being technology-literate to using computational tools to solve problems.

The report [28] describes the policy trends and national (the USA) momentum over the past 12 months (meaning the year 2018) in computing education. It contains (i) an analysis of national and state trends in advanced placement (AP) computer science (CS) participation by gender and race, including the relationship with policies; (ii) a policy summary for each of the nine policies displaying a map of the states that have enacted the policy, including highlighted states and related resources.

The next few works deal with models and modelling more specifically. Hallström and Shornbörn [16] call for the reinforcement of models and modelling for authentic integrated STEM education. It is so because models and modelling are important tools for problem solving, prediction, decision making, and communication and have been studied in various fields, especially in science, mathematics, and engineering. In addition, the authors provide an extensive analysis of types, roles, functions, strategies, and recommendations for using models and modelling in the context of authentic STEM education and literacy. Authors indicate on the following types of models used: the concrete; the verbal; the symbolic; the visual; the gestural and physical. However, this list of models is not comprehensive. In the context of our research, it should be mentioned feature-based variability modelling in e-learning [29–31]. Note that this approach has been borrowed from software engineering. One can get more knowledge on pedagogical aspects of using models and modelling in [32] and conceptual modelling in [33].

Stream B. Nowadays, we are witnessing a new wave of the technological revolution resulting in producing an innumerable quantity of data that requires sophisticated tools and skills to be adequately analysed, processed, and interpreted. The Data Science (DS) community describes this situation broadly by the term "data deluge" [13, 14]. Researchers define DS differently either as "an interdisciplinary subject that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from a variety of structural and unstructured data" [12], or as an emerging area that "involves principles, processes, and techniques for understanding phenomena via the (automated) analysis of data." [34]. Typically, DS uses theories and techniques drawn from many fields, including mathematics and computer science. Job opportunities in this sector are currently booming and this growth is expected to continue during the next years [35]. Considering these opportunities for young generation, educational experts are now investigating different ways to introduce DS in schools [36]. However, research concerned with the teaching of DS or Big Data (BD) concepts in secondary education is currently in its early stages. As indicated by [36–38], the studies are more extensive in universities, business schools or other profession-oriented institutions. Nonetheless, experts try to formulate examples of DS curricula for schools and address the different aspects involved in teaching DS. In a narrow sense, DS involves a combination of three scholastic subjects, i.e. mathematics, statistics, and computer science. Therefore, [39–41] present DS as an additional component of these subjects, providing an opportunity to modernize their content and giving to students an opportunity to understand how BD affects their own life. Putting this into practice, however, has revealed some considerable shortcomings. The first arises from teachers' competences since they appear to lack statistical thinking. Thus, a proper training has been suggested [42] and adopted in one pilot project ("The Mobilize Introduction to Data Science (IDS)") [39]. Nonetheless, the lack of a statistical mind set should not be underestimated, as other studies noted a weak scientific and statistical reasoning also in students. It appears that secondary school students struggle with interpreting and using statistical information, understanding variability in data, and extracting pertinent information from graphical representations [43]. To overcome these barriers, it is important to develop and adopt proper teaching models. For example, the PPDAC (Problem – Plan – Data – Analysis – Conclusion) cycle has been adopted as a teaching method in the previously mentioned IDS project, but it has proven quite challenging for both teachers and students [39].

Another aspect of teaching DS is its relationship with CS, which provides the necessary toolbox to perform any DS related activity. Obviously, when it comes to design a DS curriculum for secondary schools, it is important to know which software should be used for teaching DS in schools in this regard. Decisions can be based on which notions or skills students are expected to learn, hoping that teachers have the adequate digital skills [44]. This consideration leads to the question whether DS classes should be aimed at teaching more technical/digital skills, or they should be focused more on teaching the concepts? Most projects and curricula designed seem to reach for an even balance between theory and practice, as all of the DS implementations analysed devote considerable time to building the knowledge and developing the adequate skills [41, 45]. However, the Royal Society report presents the potential for DS to improve teaching also in other subjects beside CS, i.e. statistics and informatics [42]. In fact, to enlarge the benefits of teaching DS in schools, a good idea is connecting DS teaching with other data-driven subjects, such as biology, history, and geography. In this way, DS could also provide larger benefits in promoting a more data-driven culture by showing the application of mathematics to specific questions in science. Other studies highlight the importance of balancing the teaching of data applications with other modules concerned with societal and ethical issues [46], in order to make students aware of the many issues surrounding BD (privacy, confidentiality, transparency, identity, etc.) and to spread a culture of responsible data use. Discussing about DS in schools could be also an opportunity to raise more aware citizens who understand today's technology and know-how to point to their unethical applications, which, as some publications have shown, can be hard to see and understand [47]. Students should be aware about benefits of teaching and learning DS in high school. That increases motivation to learn. One factor is that data scientists are among the most requested professionals along with STEM in the job market. However, this is relevant for presenting DS not only as a job outlet, but also as a steady and exciting subject. Now the most crucial areas of data-centric applications relate to security, energy, social wellbeing, and health issues.

These are just some of many examples proving how pervasive BD and DS have become the reality in our surrounding. By incorporating DS into other data-driven subjects, new insights may be investigated, enriching the general teaching offer. Students, for example can also examine the threats of wildfires or the ways social media are tracking their data, learning how to apply math to real-world issues [48]. The possibilities are endless indeed. This multidisciplinary and the far-reaching applications make DS also an interesting subject of investigation.

This overview of the related work, as presented in each category, is by no means comprehensive. However, we treat this analysis as sufficient to support and motivate our approach by summarizing the following.

(1) All analysed educational fields (STEM, Computer Science, and Data Science) provide students with knowledge and skills that are of the highest value for current and future young generations in terms of their professional development and fulfilling the social-economic needs.

(2) All fields are multidisciplinary in nature; all rely on real-world problem solving; all of them have the wide capabilities to develop and enhance such skills and competences as computational thinking, scientific thinking, algorithmic thinking, and design thinking.

(3) The inherent interdependency among these fields and revealed properties create good preconditions for their integration into a single course. Therefore, the findings of provided analysis encourage and motivate our approach towards extending the integrated STEM driven CS education with the Data Science concepts.

3. Background and Motivation

We aim at broadening the previously researched approach presented in the book [15]. The basis of that approach is the use of robotics and smart devices (sensors, cameras, etc.) possibly connected to robots to form the smart learning environment (SLE). The basic teaching content is robot's control programs (RCP) to provide learning and teaching of the CS programming course, by dealing with real world tasks (their prototypes) and using design-based, inquiry-based, and project-based pedagogy. Students are engaged in constructing robots from available parts, developing/modifying RCP, modelling the functionality of characteristics of smart devices, representing modelling results. By providing these activities, students can get different knowledge, identified as S-knowledge, T-knowledge, E-knowledge, and M-knowledge. Typically, the S-knowledge is from physics (such as obtaining physical characteristics of different sensors) and CS (such as data, algorithms, and RCP). The T-knowledge includes properties and characteristics of robots such as LEGO [49–50], or micro controllers such as ARDUINO [51]. E-knowledge is from mechanical, electrical and software engineering. M-knowledge comes when students are engaged in formulating real-world tasks, representing project outcomes in their reports, etc. By investigating the S-, T-, E- and M-components, typically students work with *structured data*, such as an instructive material on how to construct a robot, predefined values of robots' physical characteristics (speed, voltage). Structured data has well-defined format such as computer's or robot's (it has computing and processing units too) instruction to form a program in RobotC or other language. However, real-world tasks and huge number of applications, in their initial state are concerned with a large amount of the *raw data*. Extracting this data from the environment or existing ecosystems for building applications (in a variety of fields) is a primary concern. It may be realized, for example, using the same smart devices (programmable sensors) taken from the smart educational environment [15]. On the other hand, this process along with the following activities (for collecting, storing, transforming, and analysing) is a matter of Data Science (DS). The latter typically relies on using and exploring a huge amount of data identified as Big Data (BD). Therefore, using physical components (sensors and their supporting hardware and software facilities) from the previously developed environment and new ones such as tools for data transforming and analysis, we can introduce a new STEM component, identified as DS-component (meaning Data Science concepts) to enrich each (S-, T-, E-, M-component) here. Thus, we can extend the STEM paradigm by the possibility to deliver the DS-knowledge, the basic concepts of DS.

4. Research Methodology

The research methodology we use here relies on using models and modelling as a cornerstone approach to deal with the integrated STEM educational research and practice. In this regard, we present the following excerpt from the paper [16].

> "Models and modelling processes can bridge the gap between STEM disciplines through authentic practices. Models and modelling should be used as a means to promote STEM literacy and the transfer of knowledge and skills between contexts, both in and out of the STEM disciplines. Modelling activities can serve as a meaningful route toward authentic STEM education. <. . .> If this vision is to be reinforced, it is of utmost importance that implementing any model-based authentic educational activities are underpinned by evidence-based frameworks and recommendations for teaching practice".

At the core of our approach are DS concepts modelling in the context of formulated research objectives. Modelling is a process of identifying properties and relationships among concepts using some techniques and approaches. There are many approaches and kinds of models to provide modelling (some indicated in [16], for others see related work). Regarding our previous work [15] and considering such properties as intuitiveness of the graphical notion to express learning variability, easiness of use, accessibility of tools for checking models' correctness, etc., we rely on using the feature notation and feature-based models [52], where they are relevant. Here, we combine three modelling approaches as ingredients of the proposed methodology (conceptual, feature-based, and physical modelling). By the proposed methodology we mean the development of the conceptual model and three-layered framework and its implementation through conceptual, feature-based, and physical modelling. We discuss them in detail as follows.

5. Conceptual Model for Introducing DS Concepts

The conceptual model explains how to introduce DS concepts for the inclusion them into STEM-driven CS curriculum topics to provide teaching and learning processes in K-12 classrooms. Analysis of the domains (DS, STEM, CS) and our previous research are prerequisites to define this model. It consists of two parts: STEM components and Data Science (DS) basic concepts. The first part includes components taken from the STEM acronym and identified as S-, T-, E-, and M-component here. In Fig, 1, there is the full name of each STEM component. The second part includes DS processes (data collecting, data storing, transforming, and analysing).

The core aspect of this model is the relationship among STEM components and DS processes. Table 1 explains this relationship. Note that there are two science components (S-component). We identify the first as *Science* and the other – as *Computer Science* (CS). Here, by Science we mean one, two or all from the list of subjects (Physics, Biology, and Chemistry). What aspects of these subjects we need to cover that depends on the real-world tasks and learning objectives? For example, human health monitoring tasks (e.g., pulse, temperature measurement, etc.) more relate to biology while tasks dealing with monitoring of the environment more relate to chemistry. Constructing an experimental system for collecting data is a matter of engineering. Note that the non-empty element of Table 1 defines this relationship by informally stating what activities are taken into consideration and therefore what kind of knowl-

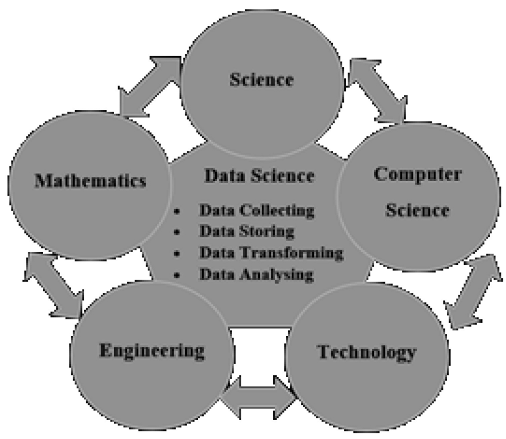

Fig. 1. Conceptual model for introducing DS concepts into STEM-driven CS education.

Table 1. DS concepts integration into STEM through real world task solving

DS	STEM components				
	(S-component)	**CS (S-component)**	**T-component**	**E-component**	**M-component**
Data Collecting	Origin of data & Data sources (context of a real-world task)	Data collection programming	Data collection technologies applied	Data collecting modelling and system designing	
Data Storing			Storage technologies applied	Data storage infrastructure modelling and system designing	
Data Transforming		Characteristics of data Transforming procedures			Representation of data transforming procedures
Data Analysing	Results of data analysis	Methods of data analysis	Data mining methods/ technologies applied		Methods of data analysis

Here, we present Table 1 formally by matrix $T1(i, j)$, where the index i ($i = 1 .. 4$) represents DS processes and the index j ($j = 1 .. 5$) – STEM components. Therefore, this model reveals what method, technology, or activity (again, implicitly at this level) the actor (teacher, student) needs to apply to solve the given task. Note that the way for the explicit knowledge delivery will appear gradually at the next stages as we explain in the following.

edge it is possible to extract (implicitly) and then to deliver for learning.

6. Three-Layered Framework

Here, we present the next part of our methodology-the three-layered framework (Fig. 2). It consists of three layers, i.e. *top*, *intermediate* and *implementation*. The top layer uses conceptual model (Fig. 1), curriculum requirements regarding DS topics and is responsible for the development of the DS model and SLE extension. The intermediate layer interpreted as "a generalized model of pedagogical domain", stands for identifying the real world task, models of learning activities and processes, including the assessment model development. Finally, the implementation layer describes the scenario on how the processes and activities are to be implemented. Two-sided fat arrows identify the possible feedbacks among layers. We define the internal part of each layer through a set of sub-processes enumerated from 1 to 8 in Fig. 2. We define each sub-process by its functionality along with adequate inputs and output. Typically, inputs are of two types within layers: from external sources (see, shifted arrows in Fig. 2) and internal ones, i.e. from the other sub-processes. We define the output of a process as the adequate model. The use of the framework starts with curriculum analysis, research objectives analysis, STEM and DS concepts domain tasks analysis and expected outcomes of the research. The framework ends with implemented processes and learning findings.

By this framework, however, we have still presented our approach conceptually. We clarify the real value of the proposed methodology and modelling approaches in the next few sections by explaining each process in more details. Knowing

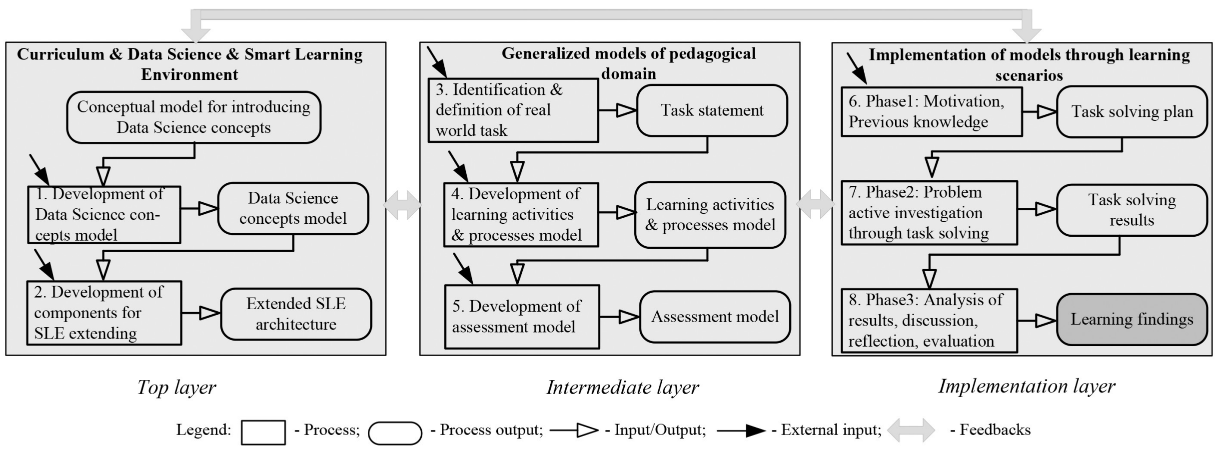

Fig. 2. A three-layered framework to present the proposed methodology.

these details, different stakeholders might be interested in using this methodology too, not only researchers and teachers. For example, in our view, the top-layer is most useful for educational strategists and curriculum designers to form an educational policy. The intermediate layer provides the useful information for teachers for curriculum improvements. The implementation layer is concerned with teachers and learners.

Next, we provide a detailed description of the proposed methodology.

7. Development of DS Model

As indicated previously, we rely on using feature-based models borrowed from SWE and CS domains. The main activity to build these models is domain analysis. FODA (Feature-Oriented Domain Analysis) is the primary source that introduces this approach [53]. In Fig. 3, we present the feature model to define the DS domain aspects that are relevant to our context. In general, feature model is the structural representation of a domain

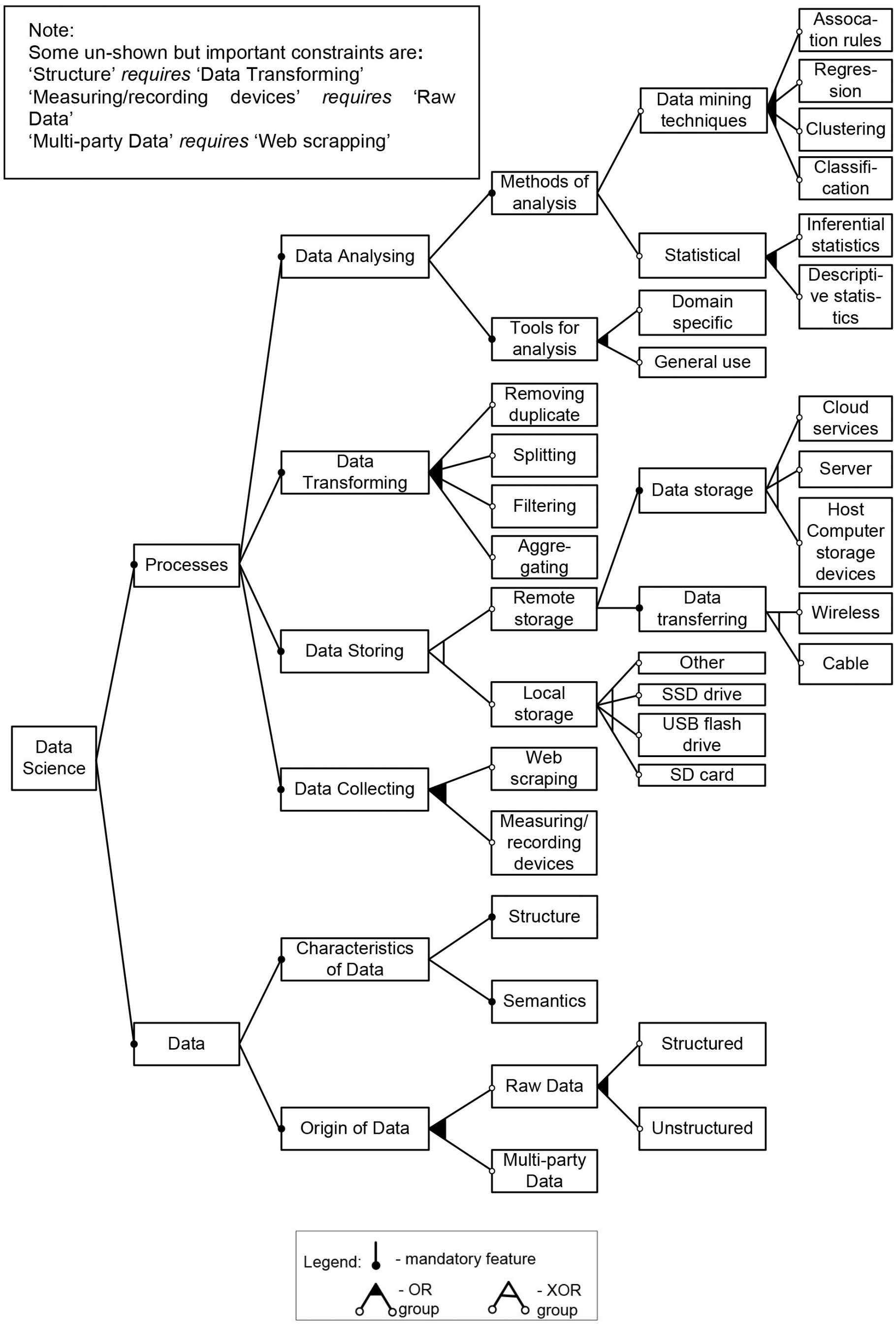

Fig. 3. Abstract feature-based model to represent Data Science (DS) domain semi-formally.

under consideration in the form of tree-like feature diagrams.

Feature is a distinguishing characteristic of a domain, though there are dozens of other definitions of this concept [52]. The latter property indicates on the universality of this approach to express the structural aspects of any domain. Features (boxes in the diagram) are arranged as parent-child relationships in the hierarchy to narrow the scope of domain aspects. One can find more information on feature types, constraints, and models also in [15, pp. 72–98].

Our feature model (see Fig. 3) is abstract, i.e. without the identification of concrete values of terminal nodes. This model is a pre-specified set of DS features in two categories (*Data* and *Processes*) along with relationships and constraints among features and/or sub-features. Note that we present constraints '*requires*' not graphically, but as a text (for better readability, see Note in Fig. 3). Therefore, this model outlines a common picture to understand this domain by educational strategists, curriculum designers and teachers. It is a semi-formal abstract representation of the topics for including them into a curriculum. Note that for doing so we yet need to add the concrete features through splitting the given ones. This model provides researchers, as well teachers and to some extent students, with the needed resources, though abstractly (measuring/recording devices, analysing tools, see adequate feature boxes in Fig. 3) in the route of solving a given task.

We outline that in the next section.

8. Components for extending SLE

Structurally, Smart Learning Environment (SLE) for smart STEM-driven CS education contains four large components: Robot-based learning environment, PC, Server and monitoring system [15, pp. 279–303]. Here, we have extended this environment with new capabilities to provide learning research activities in the Data Science (DS) field. In Fig. 4, we present these new capabilities within three large components. The made extension includes: (1) Hardware for data collecting and storing incorporated in the Robot-based learning environment as a white box; (2) Software for data transforming and analysing within the PC component and (3) Assessment facilities within Server, also as a white box.

We define the functionality of the extended environment through a set of external processes P1, . . . , P8 to provide the input/output stream among components. Note that numbering of the process here does not correspond to the one given in the initial SLE structure. Here, the indicated processes outline new functional capabilities of the extended environment. With the extended SLE, we have finished considering the top-layer of our framework (see Fig. 2). In the next section, we

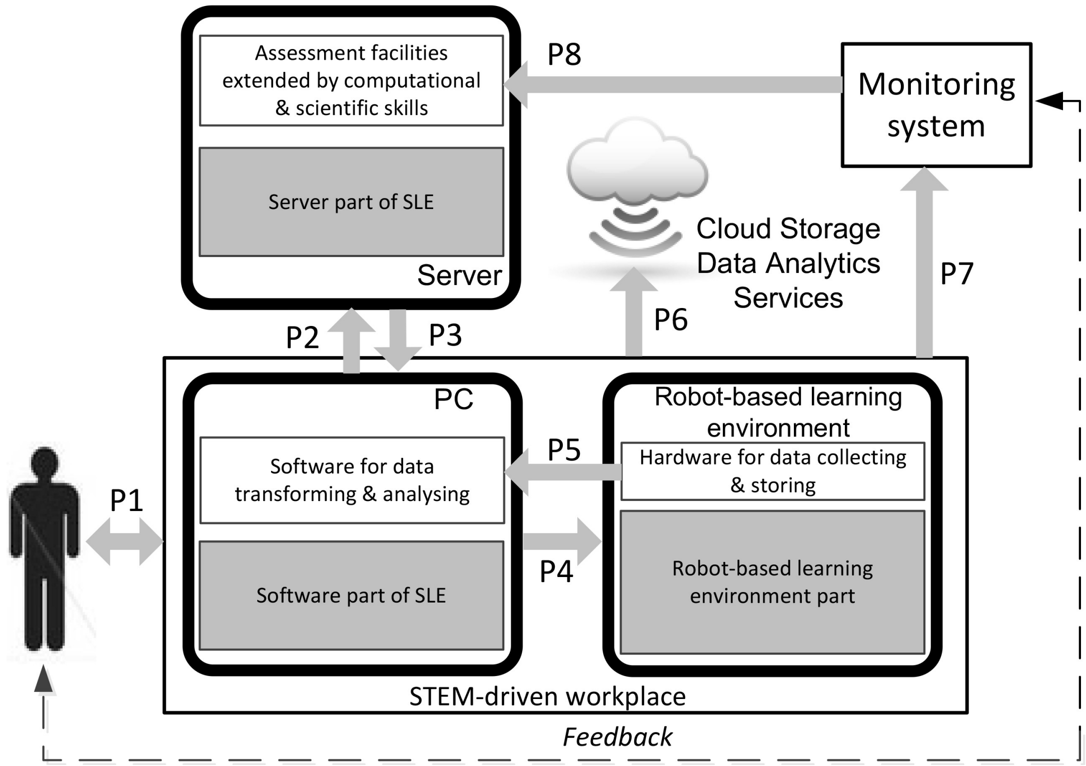

Fig. 4. Extended SLE with components for integrating DS concepts.

discuss the intermediate layer, i.e. modelling aspects as a cornerstone attribute of our approach.

9. Modelling Activities for the Development of Task Solution System

At this point, readers have a common understanding of the DS aspects through its model (Fig. 3) and the extended SLE with resources to deal with the basic DS concepts (Fig. 4). The next step is to put all this into action. As it is clear from the findings of multiple research [2, 3, 6, 30], STEM education focuses on solving the real-world tasks or their prototypes (in terms of learning and teaching capabilities). The same is with the DS education. In this case, however, the application to be investigated should contain the raw data in the necessary amount for applying DS approaches. Therefore, we start with a domain-specific task definition (in terms of data richness, Big Data (BD) concepts and specific measures to access to them) and the students' engagement in considering this task. We identify that within the first component as a task-solving plan in Fig. 5. In fact, this is the same as presented in our framework. The task-solving plan can be presented in a variety of ways. For example, it can be done through discussions on the weather pollution, its influence on human health and therefore the need of accumulating a huge amount of data and then providing processing and measuring its parameters. The other example may be related to the COVID-19 problem testing by explaining how it is important to measure the amount of oxygen in human's blood now worldwide.

Therefore, the task solving plan (among other items, it defines the necessary resources from the extended SLE) serves as the input to start modelling procedures. We identify the next bound of activities as *Task-Solution system modelling* (see Fig. 5). Indeed, if we want to solve the real-word task, we need first to build a system for that. The primary stage in doing so is modelling. Two basic attributes of any system are its structure and functionality. Here, by the system we mean hardware and software parts to cover the basic DS processes (data collecting, storing, transforming, and analysing; see also a conceptual model in Fig. 1). Note that data collecting and storing require both the adequate hardware and software, while data transforming/analysing require the only special software. One could see this property as adequate resources in Fig. 4. Here, we uncover this property in more detail. Process 2 (see Fig. 5) includes the system's functionality modelling resulting in the creation of the data collecting model. Process 3 includes the system's structure modelling resulting in the creation of the data storing model. How these models look like in practice, we disclose that in our case study. Data

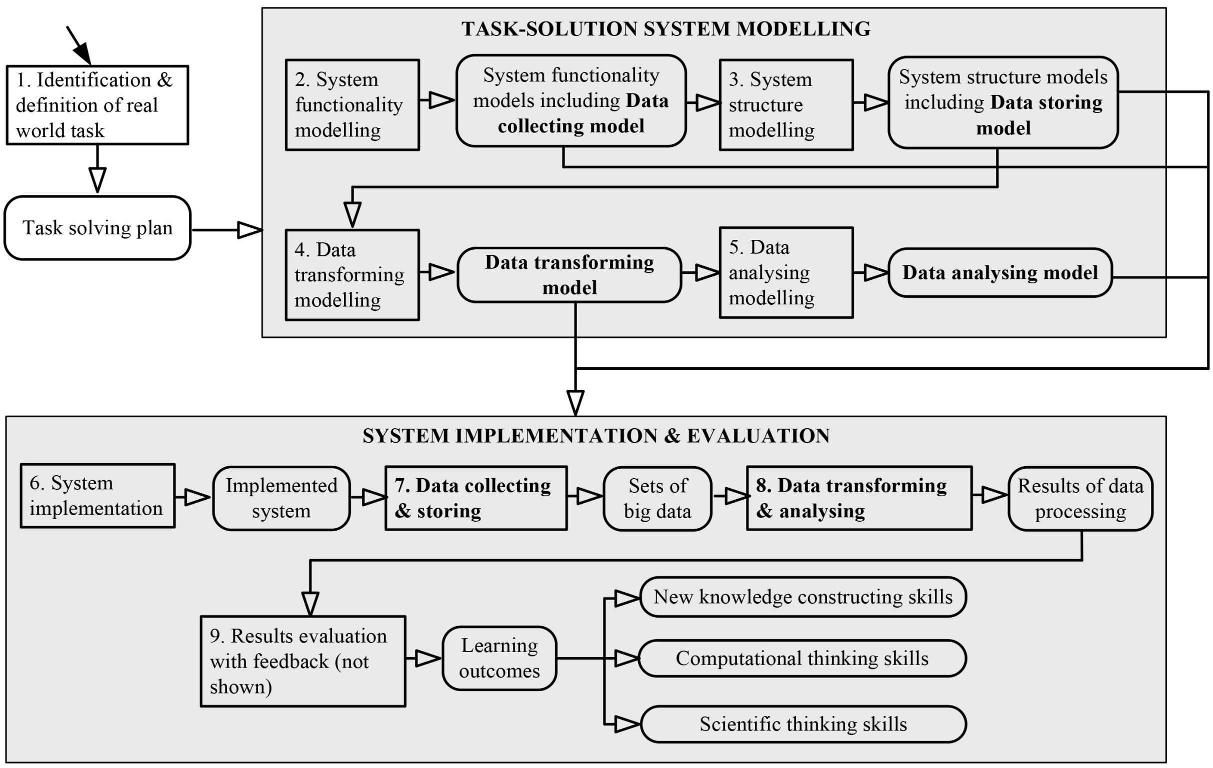

Fig. 5. General model of learning processes and activities as learning scenarios to focus on DS aspects according to the proposed framework.

transforming and analysing modelling (processes 4 and 5) rely on functional modelling only. It includes the selection of the transforming mode and adequate methods and tools (see Fig. 3). Each process (2–5) results in creating the adequate model. With these models, it is possible to implement the system (process 6). The implemented system in this way through modelling enables to realize DS processes (process 7 and 8) in the real education setting. However, the learning process starts far earlier, in fact, including all processes outlined in Fig. 5 starting from process 1 and ending with process 9. It is so because the teacher uses STEM-based pedagogy that includes a set of approaches (problem-based, learning, learning-by-doing, and inquiry-based learning by involving *active students* even in the Task-Solution system modelling, not only in implementation processes). We extend the discussion on this later, after introducing the case study.

At this stage, the essential activity is the identification what knowledge and skills students can obtain, and how to assess the efficiency of the learning process. As indicated in Fig. 5, learning outcomes include three entities: (1) new knowledge constructing skills, (2) computational thinking skills and (3) scientific thinking skills. Here, by new knowledge skills we mean the direct students' involvement in basic DS processes, their experience to provide modelling activities and to construct the system, their experience in using novel data processing methods and tools, their capabilities to repeat experiments on their own pace and relevant time. Scientific thinking skills, among others, include developing and using models, planning, and providing experimental investigation, evaluating evidence of the outcomes [54]. Computational thinking skills, on the other hand, cover the knowledge about abstraction, decomposition, generalization, data representation, and algorithms [55]. The model-driven methodology discussed so far along with STEM-based pedagogy applied are the key attributes to achieving these skills. However, there is a question: how to evaluate the extent of this knowledge students can achieve? In our approach, for assessing learning outcomes (process 9), we focus on the specific assessment model to test computational thinking and scientific skills as described in the next section.

10. Assessment Model Development

To design the assessment model, we need to possess the relevant constituents for that. The first constituent, as a basis for assessment of learners' knowledge, is the revised Bloom's taxonomy [56]. It focuses on representing any learning objective and outcomes in two dimensions, i.e. *Knowledge* and *Cognitive Process*. The Knowledge dimension consists of (1) factual knowledge, (2) conceptual knowledge, (3) procedural knowledge and (4) metacognitive knowledge. Factual knowledge includes the terminology and specific details and elements. Conceptual knowledge covers classifications and categories, principles and generalizations, theories, models, and structures. Procedural knowledge defines subject-specific skills and algorithms, techniques and methods, criteria for using appropriate procedures. Metacognitive knowledge "is the knowledge of one's own cognition and about oneself in relation to various subject matters" [56]. The listed knowledge types range from concrete to abstract.

The Cognitive Process dimension introduces a continuous sequence of cognitive complexity. Cognitive processes are divided into six categories – from the lower-order to the higher-order skills (*Remembering*, *Understanding*, *Applying*, *Analysing*, *Evaluating* and *Creating*). The next constituents are skills, such as *Computational Thinking* (CT), students able to acquire through the learning process. For example, [57] and [55] express CT skills as a set of basic concepts (*Abstraction*, *Decomposition*, *Generalisation/ Pattern recognition*, *Data representation* and *Algorithm*). On the other hand, Scientific Thinking (ST) skills include (Forming and Refining Hypotheses, Developing and using models, Investigation Skills, and Evaluating Evidence) [54].

In general, the assessment has a structural part and process-based part. The first part consists of constituents mentioned above that represent the generic aspects of any knowledge assessment. In addition, it may include specific aspects related, for example, with the real-world problem (e.g., complexity, environmental factors, etc.). The process-based part covers the assessment methods (observation, questioning, report, product review, structured activities (projects, scenarios, case study, group discussion), self-assessment, etc.) for realizing the knowledge evaluation in practice.

The next step is to put all aspects together and represent their relationships. In Fig. 6, we propose the explicit knowledge assessment model represented using the feature-based notation.

The root of the tree always represents the whole entity (in our case – the assessment model). At the top level (we mean the next to the tree's root, see Fig. 6), we introduce two common features identified as "Structure" and "Process". By the first, we aim at representing all structural features by hierarchically splitting them as the more concrete features. By the second, we aim to do the same to express the process-based vision of this model in more details. The feature "Structure" has two sub-

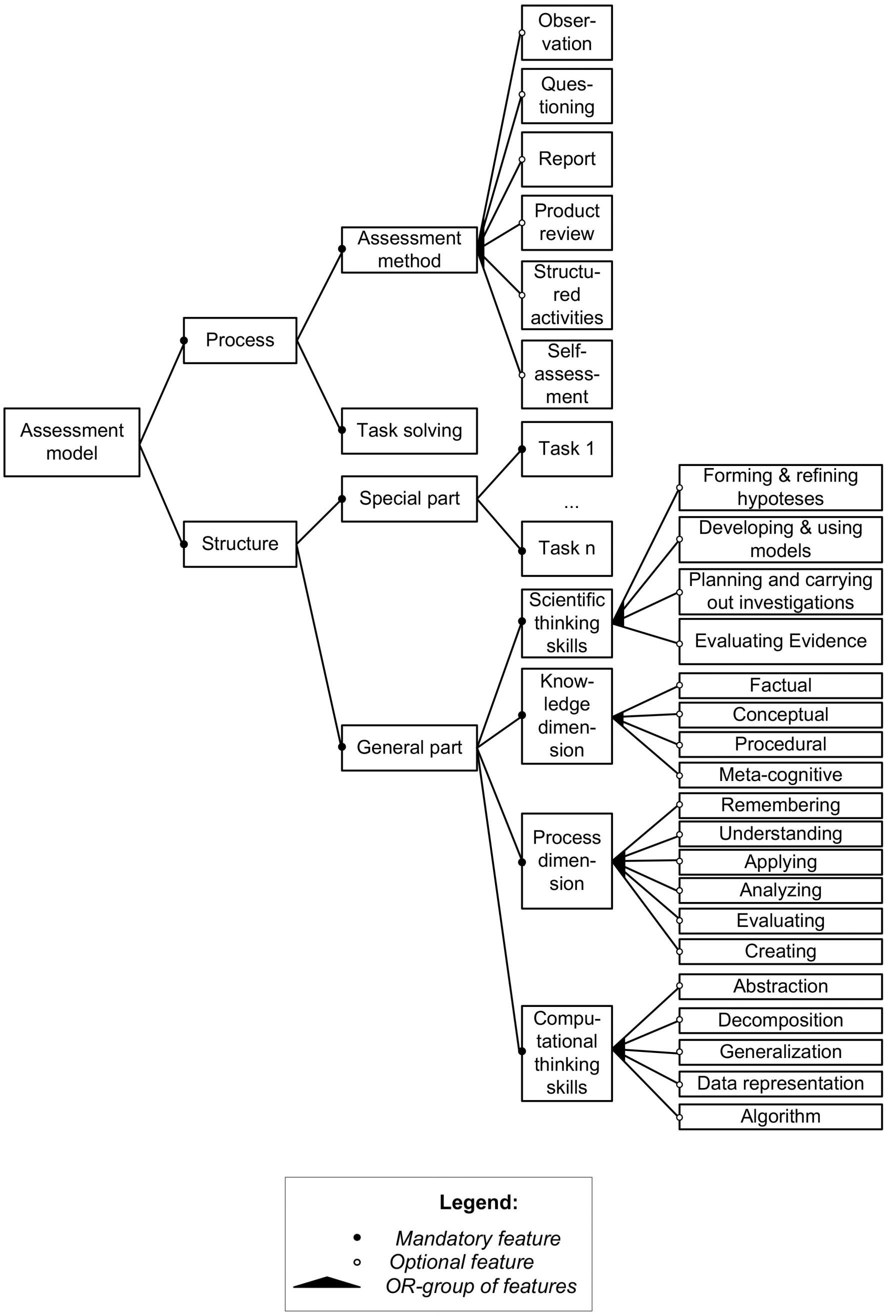

Fig. 6. The abstract feature-based assessment model: revised Bloom taxonomy combined with scientific and computational thinking skills.

features identified as "Generic part" and 'Specific part'. In this branch of the diagram, the role of the remaining sub-features is clear, though we are not able to present a concrete task here (identifying it as Task *n*).

The feature "Process" has two sub-features too. The first sub-feature is "Task solving". The second – "Assessment method" for realizing the knowledge evaluation in practice. The latter includes a set of the grouped features (ordered according to increasing complexity) for possible selection as follows: *observation, questioning, report, product review, structured activities* (projects, scenarios, case study, group discussion), *self-assessment*, etc.

Note that this model enables to express relationships among structural and process-based features through parent-child dependencies (see Fig. 6).

11. Case Study and Experimental Investigation

We have adapted the generative STEM scenario [15, pp. 259–276] for Data Science (DS) aspects and have presented a case study to demonstrate the gaining DS knowledge and skills in the context of integrated STEM. The header part of this scenario includes the following components.

The course applied – *Robotics* (optional module of technology course for 10th grade) with the focus on STEM-driven CS education.

Participants were 48 students of 10th grade (15-16 years) from one gymnasium (in other terms, secondary school). According to the available knowledge and skills, students were *beginners* [15].

Topic as a real-world task follows: *Air temperature & relative humidity measurements in the classroom environment and analysis of the results.*

The main part of the scenario consists of three phases (see also the implementation layer in Fig. 2).

(1) Motivation phase includes a discussion on the real task solution plan and its implementation:

- Selecting devices for task solution and developing structural and physical models of the experimental system. Students have applied the previous knowledge from science (physics) and have gained new knowledge about the temperature-humidity sensor measurement principles and data collecting and storing on the micro-SD card.
- Discussing an algorithm of the experimental system functionality. Students have manipulated the previous knowledge from computer science (programming) and were able to get new knowledge how to use functions of special libraries to retrieve data from the temperature-humidity sensors and write data on the micro-SD card.
- Considering aspects of the experimental measurements and presenting and explaining the results obtained. Students have operated the previous knowledge from science to produce new knowledge about data transforming procedures with explanation of the physical phenomena.

At the motivation phase, collaborative learning was dominating, because the students have developed the task-solving plan through the group discussions.

(2) Problem active investigation through task solving consists of following learning activities:

- Students have constructed an experimental system that consists of the air temperature & humidity sensor and ARDUINO microcontroller. Students first have developed a structural model of the system (see Fig. A-1 in Appendix). Next, having this model, they have had to implement it through physical modelling (see Fig. A-2 in Appendix). In this stage, students were able to improve their knowledge and skills in physics, e.g., electricity, sensor‘s functionality principles, an explanation of the physical phenomena (S-component). In addition, students have analysed the technology-driven environment (T-component), so being introduced to applied engineering (E-component). From the DS viewpoint, students were introduced with data collecting and storing processes; thus they have acquired the knowledge about data collecting and storing modelling for the given task and learning about data storage technologies.
- After implementing the physical model of the system, students had to program the system's functionality in ARDUINO Software (IDE) (see Fig. A-3 in Appendix). Then they have had to verify whether the system is working properly. In this stage, students could be able to improve programming skills (CS-component), to deepen electrical circuit design skills (E-component) and to learn assessing the suitability of the system for the task solving (interdisciplinary STEM knowledge).
- In the third stage of this phase, students have had to model and implement experiments and perform air temperature and humidity measurements over time. In our case, air temperature and humidity were measured in the classroom when the window was opened or closed. The results were stored on micro-SD card as CSV (Comma Separated Values) file. At the end of the experiment, the students processed the results using a spreadsheet and presented them graphically (see Fig. 7). From the DS viewpoint, students were introduced with data transforming procedures. In the fourth stage, students have had to analyse the obtained results and formulate conclusions.

At the problem-solving phase, the personalized learning was dominating, because the students have developed experimental systems and performed measurements individually. All students have been involved in active learning activities covering inquiry-based, problem solving, and design-based learning methods.

(3) Analysis of results, discussion, reflection and evaluation phase consists of two parts.

- Firstly, students have presented and discussed results and formulated conclusions. In Fig. 7, we present time-temperature (T) & time-relative humidity (RH) dependencies obtained by one student because results of others were very similar. At the initial time of measurement, the window in the classroom was open (T = 19°C, RH = 53%). When the window was closed, the temperature in the classroom has raised to 21.7°C and the humidity – to 61% and in the course of

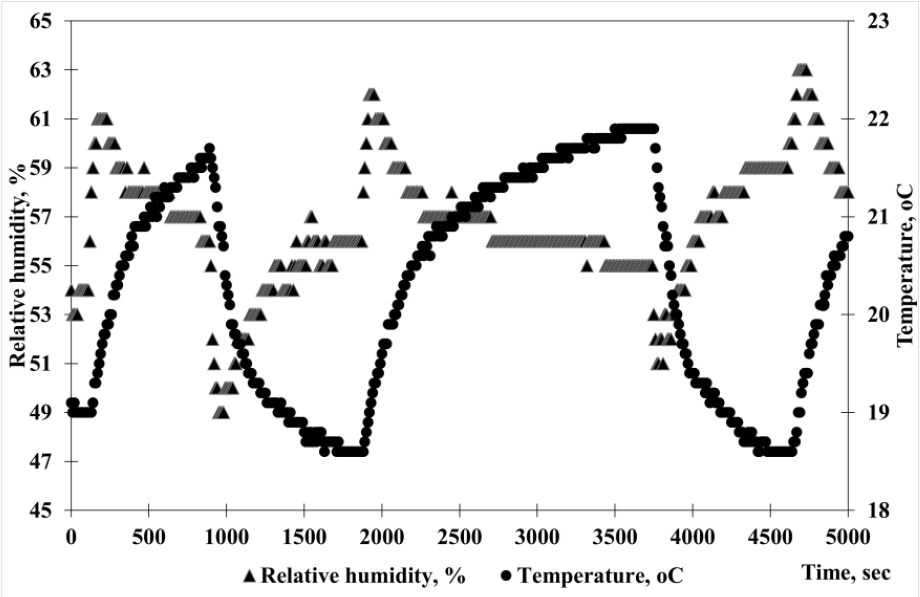

Fig. 7. Time-Relative humidity, Time-Temperature dependencies in the classroom premises.

the time started to decrease achieving 49%. By opening-closing the window, the regularity of temperature and relative humidity changes have been repeating. Students have applied the physics knowledge to formulate conclusions (S-component).

At this phase, the collaborative learning dominated, as students presented, discussed results and defined conclusions working in groups.

- The teacher has evaluated the progress made by students using the developed assessment model (see Fig. 6). Note that this is a subjective assessment applying observation and structured activities assessment methods. It depends on the teacher's experience and the pedagogical specification of the task by fixing the learning progress at different stages of task solving (see Table 2).

The results of the assessment are presented in Table 3. In addition, the teacher has been convinced that learning motivation greatly improved since students also performed similar experiments at home on their own initiative.

12. Discussion and Evaluation

In this paper, we have presented a methodology describing the possibilities for seamless integration of Data Science concepts (data collecting, transferring, transforming, and analysing) into the STEM-driven CS course at the high (secondary) school. The methodology includes two things – the conceptual model of the approach and the three-layered framework to define the whole activities for achieving the research objectives using modelling. Our methodology covers conceptual, feature-based, and physical modelling. By passing through this sequence, we were able to implement our approach and test it as we described in the case study and experimental investigation. Devised models, especially physical modelling and design and testing activities provided by students largely contribute to engineering education too. Based on achieved outcomes in this research, our previous research and works of others, we can confirm the power of modelling in multiple aspects, such as the implementation of educational systems, students' higher motivation, better involvement in learning through learning-by-doing and inquiry-based learning. The important aspect of the discussed approach is its inheritance to real-world task solving in both STEM-driven CS education, and DS. Regarding the latter field, however, there is some specificity due to the need to focus on raw data identified as Big Data [12]. Typically, sensors stand for the facility to collect the raw data. On the other hand, often sensors are nodes of Internet of Things (IoT), the emerging technology in the 21st century [58]. As the IoT concept attracted the attention of educational researchers too, we can consider that sensors used in our approach are nodes of the virtual or real IoT for educational purposes [59]. Therefore, the scenarios of choosing real-world tasks may vary from a simple application (the case in our approach) to the IoT-based applications.

Next, we focus on dealing with advantages of the proposed approach in two aspects: *methodological* and *pedagogical*. Methodological aspects include:

Table 2. Pedagogical specification of the task according to assessment model (see Fig. 6)

Assessment criteria	Explanation of criterion
Revised Bloom's taxonomy (cognitive processes)	
Remembering	Knowledge about temperature-humidity sensor measurement principles and data collecting and storing on micro-SD card, how to use functions of special libraries to retrieve data from temperature-humidity sensor and write them on micro-SD card, knowledge about data transforming procedures and explanation of the physical phenomena.
Understanding	Understanding temperature-humidity sensor's functionality, data collecting and storing on micro-SD card principles, understanding that exist dependencies between temperature and relative humidity.
Applying	Developing an experimental system and retrieving data through automatized measurements of relative humidity and temperature by opening-closing window in the classroom.
Analysing	Transforming results into spreadsheet and presenting them graphically. Explaining time-relative humidity and time-temperature dependencies.
Evaluating	Evaluating the task solution plan, the developed system, and defining advantages and disadvantages of the system.
Creating	Students demonstrated their ability to create own systems using additional sensors (e.g., noise, gas, pulse, etc.).
Computational thinking skills	
Abstraction	Extracting the physical phenomena from the retrieved data, developing similar experimental systems using additional sensors.
Decomposition	Testing of temperature-humidity sensor functionality, testing of data collecting and storing on micro-SD card.
Generalisation/ Pattern recognition	Developing similar experimental systems using additional sensors and providing a similar data collecting and storing.
Data representation	Representing Comma Separated Values (CSV) data graphically.
Algorithm	Developing an algorithm of the experimental system functionality.
Scientific thinking skills	
Using models	Developing structural and physical models.
Planning and carrying out investigations	Developing an experimental system and retrieving data through automatized measurements of relative humidity and temperature by opening-closing window in the classroom.
Analysing and interpreting data/ evidence	Transforming results into spreadsheet and presenting them graphically. Explaining time-relative humidity and time-temperature dependencies.
Constructing explanations	Transforming results into spreadsheet and presenting them graphically. Explaining time-relative humidity and time-temperature dependencies.

Table 3. Assessment of students' progress obtained through teacher's monitoring and fixing the outcomes (see Table 2)

Revised Bloom's taxonomy		**Computational thinking**		**Scientific thinking**	
Process dimension	**% of students**	**Skills**	**% of students**	**Skills**	**% of students**
Remembering	100	Abstraction	70	Using models	100
Understanding	100	Decomposition	70	Planning and carrying out investigations	100
Applying	100	Generalisation/ Pattern recognition	50	Analysing and interpreting data/ evidence	70
Analysing	70	Data representation	50	Constructing explanations	50
Evaluating	30	Algorithm	50		
Creating	10				

(i) Logical connectedness of different steps in terms of produced outcomes. (ii) Strong focus on modelling processes and explicit model creation using different modelling approaches. (iii) The proposed methodology covers the full cycle of data processes (collecting, transferring, transforming, and analysing). (iv) Strong adherence to BD domain and solving real-world problems. (v) Methodology extends STEM-driven CS education capability significantly due to modelling and contributes to engineering education. (vi) The proposed methodology ensures a flexible configuration by adding new components into previously developed SLE.

From pedagogical perspective, this approach (i) enables to enact collaborative learning. (ii) It enhances computational thinking and scientific thinking skills. (iii) Practically, the approach is independent upon the previous knowledge (in our experiment all students were beginners). (iv) Approach contributes to the increased motivation (all students have passed the full cycle of DS processes). (v) Approach also supports personalized learning since students can work on own pace and relevant intensity and at home. (vi) Approach covers a set of pedagogical approaches such as learning-by-doing, problem-based and inquiry-based adding the pedagogical value. (vii) In terms of modelling, this approach enables student to fasten the move from the beginner's state to the advanced state after completing the introduced learning scenario. In the latter state, active students can collaborate by extending the functionality of the system through modelling and thus to gain engineering skills. (viii) Approach implements the extended assessment model that also includes assessment of computational and scientific thinking skills.

The approach has also some drawbacks. (1) Looking at the multiple steps or stages within this methodology, one may treat it as being complex enough. Its complexity, however, not so much relates with students (because of this methodology is well structured) but, largely, complexity issues may occur for teachers (due to their insufficient professional preparedness in interdisciplinary knowledge to manage the multiple processes properly before delivering this knowledge to students). (2) So far, we have still a restricted experience of using this approach in terms of durability, scope of experiments, the number of students involved, due to the novelty of the approach. (3) To support the approach, we need to apply the modern facilities and technological resources, which typically cannot be found in already existing educational systems.

13. Conclusions

Three fields – STEM, Computer Science (CS) and Data Science (DS) – have inherent relationships, making the seamless integration of all into one course possible. We have proposed a novel three-layered framework to do that systematically. The basic idea relies on using models and modelling. We have applied the three types of modelling in our framework – conceptual at the top layer, feature-based at the intermediate layer and physical at the implementation layer. All these have enabled to achieve the systemisation and coherence of the proposed methodology. The provided research has confirmed once again the role, the power, and the relevance of modelling-oriented methodologies for education in general and for STEM-driven CS education with multiple engineering approaches. The multi-stage modelling and the focus on data-centric real-world tasks enable to achieve a higher degree of integration what Hallström & Schönborn call authentic STEM education. Our case study shows that solving authentic tasks through modelling is highly influential on students' interdisciplinary knowledge gaining due to the enforced interest and motivation to learn. The DS component, with its processes and supporting facilities is, in fact, an enabler and driver in gaining this knowledge. The separate DS processes, however, affect the STEM components and therefore interdisciplinary knowledge gaining, differently. For example, data collecting, transferring, and storing are largely influencing the S-, T- and E-components and adequate knowledge, while data transforming, analysing (representing) are more influential on the M-component. What is common for all components (S, T, E, M and DS) is the possibility to get and to enrich gradually the scientific and computational thinking skills of learners. In addition, our experiment through the case study shows that the extent of the interdisciplinary knowledge is extended due to the related DS processes and new facilities used.

From the pure technological and engineering viewpoint, when students hold in their hands sensors and provide collecting data automatically from classroom or other environments outside, one can treat that as simplified nodes of the virtual Internet of Things (IoT), the most striking technology of the 21st century. Of course, there is a long way to apply this technology in the educational sector effectively, though one can find successful attempts in the literature already now.

At the end, we need to emphasise yet another important finding – the successful implementation of this methodology in reality can take place only if the teacher has the adequate competence in the related fields, or at least can be ready to accept and overcome the challenges of interdisciplinary teaching. Therefore, the teacher's professional preparation could be a serious obstacle to accept and use this methodology in a wider context.

The future work will be directed on a wider exploration of big data capabilities in our vision on STEM-based research and practice.

References

1. National Research Council, *Next Generation Science Standards: For States, By States*, Washington, DC: The National Academies Press, https://doi.org/10.17226/ 18290, 2013.
2. B. Freeman, S. Marginson and R. Tytler, An international view of STEM education, in *STEM Education 2.0*, Brill Sense, pp. 350–363, 2019.

3. S. C. Fan, K. C. Yu and K. Y. Lin, A Framework for Implementing an Engineering-Focused STEM Curriculum, *International Journal of Science and Mathematics Education*, pp. 1–19, 2020.
4. O. Fragou, C. Goumopoulos and C. Tsompanos, STEM Oriented Online Platforms Embracing the Community of Practice Model: A Comparative Study and Design Guidelines, *J. UCS*, **25**(12), pp. 1554–1588, 2019.
5. S. Sutaphan and C. Yuenyong, STEM Education Teaching Approach: Inquiry from the Context Based. In *Journal of Physics: Conference Series*, **1340**(1), IOP Publishing, p. 012003, 2019.
6. T. Verily, N. Celeste, C. Scribner, J. A. Francis and D. Cross, Enhancing STEM Learning through an Interdisciplinary, Industry-Generated Project: The Project Required Students to Solve a Complex Problem by Integrating and Applying a Range of Knowledge and Skills across Different Disciplines, *Technology and Engineering Teacher*, **79**(1), pp. 26–31, 2019.
7. H. Kimmel, L. Hirsch, L. Burr-Alexander and R. Rockland, Engineering & STEM: Complementary areas of study, *International Journal of Engineering Education*, **33**(1), pp. 287–294, 2017.
8. M. Ryan, J. Gale and M. Usselman, Integrating engineering into core science instruction: Translating NGSS principles into practice through iterative curriculum design, *International Journal of Engineering Education*, **33**(1), pp. 321–331, 2017.
9. J. M. Wing, Computational thinking, *Communications of the ACM*, **49**(3), pp. 33–35, 2006.
10. M. Guzdial and B. Morrison, Growing computer science education into a STEM education discipline, *Communications of the ACM*, **59**(11), pp. 31–33, 2016.
11. P. Zikopoulos and C. Eaton*, Understanding big data: Analytics for enterprise class hadoop and streaming data*, McGraw-Hill Osborne Media, 2011.
12. V. Dhar, Data science and prediction, *Communications of the ACM*, **56**(12), pp. 64–73, 2013.
13. G. Bell, T. Hey and A. Szalay, Beyond the data deluge, *Science*, **323**(5919), pp. 1297–1298, 2009.
14. K. M. Tolle, D. S. W. Tansley and A. J. Hey, The fourth paradigm: Data-intensive scientific discovery [point of view], *Proceedings of the IEEE*, **99**(8), pp. 1334–1337, 2011.
15. V. Štuikys and R. Burbaitė, *Smart STEM-Driven Computer Science Education: Theory, Methodology and Robot-based Practices*, Springer, 2018.
16. J. Hallström and K. J. Schönborn, Models and modelling for authentic STEM education: reinforcing the argument, *International Journal of STEM Education*, **6**(1), pp. 1–10, 2019.
17. J. K. Gilbert and R. Justi, *Modelling-based teaching in science education*, **9**, Basel, Switzerland: Springer international publishing, 2016.
18. L. S. Nadelson and A. L. Seifert, Integrated STEM defined: Contexts, challenges, and the future, *The Journal of Educational Research*, **110**(3), pp. 221–223, 2017.
19. M. Honey, G. Pearson and H. Schweingruber (eds.), *STEM Integration in K-12 Education: Status, Prospects, and an Agenda for Research. National Academy of Sciences*, The National Academies Press, Washington, D.C., 2014.
20. U. Hasanah, Key Definitions of STEM Education: Literature Review, *Interdisciplinary Journal of Environmental and Science Education*, **16**(3), e2217, 2020.
21. U. Kubat and E. Guray, To STEM, or not to STEM? That is not the question, *Cypriot Journal of Educational Sciences*, **13**(3), 388-399, 2018.
22. L. Thibaut, L., S. Ceuppens, H. De Loof, J. De Meester, L. Goovaerts, A. Struyf, J. Boeve-de Pauw, W. Dehaene, J. Deprez, M. De Cock, L. Hellinckx, H. Knipprath, G. Langie, K. Struyven, D. Van de Velde, P. Van Petegem and F. Depaepe, Integrated STEM education: A systematic review of instructional practices in secondary education, *European Journal of STEM Education*, **3**(1), p. 2, 2018.
23. L. Hobbs, J. C. Clark and B. Plant, Successful students–STEM program: Teacher learning through a multifaceted vision for STEM education, in *STEM education in the junior secondary*, Springer, Singapore, pp. 133-168, 2018.
24. J. Krajcik, Three-Dimensional Instruction: Using a New Type of Teaching in the Science Classroom, *Science Scope*, **39**(3), pp. 16–18, 2015.
25. B. Priemer, K. Eilerts, A. Filler, N. Pinkwart, B. Rösken-Winter, R. Tiemann and A. U. Zu Belzen, A framework to foster problem-solving in STEM and computing education, *Research in Science & Technological Education*, **38**(1), pp. 105–130, 2020.
26. H. Jang, Identifying 21st Century STEM Competencies Using Workplace Data, *Journal of Science Education and Technology*, **25**, pp. 284–301, 2016.
27. A. Yadav, H. Hong and C. Stephenson, Computational Thinking for All: Pedagogical Approaches to Embedding 21st Century Problem Solving in K-12 Classrooms, *TechTrends*, **60**, pp. 565–568, 2016.
28. Code.org Advocacy Coalition, Computer Science Teachers Association, and Expanding Computing Education Pathways Alliance, "2019 state of computer science education", https://advocacy.code.org/2019_state_of_cs.pdf, Accessed 1 September, 2021.
29. S. Azouzi, S. A. Ghannouchi and Z. Brahmi, Towards supporting modeling variability in e-learning application: a case study, *18th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT)*, pp. 488–494, 2017.
30. H. Sebbaq, A. Retbi, M. K. Idrissi and S. Bennani, Software Product Line to overcome the variability issue in E-Learning: Systematic literature review, *Proceedings of the 12th International Conference on Intelligent Systems: Theories and Applications*, pp. 1–8, 2018.
31. E. F. Coutinho and C. I. Bezerra, A study on dynamic aspects variability in the SOLAR educational software ecosystem, *Journal of the Brazilian Computer Society*, **26**(1), pp. 1–19, 2020.
32. N. M. Seel, Model-based learning: A synthesis of theory and research, *Educational Technology Research and Development*, **65**(4), pp. 931–966, 2017.
33. B. Thalheim, Towards a theory of conceptual modelling, *J. UCS*, **16**(20), pp. 3102–3137, 2010.
34. F. Provost and T. Fawcett, Data science and its relationship to big data and data-driven decision making, *Big Data*, **1**(1), pp. 51–59, 2013.
35. A. Deen, Why Data Science Will be Among the Most Promising Careers in 2020, https://www.equities.com/why-data-science-will-be-the-most-promising-career-in-2020, Accessed 1 September, 2021.
36. I. Song and Y. Zhu, Big data and data science: what should we teach?, *Expert Systems*, **33**(4), pp. 364–373, 2016.
37. R. J. Brunner and E. J. Kim, Teaching Data Science, *The International Conference on Computational Science*, **80**, pp. 1947–1956, 2016.

38. S. C. Hicks and R. A. Irizarry, A Guide to Teaching Data Science, The American Statistician, **72**(4), pp. 382–391, in Rolf, B. et al, Paderborn Symposium on Data Science Education at School Level 2017: The Collected Extended Abstracts. Paderbon: Universitätsbibliothek Paderborn, pp. 1–14, 2018.
39. R. Gould, S. Machado, C. Ong, T. Johnson, J. Molyneux, S. Nolen, H. Tangmunarunkit, L. Trusela and L. Zanontian, Teaching Data Science to Secondary Students: The mobilize introduction to data science curriculum, *Iase-Web. Org*, pp. 1–11, 2016.
40. C. Ridslade, J. Rothwell, M. Smit et al., *Strategies and best practices for data literacy education: knowledge synthesis report*, Dalhousie University, 2015.
41. R. Biehler and C. Schulte, Perspectives for an Interdisciplinary Data Science Curriculum in German Secondary Schools, *Paderborn Symposium on Data Science Education at School Level*, pp. 2–14, 2017.
42. V. Pittard, The Integration of Data Science in the Primary and Secondary Curriculum, *Final Report to the Royal Society Advisory Committee on Mathematics Education*, 2018.
43. T. Irish, A. Berkowitz and C. Harris, Data Explorations: Secondary Students' Knowledge, Skills and Attitudes Toward Working with Data, *EURASIA Journal of Mathematics, Science and Technology Education*, **15**(6), pp. 1–15, 2019.
44. F. F. Cruz and M. F. Diaz, Generations Z's Teachers and their Digital Skills, *Comunicar, Media Education Research Journal*, **24**(1), pp. 97–105, 2016.
45. P. Tong and F. Yong, Implementing and Developing Big Data Analytics in the K-12 Curriculum: A Preliminary Stage, *Conference Paper EdCon*, 2015.
46. S. Sentance, Data Science and Data Literacy in School: Opportunities and Challenges, *Proceedings of the 18th Koli Calling International Conference on Computing education Research*, pp. 84–89, 2017.
47. C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown Books, Washington D.C., USA, 2016.
48. How should high school students learn data science?, https://www.quora.com/How-should-high-school-students-learn-data-science, Accessed 1 September 2021.
49. I. M. Souza, W. L. Andrade, L. M. Sampaio and A. L. S. O. Araujo, A Systematic Review on the use of LEGO® Robotics in Education, in 2018 IEEE Frontiers in Education Conference (FIE), pp. 1–9, 2018.
50. M. F. Iskander, J. Baker, J. K. Nakatsu, S. Y. Lim and N. Celik, Multimedia Modules and Virtual Organization Website for Collaborative Research Experience for Teachers in STEM, *J. UCS*, **17**(9), pp. 1347–1364, 2011.
51. D. Tsiastoudis and H. Polatoglou, Inclusive education on stem subjects with the arduino platform, *Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion*, pp. 234–239, 2018.
52. P. Y. Schobbens, P. Heymans and J. C. Trigaux, Feature diagrams: A survey and a formal semantics, *14th IEEE International Requirements Engineering Conference (RE'06)*, pp. 139–148, 2006.
53. K. Kang, S. Cohen, J. Hess, W. Novak and S. Peterson, *Feature-oriented domain analysis (FODA) feasibility study*, Software Engineering Institute, Carnegie Mellon University, 1990.
54. C. Zimmerman and D. Klahr, Development of scientific thinking, *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, **4**, pp. 1–25, 2018.
55. K. Cummins, *Teaching Digital Technologies & STEM: Computational Thinking, coding and robotics in the classroom*, Retrieved from Amazon.com, 2016.
56. L. W. Anderson and B. S. Bloom, *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*, Longman, 2001.
57. S. Atmatzidou and S. Demetriadis, Advancing students' computational thinking skills through educational robotics: A study on age and gender relevant differences, *Robotics and Autonomous Systems*, **75**, pp. 661–670, 2016.
58. P. P. Ray, A survey on Internet of Things architectures, *Journal of King Saud University-Computer and Information Sciences*, **30**(3), pp. 291–319, 2018.
59. M. Kassab, J. DeFranco and A. Laplante, A systematic literature review on Internet of things in education: Benefits and challenges, *Journal of Computer Assisted Learning*, **36**(2), pp. 115–127, 2020.

Appendix

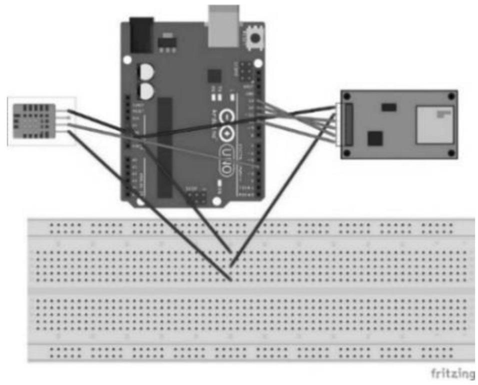

Fig. A-1. A structural model of system developed using fritzing software *(DHT11 Temperature & Humidity sensor on left, SD Card Module on right and ARDUINO in the centre)*.

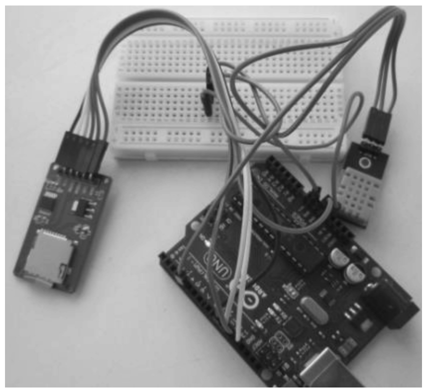

Fig. A-2. An implementation of the structural model *(SD Card module on left, ARDUINO in the centre and DHT11 Temperature & Humidity sensor on right)*.

```
#include "DHT.h"
#include <SPI.h>
#include <SD.h>

#define DHTPIN 4
#define DHTTYPE DHT11
DHT dht(DHTPIN, DHTTYPE);

const int chipSelect = 10;
File myFile;

void setup()
{
  Serial.begin(9600);
  while (!Serial) {
    ;
  }

  Serial.print("Initializing SD card...");

  if (!SD.begin(chipSelect))
  {
    Serial.println("Initialization failed!");
    return;
  }
  Serial.println("Initialization done.");
  dht.begin();
}
```

```
void loop()
{
  delay(2000);
  myFile = SD.open("results.csv", FILE_WRITE);
  float h = dht.readHumidity();
  float t = dht.readTemperature();
  if (isnan(h) || isnan(t))
  {
    Serial.println("Error!");
    return;
  }

  if (myFile)
  {
    Serial.print("Temperature ");
    Serial.println(t);
    Serial.print("Humidity ");
    Serial.println(h);

    myFile.print(t);
    myFile.print(";");
    myFile.println(h);
    myFile.close();
  }
  else
  {
    Serial.println("Error!");
  }
}
```

Fig. A-3. Program of the system functionality.

Vytautas Štuikys, Professor, holds PhD and Doctor Habilitatus degrees in CS and Software Engineering since 1970 and 2003 adequately. His research interests include design of educational software systems (eco-systems) based on using high-level modelling, model transformations (including meta-programming), knowledge-based and Learning Analytics approaches. He is author/co-author of three monographs published by Springer in 2013, 2015, 2018.

Renata Burbaitė, PhD is currently an Associated Professor at the Software Engineering Department of Kaunas University of Technology and Informatics teacher at the High School. Her research interests include CS educational domain modelling and STEM-oriented robot-based CS education methodologies. She is a co-author of the monograph "Smart STEM-Driven Computer Science Education: Theory, Methodology, and Robot-Based Practices", published by Springer in 2018.

Giedrius Ziberkas received PhD degree in information technology in 2001 from Kaunas University of Technology. Currently he is an Associated Professor at Software Engineering Department of Kaunas University of Technology. His research interests include automated program generation, metaprogramming, and development of learning systems. In addition, he has published a few works in the energy efficiency area. Since 2011, he is accumulating experience in the field of personal data security.

Ramūnas Kubiliūnas received PhD degree in information technology in 2009 from Kaunas University of Technology. Currently he is a Lecturer at Software Engineering Department and a Director of Study Trend "Software Systems" at Faculty of Informatics at Kaunas University of Technology. His research interests include distance learning, e-learning, e-learning content preparation in various forms including learning objects since 2001. He is an author and co-author of more than 30 various related publications and is responsible of the master's degree study program "Information Technologies of Distance Education".