

FOSTERING DATA LITERACY COMPETENCIES IN SCHOOL

How can larger amounts of data be analysed in schools? Why is this topic important for students? Current research on data literacy gives an answer to these questions

Today, data analyses are everywhere. For example, large companies collect masses of data and analyse it with the goal to systematically promote their products, whilst social networks use data analyses to suggest ‘friends’.

Nowadays, everyone is confronted with new challenges due to the increasing and widespread relevance of data, for example the necessity to decide which personal and foreign data are shared with others (including services on the internet), under which conditions, and for which purpose. At the same time, the question arises what others can do with, and read from, this data. Accordingly, not only do students need to acquire skills in this area, we all need to become data literate.

Data literacy competencies

But what does data literacy mean?

According to widely accepted definitions, data-literate people are able to work with and handle data in a meaningful way. For example by acquiring, structuring, or analysing it.

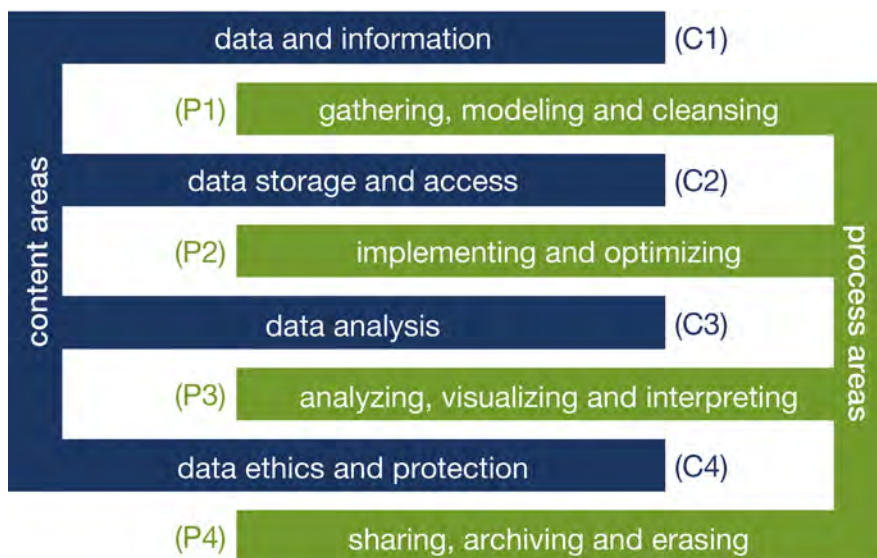
In recent years, we further investigated this topic from a computing education point of view, taking into account various perspectives. On the one hand, we regarded the technical perspective on the large topic data, and on the other hand we also considered the students’ and teachers’ perspectives, as well as requirements coming from society. On this basis, we developed a data literacy competency model.

This model characterises competencies related to data from two different perspectives resulting in two areas: the content areas clearly emphasise technical aspects, and hence are focused on the computer science content, while the process areas take a rather practically oriented perspective, and illustrate what can be done with data.

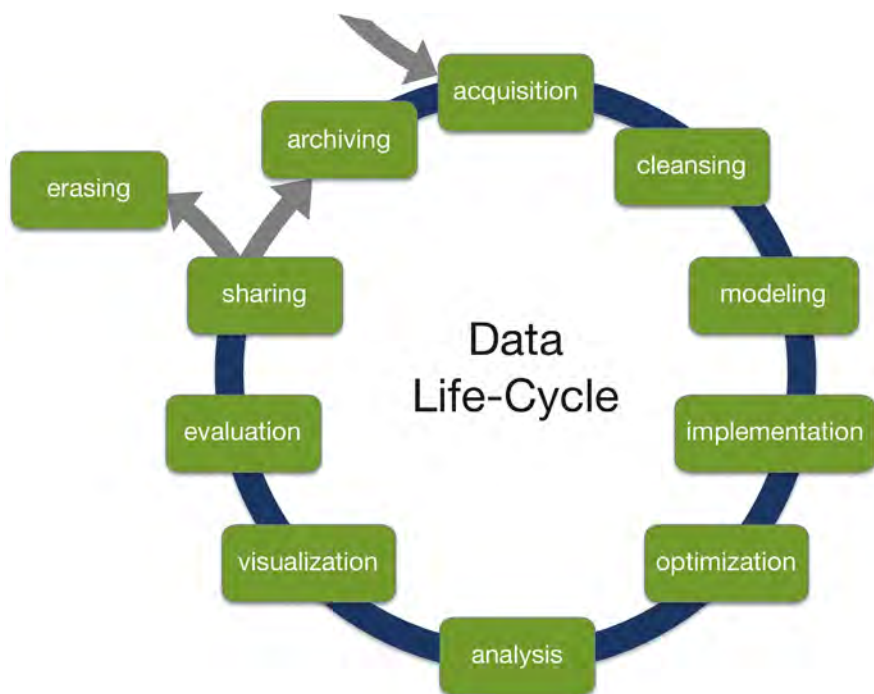
The two types of competency areas are closely intertwined, thus each data literacy competency has to connect to at least one content and one process area. For example, the competency to visualise data and analysis results incorporates both a content aspect (such as knowing different visualisation methods and their purpose) and a process aspect (covering being able to prepare data in a way suitable for visualising them and creating the aspired visualisation). Although the competency model was developed with a focus on computing education, due to its structure it can also be adapted to incorporate aspects from other subjects. After all, computer science is not the only subject that has to deal with data today. Other subjects can contribute important aspects too, particularly to the content areas, hence enriching the model and extending its usability beyond computing education.

The life cycle of data

When trying to include data literacy competencies in school teaching, often the question arises as to where to start. Most computing lesson plans likely have various connection points to data literacy,



■ According to the research-based data literacy competency model, each data literacy competency must have both content-related and process-related aspects



■ The data life cycle model gives an orientation for data-literacy-oriented teaching, for both teachers and students

so there are various possibilities for data literacy teaching. Yet, it is important to keep in mind the whole process of working with data. When only discussing distinct parts of this topic, for example the analysis, other important aspects are missing (such as gathering data or justifying the analysis from an ethical perspective).

Hence, as a guideline for data literacy teaching, we developed the data life cycle model. This model gives teachers and students an orientation when working with data and sets its emphasis on the whole process of working with data, not just a small excerpt of it. Of course, not all aspects can be considered in the same depth in school, but using the data life cycle as an orientation helps to bring together all the knowledge and skills students acquire throughout computing education.

For example, aspects related to data modelling, implementation, and optimisation are typically already there in most computing curricula, and hence only need to be brought into connection with other aspects of the data life cycle. Also, when working with databases, real data could be acquired in class and structured for efficiency, storing them in the database instead of discussing rather fictitious examples. The important question,

therefore, is not where to start with teaching data literacy, but where to connect it to what we teach already. The data life cycle helps to identify such connection points.

Fostering data literacy in school

When data literacy competencies are to be fostered in school, several challenges have to be overcome. Suitable tools have to be identified, appropriate examples that can motivate students and that concern them have to be selected, and concepts need to be worked out on how to foster these skills.

Particularly, as most data literacy competencies cannot be gained by theoretical considerations only, suitable examples and appropriate data play an important role for teaching data literacy. Such data may be acquired from various sources today. For example, programming interfaces (APIs) of widely known services on the internet (such as Twitter) aren't the only data sources that can be used. There are also rich and easy-to-use data sets that are released, such as by public administrations as part of open data projects (for example, open data can be found at data.gov.uk).

An exemplary project, particularly considering data analysis, is based on



DR. ANDREAS GRILLENBERGER

Andreas Grillenberger is a research fellow at Freie Universität Berlin in Germany. Since 2013, he's worked on the topics of data, data management, and data literacy. He investigates these topics from a computing education perspective, particularly with focus on secondary education.

using real data about school students (e.g. a dataset about Portuguese students that was released on the UCI Machine Learning Repository, that can be found at archive.ics.uci.edu/ml/datasets/Adult), which can be analysed by students using simple tools (such as Orange, at orange.biolab.si), with the purpose to predict students' grades based on the information contained in the dataset.

As this setting directly concerns students, particularly if it is presented by the teacher as a possible new way to grade them, several ethical problems are raised that directly affect the students. The resulting discussions on challenges, risks, and opportunities that arise, along with the possibility to work with and analyse large amounts of data directly, lead to another challenge. Although this article has taken a computing education perspective on data literacy, this topic also affects and is relevant for many other subjects. Data literacy should therefore be considered an interdisciplinary topic and taught accordingly in school. Fostering data literacy in school is an open challenge to which we all can, and must, contribute in order to prepare our students for a life in a world where data is used continuously and everywhere. (HW)