Big Data and Data Management:
A Topic for Secondary Computing Education

Andreas Grillenberger
Computing Education Research Group
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
Martensstraße 3, 91058 Erlangen, Germany
andreas.grillenberger@fau.de

ABSTRACT

The topics data management and data analysis are currently discussed in various contexts, e.g. in Computer Science but also in daily life and society. The recent developments in this field, which are often summarized under the term *Big Data*, did not only lead to the emergence of new database models, but also comprise new threats, e.g. for data privacy. These topics include many aspects that are important for everyone, but they only gain in relevance slowly in higher education and hardly in secondary education. Hence, I will evaluate data management as a topic for secondary education, with a view on the long-lasting concepts and aspects in this field.

Categories and Subject Descriptors

K.3.2 [**Computers and Education**]: Computers and Information Science Education—*Computer Science Education*

Keywords

Data Management; Big Data; Secondary Schools; Daily Life; Databases; Data Privacy; Data Analysis; NoSQL

1. RESEARCH SITUATION

I am a first-year doctoral candidate in computing education research at the University of Erlangen-Nürnberg, Germany, under supervision of Prof. Dr. Ralf Romeike. As usual in Germany, my doctoral study is not part of a PhD program, instead it is an individual part-time study alongside a (full-time) employment as research assistant.

Since my work is right at its beginning, I expect to defend my thesis in about 4.5 years. Until now, I focused especially on narrowing the field, as data management covers various different aspects. Yet, some first ideas on this topic have been developed and published as conference papers [5, 6].

2. CONTEXT AND MOTIVATION

Despite their influences on CS as well as on daily life and society, the topics data management and Big Data only play a minor role in current CS education. Additionally, the relevance of various concepts and topics currently taught in computing education is challenged by the recent developments. Today, education in this field focuses on "databases", while in the future the impact of other aspects will increase: data privacy, data safety/security, data analysis, Open Data, distributed data, cloud storage, data backup, and so on. Therefore, CS education research needs to analyze the influences of these developments on current teaching. Overall, the current emphasis on relational database management systems (RDBMS) results in the main topics data modeling, normalization, redundancy-free and consistent data storage, SQL, and so on. While these aspects are clearly fundamental to RDBMS, their relevance is changing when considering other modern databases, like the NoSQL[1] ones. For example, such newly emerging databases often use redundancy for speeding up the access to distributed stored data. As topics considered as fundamental so far may lose importance in the future, the current focus of database education is being questioned.

Therefore, current education on this topic is not able to fulfill the task of preparing students for handling the large amounts of data in their daily and probably in their later vocational life. Nor can it sufficiently support understanding of current issues like the threats of large data collections.

3. BACKGROUND & RELATED WORK

This focus of teaching was last discussed in the 1990s: e.g. Witten [8] emphasizes the relevance of (R)DBMS for CS education. Later, mainly learning environments (especially for teaching SQL, e.g. [4]) were presented, while only few work on approaches for teaching databases and SQL was published: e.g. Antonitsch [1] questions current approaches for database education by discussing databases in context of information retrieval. At the moment, a discussion on these topics seems to emerge: at the SIGCSE'14 multiple papers in this field were presented, mainly suggesting data management as a topic for higher education. However, Buffum et. al. [2] set their focus on middle schools, especially on integrating Big Data aspects into other subjects in order to teach CS principles by using these as a tool.

Another field that is also strongly related to data management and (CS) education is learning analytics / educational data mining. This topic is already being discussed in detail in various other publications and research projects. Hence, I will not cover such aspects in my research.

---

[1]Nowadays, NoSQL is commonly interpreted as "not only SQL" [3] in contrast to its original meaning "no SQL".

4. STATEMENT OF THESIS/PROBLEM

The shortage of publications shows that data management is only considered as a marginal topic so far. Today, data become visible for everyone, as their value is clearly increasing and even popular media are talking about data analysis. In this context, everyone needs knowledge, skills and competencies for handling data. Therefore, the key question guiding my research is: "*Which are the influences of current developments in data management on secondary computing education and how can CS education handle the newly arising requirements in this field?*". This especially includes a broader view on the topic "data" in future CS education, instead of focusing mainly on databases.

5. RESEARCH GOALS & METHODS

In order to answer the key question stated before, I will focus on the following research questions:

What knowledge and skills does everyone need to have concerning data management? A first step to answering this question was having a view on current topics in society and daily life that are often strongly related to data and data management. On this basis, some ideas on key competencies for handling data in a proper way were derived inductively [6]. In the future, it is planned to expand and systematize this catalog, e.g. by applying a competence model. Additionally, I plan to design some checklists that can be used as learning material (for example on topics like "*What can I do to make my data secure?*") in order to raise learners' and teachers' awareness on this topic.

Which are the fundamental and long-lasting aspects of data management that are important for computing education? Data management is a topic that is continuously changing and developing. In contrast, the aspects brought to school must be stable during these changes in order to prevent teaching of outdated knowledge and skills. A starting point for analyzing such aspects is focusing on databases: today, various concepts supposed to be fundamental so far are changing; in this context, stable concepts can be found by comparing the concepts of newly arising database models with the proven ones.

Which attitudes do learners currently have on the chances and threats of data management? Today, learners are mostly confronted with data management long before secondary and even before primary education. Hence, an important basis for teaching are their prejudices and preknowledge on this topic, which may range from "excitement" to "helplessness". Dealing with these attitudes is fundamental for finding the main aspects of data management for secondary education. Therefore, analyzing these attitudes and knowledge by a questionnaire study is an important task.

Which are the new requirements computing education has to deal with in the future? For finding these requirements, a first step was taken: first ideas on challenges for CS education were collected based on the literature review and will be presented at ISSEP'14 [5]. In the future, these challenges need to be detailed and more well-founded, e.g. by comparing the current CS education with the findings in the other questions described before.

6. DISSERTATION STATUS

As my research is at a preliminary stage, writing the thesis was not started yet. However, some first ideas were gained while narrowing the field of research and were / will be presented in form of conference papers. These include a catalog of key competencies everyone needs for handling data that was presented at KEYCIT'14 [6], as well as first ideas for the challenges that lie ahead for CS education [5]. These will be published at ISSEP'14 together with a poster that discusses the influences of data management on secondary education in general and outlines the gaps of current education on these topics [7].

At the doctoral consortium I am looking forward to others' estimations on my research topic and would be interested in additional ideas or aspects I have not considered so far. Especially, as no empirical research has been done yet, the doctoral consortium may help refining the research questions stated above. By having a look on others' ongoing work, I also expect having new ideas for how to deal with my own research. At last, meeting people who possibly work at related topics and establishing contacts is an important reason for attending the conference for me.

7. EXPECTED CONTRIBUTIONS

The ongoing work provides a clear contribution to computing education: a main topic of CS, "data", will be analyzed for teaching purposes in the context of current developments but with a strong emphasis on the long-lasting ideas and concepts of data management. This will not only prepare including these ideas into curricula, but can also enable students to understand current developments and handle data in a more responsible way in their daily life.

8. REFERENCES

[1] P. Antonitsch. Databases as a tool of general education. In R. Mittermeir, editor, *Informatics Education - The Bridge between Using and Understanding Computers*, volume 4226 of *Lecture Notes in Computer Science*, pages 59–70. Springer Berlin Heidelberg, 2006.

[2] P. Buffum et al. CS Principles Goes to Middle School: Learning How to Teach "Big Data". In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education*, SIGCSE '14, pages 151–156, New York, NY, USA, 2014. ACM.

[3] S. Edlich et al. *NoSQL [in German]*. Hanser, Carl Gmbh + Co., 2011.

[4] A. Grillenberger and T. Brinda. eledSQL: A New Web-based Learning Environment for Teaching Databases and SQL at Secondary School Level. In *Proceedings of the 7th Workshop in Primary and Secondary Computing Education*, WiPSCE '12, pages 101–104, New York, NY, USA, 2012. ACM.

[5] A. Grillenberger and R. Romeike. Big Data - Challenges for Computer Science Education [in print]. In *Proceedings of ISSEP 2014*.

[6] A. Grillenberger and R. Romeike. Teaching Data Management: Key Competencies and Opportunities [in print]. In *Proceedings of KEYCIT 2014*.

[7] A. Grillenberger and R. Romeike. Data Management: More Than a Matter of CS. In *Local Proceedings of ISSEP 2014*, 2014.

[8] H. Witten. Datenbanken - (k)ein Thema im Informatikunterricht? [Databases - (No) Topic in Computer Science Education?]. *LOG IN*, 2, 1994.