

Bringing the Innovations in Data Management to Secondary CS Education

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

ABSTRACT

Data Management is a field of CS that has changed significantly with the innovations of recent years. With increasing amounts of data that is being collected, stored and processed today, this field evolved from the traditional field databases. While storing data in databases is a typical topic in (secondary) CS education, more modern aspects are hardly covered at all. However, as these innovations have a high impact on our daily lives and in society, not only can they help convey a realistic impression of CS, but they have also the potential to support students when handling (their own and others') data. In my research project, I investigate the field data management from an educational point-of-view in order to identify central aspects relevant and practical for CS classes in secondary schools. To ensure that the identified concepts and ideas are central to CS education, the field will be analyzed not only from a CS point-of-view, but also from the students', teachers' and the societal perspective.

Categories and Subject Descriptors

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

Keywords

data management; databases; big data; secondary CS education; educational reconstruction

1. PROGRAM CONTEXT

I am a doctoral student at the computing education research group at the University of Erlangen-Nürnberg, Germany, since 10/2013. My advisor is Prof. Dr. Ralf Romeike. As it is common in Germany, I am not enrolled in a PhD program, instead I work as research assistant at the university. I intend to finish my work by late 2018. So far, I have set my focus on finding an appropriate research framework [7], analyzing the science content structure of data management as well as on characterizing the gap between data manage-

ment in CS and in CS education [5]. The next step is to go through a first iteration of the research framework.

2. CONTEXT AND MOTIVATION

Besides algorithms and data structures, databases have been a central topic of CS education (CSE) for years. In CS, various innovations in this field led to a shift of focus from databases to data management in general. Not only have new topics evolved, but various of the topics in this field are also considered from a wider perspective today. These innovations strongly affect society and our daily lives—e. g. when talking about systems like online shops or social media. Yet, they are hardly covered in current CSE: aspects like handling meta-data are only considered marginally, new methods of data analysis are missing, and the increasing parallelization of data processing is not part of current teaching, even though these aspects are generally accepted as being central to CS. Additionally, next to providing a realistic perspective on CS, an important task of CS teaching is fostering competencies that are needed for daily life, like using and handling data in a self-determined and responsible way.

3. BACKGROUND & RELATED WORK

Data Management is frequently discussed from various perspectives. This field was characterized in detail by the Data Management Association in their Data Management Body of Knowledge [2]. Also, established database textbooks are central to data management. In CSE, many publications on big data / data management can be found, too. However, most of them are related to learning analytics, which is not covered by my research. Only few articles can be found on data management as topic in class: while databases were discussed intensively in the early 1990s (e. g. [8]), there have hardly been any publications for years. But in recent years, the attention on this topic has increased: e. g. Buffum et al. describe an approach for integrating big data principles into other subjects [1], and the topic is also mentioned in the draft for the revised K–12 CS Standards¹.

4. STATEMENT OF THESIS/PROBLEM

As shown in [5], there is a clear gap between the current state-of-the-art in data management and its representation in CS education. In parts this gap may be explained by the complexity of some topics or because they are only

¹<http://www.csteachers.org/page/SubmitYourFeedback> (checked on 19/05/2016)

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

ICER '16, September 9–11, 2016, Melbourne, Australia.

© Copyright held by the owner/author(s).

ACM ISBN 978-978-1-4503-4449-4/16/09.

<http://dx.doi.org/10.1145/2960310.2960341>.

relevant in specific contexts. But as described in [6], students also need to acquire competencies on various aspects of data management in their own daily lives. Topics like synchronization and backup can serve as interesting contexts, e.g. for discovering the problems but also the advantages of redundant data storage, while today's teaching focuses only on avoiding redundancy (e.g. when normalizing data schemes). Finding such concepts and ideas of data management is a central aspect of my work. Hence, my key question is: "Which are the influences of the current developments in data management on secondary CS education, and how can CS education handle the challenges and requirements that are accompanying these changes and innovations?"

5. RESEARCH GOALS & METHODS

As a research framework for my project, I will use the model of educational reconstruction for CSE [3]. A detailed research plan for my project was described in a work-in-progress paper [7], so I will not describe the application of the framework here. Instead, I will concentrate on the goals and methods. The framework considers four perspectives: science, society, students and teachers. While in the model all these perspectives are equally relevant, not all of them can be examined during one PhD project in equal depth. Hence, I set my focus is on the content-related aspects of the model in particular and on the science perspective, while other aspects are covered in a less detailed way. Consequentially, the most important questions are:

1. How is everyone affected by data management in their daily lives ?
2. Which are the fundamental and long-lasting aspects of data management that are central to CS education?
3. What knowledge, competencies and skills concerning data management does everyone need to have?
4. Which prior knowledge, attitudes and perceptions do learners currently have on this topic as well as on the opportunities and threats of data management?

For answering these questions, I will use several methods: As the focus of the project is on the content-oriented aspects, methods like the qualitative content analysis are central. For investigating questions 2 and 3, I plan to analyze documents like current curricula, educational standards, but also scientific material. For example, to determine the gap between the scientific point-of-view and CS teaching on this topic [5], I conducted an analysis of current CS curricula and standards in comparison to the the DAMA-DMBoK [2]. Also, to ensure the validity of the results, I plan some expert interviews. For dealing with question 4, I am investigating alternative methods to surveys, as directly questioning students is hardly possible in Bavaria because of regulations by the ministry of education. Yet the students' perspectives should not be neglected in this project. A possible option to include at least a general impression of the learners' attitudes and interests is a meta-analysis of studies like the German JIM Study [4] on youth, information and media.

6. DISSERTATION STATUS

In addition to choosing the model of educational reconstruction as a research framework and adapting this model for my project (cf. [7]), some aspects have already been in-

vestigated: from the societal perspective, central competencies that everyone needs for handling personal and others' data in the daily life were analyzed and proposed at the KEYCIT conference (cf. [6]). The scientific perspective was considered by characterizing the gap between current CS education and data management in CS [5]. Also, for gaining a more comprehensive overview of data management, I started creating an overview map of of this field which should help to find central ideas as well as appropriate contexts and phenomena for linking these ideas with students' experiences. In addition I started to investigate students' perspectives by analyzing studies on the media use of teenagers, and I collected feedback from the teachers' perspectives in various teacher training workshops I conducted. At the moment, the thesis document itself only exists as an outline and partly in a rough draft version. However, several of the already published papers will be reused in the thesis.

7. EXPECTED CONTRIBUTIONS

Today, data management is hardly considered in CSE, even though it has a strong relevance in daily life as people use and produce data continuously. By analyzing data management from an educational point-of-view and by finding its central and long-lasting ideas, this project can strongly contribute to the further development of teaching in this field. This will, on the one hand, help modernizing current teaching, but, on the other hand, also prevent just following trends. In addition, by adapting the model of educational reconstruction for analyzing this complete field, it will be investigated if this model is suitable for bringing innovative topics to CSE.

8. REFERENCES

- [1] P. Buffum et al. CS Principles Goes to Middle School: Learning How to Teach "Big Data". In *Proceedings of SIGCSE '14*, pages 151–155. ACM New York, USA, 2014.
- [2] DAMA International. *The DAMA Guide to the Data Management Body of Knowledge - DAMA-DMBOK*. Technics Publications, LLC, USA, 2009.
- [3] I. Diethelm, P. Hubwieser, and R. Klaus. Students, Teachers and Phenomena: Educational Reconstruction for Computer Science Education. In *Proceedings of Koli Calling 2012*, pages 164–173. ACM New York, USA, 2012.
- [4] S. Feierabend, T. Plankenhorn, and T. Rathgeb. *JIM-Studie 2015*. Medienpädagogischer Forschungsverband Südwest, 2015.
- [5] A. Grillenberger and R. Romeike. A Comparison of the Field Data Management and its Representation in Secondary CS Curricula. In *Proceedings of WiPSCE 2014*, pages 29–36. ACM New York, USA, 2014.
- [6] A. Grillenberger and R. Romeike. Teaching Data Management: Key Competencies and Opportunities. In *Proceedings of KEYCIT 2014*, pages 133–150. University Press Potsdam, 2014.
- [7] A. Grillenberger and R. Romeike. Bringing the Innovations in Data Management to CS Education: An Educational Reconstruction Approach. In *Proceedings of WiPSCE 2015*, pages 88–91. ACM New York, USA, 2015.
- [8] H. Witten. Datenbanken - (k)ein Thema im Informatikunterricht? *LOG IN*, 2:14–19, 1994.