Bringing the Innovations in Data Management to CS Education: an Educational Reconstruction Approach

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

Ralf Romeike
Computing Education Research Group
Friedrich-Alexander-Universität Erlangen-Nürnberg
Martensstraße 3, 91058 Erlangen, Germany
ralf.romeike@fau.de

ABSTRACT
This paper describes the application of the research framework *educational reconstruction* for investigating the field *data management* under a CS education perspective. Like the many other innovations in CS, *Big Data* and the field *data management* have strong influences on students' daily lives. In contrast, school does not yet sufficiently prepare students to handle the arising challenges. In this paper we will describe how we apply an educational reconstruction approach to prepare the teaching of essential data management competencies. We will summarize the main goals and principles of educational reconstruction and discuss the application of the framework to the topic data management, as well as first outcomes. Just as educational reconstruction is suitable for finding the essential aspects for teaching data management and for designing classes/courses on this topic, it also seems promising for the curricular development of other CS innovations as well.

CCS Concepts
• Social and professional topics → K-12 education

Keywords
Data Management; Big Data; CS Education; Secondary Schools; Educational Reconstruction

1. INTRODUCTION
As Computer Science is a highly innovative subject and hence is affected by various ongoing changes, an important task of CS education research is to analyze the ongoing developments for topics that can support CS education as well as to prepare bringing aspects to school that are on the one hand central to the science content, but on the other also suitable and useful for CS teaching. Next to the most obvious current development in data management, storing and processing *Big Data*, there are also many other innovations, such as *modern database models* (*NoSQL*), *data stream systems*, *data mining* and *cloud computing*. While some of these are obviously related to aspects of the students' daily life, e. g. in relation to data privacy and security, other aspects like NoSQL are only being used in the background of many applications. Yet, even these topics potentially include ideas and concepts that can be considered as *fundamental ideas of CS* [14]. Since it is considered important *"[...] that we in computing education are focusing on what the students actually need to know, or we risk becoming ineffective and useless*" (Sally Fincher, as cited in [2]), such influences of CS topics on large parts of people's lives can serve as good starting points for a motivating, fruitful and sustainable CS education.

When discussing the relevance of data management for school education, we are confronted with several challenges, such as finding the fundamental aspects from a scientific point of view, selecting appropriate aspects for school, adapting them for CS teaching by developing suitable teaching methods, finding matching examples and connections to the students' lives, and supporting teaching by developing tools. In particular, as data management topics are often highly complex, reducing complexity and adapting them for CS education is a central task.

For handling these challenges and for incorporating aspects in CS education that are characteristic of the ongoing innovations in this field, *educational reconstruction* is an approach that meets the demands: it supports the development of teaching concepts by taking into account not only the scientific point of view, but also the students' perspectives on a topic, social demands, phenomena and the teachers' perspectives. Hence, educational reconstruction is more than the previously mentioned reduction of complexity, as it instead expands a topic by connecting it to the students' perspectives, daily experiences and the social demands. So, with educational reconstruction a topic is not only being adapted for school, but also the usefulness of data management issues for students can be evaluated in this way.

In this paper, we will describe an exemplary application of this research framework to the field *data management*. We will first summarize the central aspects of educational reconstruction and afterwards describe how this framework can support our goals and how we apply it for identifying the aspects of the topic data management that are relevant and useful for secondary school students.

2. EDUCATIONAL RECONSTRUCTION
Educational reconstruction is a research framework focused on educational design that is primarily coming from (natural) science education that *"has been developed as a theoretical framework for studies as to whether it is worthwhile and possible to teach particular content areas of science"* [6] as well as for deciding how to teach it. In order to reach this target, the framework does not only focus on the scientific content structure, but also stresses the importance of taking into consideration the pupils' perspectives and demands.

WiPSCE '15, November 09 - 11, 2015, London, United Kingdom

ACM 978-1-4503-3753-3/15/11…$15.00
DOI: http://dx.doi.org/10.1145/2818314.2818330

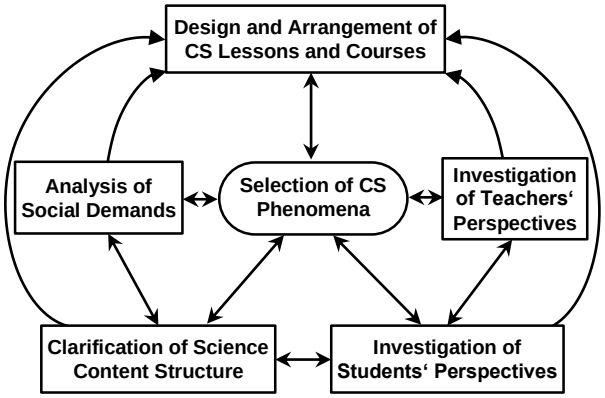

Figure 1: Model of the Educational Reconstruction, adapted for Computer Science Education [5]

According to the original German description of this framework by Kattmann et. al. [11] and the characterization in English by Duit et. al. [6], the *educational reconstruction* approach emphasizes the consideration of three components when preparing scientific content for school teaching: the *clarification and analysis of the science content*, the *investigation of students' perspectives* on this content as well as the *design of the learning environment*. While the analysis of the science content structure from an educational perspective is especially focused on creating a reliable basis for designing a teaching concept, the students' perspectives incorporate not only their perceptions on a topic but also motivational aspects as well as relations to everyday experiences. The central question of the first component is "*What does science know on the topic?"*, while the latter considers "*what do students actually need to know?"*. By comparing the students' perspectives and the scientific point of view on an issue, the selection of contents, structuring them as teaching concept, selecting appropriate teaching methods and so on can happen. These three components strongly influence each other, so much so that they cannot be considered in a separated way. Thus, all these phases need to be passed through multiple times in an iterative way when reconstructing a topic for school teaching. This clearly shows one of the main aims of educational reconstruction: instead of elaborating topics for school education guided by theory and by reducing their complexity for school teaching, this approach is based on the two pillars of scientific knowledge on a topic and the perspectives that the students have on it.

Educational reconstruction has been applied successfully numerous times in the (natural) sciences biology, chemistry and physics. However, CS education to some extent is different: in natural sciences particularly *"there is a strong consensus about the core content of the curricula"* and *"teachers are educated according to a long tradition and to curricula that are at least similar among different countries or states"* [5]. Also, the use of phenomena as starting points and orientation for teaching is very common, while this is not (yet) the case for CS. In order to address this difference, Diethelm et. al. [5] revised the educational reconstruction approach for CSE by adding three additional components: *analyzing the social demands*, selecting *CS phenomena* as starting points as well as *investigating the teachers' perspectives* (cf. fig. 1). Also, the refined model takes into account the relevance of not only developing a lesson and course design, but instead *designing and arranging CS lessons and courses*, which includes planning appropriate "*teaching methods, media or the organization of the lesson*" [5]. This revised approach meets the requirements of our research, as the three added components in particular extend the educational reconstruction by important aspects: with the increasing relevance of data management in various contexts, it is also important to consider such implications in teaching, but also, for example, the strongly differing conditions for CS teaching in different regions and countries. Also, differences in teacher education influence the extent of support and material that teachers need for teaching such modern topics, which can be considered with the teachers' perspectives. When mentioning educational reconstruction, in the following we will hence refer to the revised model for CS education.

3. APPLICATION OF EDUCATIONAL RECONSTRUCTION

In this section we will describe, how we apply educational reconstruction on the field data management in order to adapt this topic for general (secondary) CS education.

3.1 Selection of CS Phenomena

According to Diethelm et. al., *"there is a broad agreement currently within CSE that teaching units should start from a 'real-world' context or phenomenon, aiming to motivate the students, to open connections to prior knowledge or to show application situations of the intended knowledge"* [5]. As Humbert and Puhlmann define phenomena as *"occurrences of informatics in the world"* [10], this approach is also promising for bringing data management to CS education, because of its strong relation to everyone's daily lives. Diethelm et. al. [5] argue, that *"Students have contact with phenomena that are related to information technology every day, partly without even apprehending those"*, which is also the case for the innovations in data management. Hence, the main questions in this step of educational reconstruction are:

- Which phenomena can act as starting points for CS education?
- How can data management contribute to the explanation of the students' everyday lives?
- Is data management able to support tasks and activities of the daily life—even beyond such that are strongly related to CS?

As a starting point for the reconstruction of the topic data management, we selected some exemplary phenomena which most students encounter daily:

- **Cloud Storage and Transactions:** Even when the connection between a client and a cloud storage server is interrupted during a file transfer, typically the version stored on the server is neither incomplete nor faulty, but when restoring the connection, the transfer is being resumed at the point it was stopped instead of being restarted.
- **Synchronization:** Directly after synchronizing data between two devices, various errors occurred: duplicate contacts or additional files with file names like *"file (01-08-2015 11:21).txt"* appear.
- **Data Analysis and Tracking:** After looking for a product in a large online shop, a search engine presents ads with similar products—even if the user never searched for the product using this search engine.
- **Real-Time Data Analysis:** Only a short time after a big event or a catastrophe happened, Twitter presents it as a trend on its website.

This is only a small selection of phenomena in relation to data management that can show the universal relevance of data management concepts in daily life, but also the variety of concepts and topics in this field. By basing teaching on such phenomena, it can be ensured that topics are selected for teaching that are relevant for the students and that also contribute to understanding daily life.

3.2 Analysis of Social Demands

The social demands play an important role in educational reconstruction, as school education is strongly affected by external influences, e. g. legal and curricular requirements, educational standards, requirements of the society to a certain subject as well as the educational system. In particular, an important influence on CS education is the ongoing discussion on the relevance of CS for school teaching: while in this context the focus is often set on the programming aspect, innovations like data management add another perspective on this subject and hence can also emphasize the versatility and usefulness of CS in various contexts, as data management (like many other CS topics) is not only a topic of Computer Science, but also finds its way into private and professional life, e. g. with Cyber-Physical Systems, Smart Phones, Smart Homes, Smart Cars, Smart Factories, and so on. Even with this increasing datafication of our world [13] and the capturing of nearly every aspect of our life as data, today most aspects of data management are hardly considered in CS education at all (cf. [7]). Hence, an important task of education in general and CS education in particular, is to prepare students for the requirements of today's society and everyday life. In this context, important questions to answer are:

- With which demands from society and daily life is general CS education confronted today?
- Which relevance does the science content have in this context?
- What does everyone need to know about data management?

A first step for finding the requirements with which everyone is confronted in the context data management today was a discussion of several key competencies concerning data and data management [8]. Also, some basic requirements on handling data can also be found in educational documents like the CSTA K-12 CS standards: *"[…] ethical issues that relate to computers and networks (e.g., security, privacy, ownership, and information sharing)"* [15]. By considering the social demands, an orientation of the general educative aspects of CS can be ensured.

3.3 Clarification of Science Content Structure

The aim of analyzing the science content structure is not to transfer the science content structure directly to CS education, but to analyze it from an educational point of view. For this task, also taking into account the existing approaches for selecting topics for CS education, like the criteria for *Fundamental Ideas* [14] or the *Great Principles of Computing* [4], is promising. Hence, the main problems in this part are:

- Which are the central terms, ideas, concepts and methods of a topic?
- How is the topic structured from a scientific point of view?
- Which of the central ideas in this field are part of CS education yet and where is a gap between the science content and current CS education?
- How do the potential topics fit into current CS education?

As first steps concerning these questions we focused on a characterization of this field by the Data Management Association (DAMA), the Data Management Body of Knowledge [3], and compared this to the representation of data management in current CS education [7]. This leads us to the gap between science and school concerning these topics: especially, our hypothesis that data management is mainly considered from a database perspective in large parts of the curricula and educational standards was confirmed. Next to relational databases and the relational model, other topics like backup & recovery, acquiring data, meta data, data quality and so on were only mentioned marginally, while many modern approaches like cloud computing are missing. This comparison also led us to a first structure of topics, which are divided on the top-level into *Data Development, Document & Content Management, Data Security Management, Meta Data Management, Data Quality Management* and *Data Usage*. So, these first results are a promising basis for our future work, which can be further improved by analyzing additional scientific literature, such as textbooks on the topic.

3.4 Consideration of Students' Perspectives

According to Kraynova [12], studies have shown that some new teaching approaches could not show learning success because the content structure was not planned appropriately for supporting the way from pre-teaching conceptions to the scientific ones. This emphasizes the relevance of students' perspectives on a topic, which includes motivational, volitional and cognitive aspects, e. g. their pre-knowledge, perceptions and attitudes on a topic, knowledge that they need for understanding everyday experiences, designing lessons in order to ensure a practical and action-oriented teaching and to support students' self-efficacy. Hence, important questions to ask in this context are:

- Which perceptions do students have on the central concepts of data management (e. g. Big Data, Data Mining, Privacy)?
- Which ideas are useful and practicable for students?
- What pre-knowledge do students need for understanding data management concepts and ideas?

In context of educational reconstruction, the students' perspective is often considered by investigating and taking into account the students' perceptions on the concrete topic, which is also becoming increasingly relevant in CS education research today, e. g. the study described in [5]. Other studies on students' perspectives on data management topics can also be found in other disciplines: e. g. in [1], Boyd and Marwick argue that it is *"a widespread myth that American teenagers don't care about privacy"* by examining *"how teens understand privacy and what strategies they take in their efforts to achieve social privacy".*

A first step concerning the students' perspective, which was not focused on students' perceptions, was also done by the previously described analysis of the representation of data management in various CS curricula and educational standards [7]: by analyzing these educational documents, we could find a common core of pre-knowledge on which further topics can be based. These aspects are in particular *storing data in relational structures, using database management systems, aspects of data privacy and security* and some *aspects of data usage*. On the other side, the initial results of determining what students need for handling their data in an appropriate way were also summarized by the collection of key competencies in this field [8].

3.5 Consideration of Teachers' Perspectives

In addition to the students' perceptions, the model also considers the teachers' perspectives on a topic. These perspectives include on the one hand the teachers' perceptions of a topic and of the students' perceptions on it too, but on the other hand also the support, materials, examples, teacher-trainings and so on the teachers need for successfully including a topic in CS education. Incorporating such demands when analyzing a topic for its value for CS education involves the chance to support teachers from the beginning by developing suitable material and tools based on their needs. Hence, the main aspects to investigate in this context are:

- Which knowledge do teachers need about data management?
- How can developing pedagogical content knowledge (PCK) on data management be fostered?

- At which points can teachers be supported, e. g. by developing tools or by selecting suitable materials and examples?

For data management, tools play a particularly important role: although various professional tools for data analysis, visualization and so on, are available, they are often too complex to be used at school. By providing appropriate tools that are especially suitable for CS education as well as by finding matching examples, we can hence strongly support teachers when introducing this topic at school. A first result of our work in this field is the development of a small data stream system, which is optimized for educational use [9]. This data stream system, which is based on the programming environment Snap!, can be used without any special pre-knowledge on data streams, data analysis and so on. At the moment, the only data source which was implemented is the Twitter data stream, but the principle of our implementation can be transferred to many other data sources, including sensor values or website monitoring (if appropriate interfaces/APIs are provided).

3.6 Design and Arrangement of CS Lessons

Based on the previously described components and considerations, CS lessons and courses can be designed and arranged. As educational reconstruction is designed as an iterative process, the lesson/course design will also be refined in each iteration. Therefore, the results from all other components will be used for designing and arranging CS lessons and courses by selecting such topics that are relevant from both, the CS point of view as well as the students' one, but which are also feasible for the teachers. Also, selecting the appropriate phenomena as starting points for these topics and including the social demands will be done in this component. Hence, developing concrete teaching concepts is dependent on all the other components, so that at the current state of our work, describing concrete results of this phase is thus not possible.

During this phase, fitting this topic in current school teaching is an important challenge. While current CS education in this field mainly focuses on a database perspective, the main concepts that are taught are also included in other parts of data management: for example, the concept *redundancy*, which is at the moment deliberately being avoided in order to prevent inconsistencies. However, this concept can for example also be addressed from a *backup* perspective: when creating backups of data, redundant data storage is required. Hence, in our lesson design, an important aspect will be to cover the discussed concepts—wherever applicable—not from one perspective alone, in order to prevent misconceptions like "*redundancy should always be prevented"*.

4. SUMMARY

In this paper, we described how *educational reconstruction* can be applied to the task of bringing the innovations in *data management* to (secondary) CS education, as well as some initial results we achieved in our previous work on this topic. Many of these innovations comprise important concepts and ideas that are central to this topic and that are also promising to fulfill the criteria for fundamental ideas in CS education. Hence, bringing these innovative but also fundamental aspects to CS education using educational reconstruction as a framework is a promising approach. In this sense, the described approach can also be considered as exemplary project for applying educational reconstruction to the ongoing innovations in CS and for bringing the central and suitable aspects to CS education. Especially by considering the social demands but also the students' and teachers' perspectives on a topic, this framework can be used for designing CS lessons and courses with respect to the general educative purpose of secondary CS education.

5. REFERENCES

[1] D. Boyd and A. E. Marwick. 2011. Social Privacy in Networked Publics: Teens' Attitudes, Practices, and Strategies. In *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*.

[2] S. Cooper, S. Grover, M. Guzdial, and B. Simon. A Future for Computing Education Research. *Commun. ACM* 57, 11, 34–36. 2014.

[3] DAMA International. 2009. *The DAMA Guide to the Data Management Body of Knowledge - DAMA-DMBOK*. Technics Publications, LLC, USA.

[4] P. J. Denning. Great Principles of Computing. *Commun. ACM* 46, 11, 15–20. 2003.

[5] I. Diethelm, P. Hubwieser, and R. Klaus. 2012. Students, Teachers and Phenomena: Educational Reconstruction for Computer Science Education. In *Proceedings of the 12th Koli Calling International Conference on Computing Education Research*. Koli Calling '12. ACM, New York.

[6] R. Duit, H. Gropengießer, U. Kattmann, M. Komorek, and I. Parchmann. 2012. The Model of Educational Reconstruction - a Framework for Improving Teaching and Learning Science. In *Science Education Research and Practice in Europe*, D. Jorde and J. Dillon, Eds. Cultural Perpectives in Science Education. SensePublishers, 13–37.

[7] A. Grillenberger and R. Romeike. 2014. A Comparison of the Field Data Management and its Representation in Secondary CS Curricula. In *Proceedings of WiPSCE 2014*. ACM, Berlin.

[8] A. Grillenberger and R. Romeike. 2014. Teaching Data Management: Key Competencies and Opportunities. In *KEYCIT 2014 - Key Competencies in Informatics and ICT*. Commentarii informaticae didacticae. Universitätsverlag Potsdam.

[9] A. Grillenberger and R. Romeike. 2015. Analyzing the Twitter Data Stream Using the Snap! Learning Environment. In *Proceedings of ISSEP 2015*.

[10] L. Humbert and H. Puhlmann. Essential ingredients of literacy in informatics. *Informatics and Student Assessment. Concepts of empirical research and standardisation of measurement in the area of didactics of informatics*. 2004.

[11] U. Kattmann, R. Duit, H. Gropengießer, and M. Komorek. Das Modell der didaktischen Rekonstruktion - Ein Rahmen für naturwissenschaftsdidaktische Forschung und Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften* 3, 3, 3–18. 1997.

[12] Kraynova, A. 2012. *Didaktische Rekonstruktion der Nanophysik*. Univ. Oldenburg.

[13] Mayer-Schönberger, V. and Cukier, K. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt.

[14] A. Schwill. Fundamental ideas of computer science. *Bull. European Assoc. for Theoretical Computer Science* 53. 1994.

[15] Seehorn, D., Carey, S., Fuschetto, B., Lee, I., Moix, D., O'Grady-Cunniff, D., Boucher Owens, B., Stephenson, C., and Verno, A. 2011. *K-12 Computer Science Standards*. Computer Science Teachers Association, ACM.