A Comparison of the Field Data Management and its Representation in Secondary CS Curricula

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

In the last few years, the focus of *data management* has changed from handling relatively small amounts of data, often in relational databases, to managing large amounts of data using various different database types. In many secondary school curricula, *data management* is mainly considered from a *"database"* perspective. However, in contrast to the developments in computer science research and practice, the new and changing aspects of data management have hardly been discussed with respect to CS education. We suggest re-evaluating the focus and relevance of the established database syllabi, to discuss the educational value of the newly arising developments and to prevent the teaching of outdated concepts. In this paper, we will contrast current educational standards and curricula with an up-to-date characterization of data management in order to identify gaps between the principles and concepts of data management that are considered as important today from a professional point of view on the one side, and the emphasis in current CS education on the other side.

The findings of this analysis will provide a basis for aligning the concepts taught in CS education with the developments in data management research and practice, as well as for re-evaluating the educational value of these concepts.

Categories and Subject Descriptors

K.3.2 [**Computers and Education**]: Computer and Information Science Education—*computer science education, curriculum*

Keywords

Analysis, characterization, curricula, data management, databases, secondary school, standards

WiPSCE'14, November 05 - 07 2014, Berlin, Germany.

ACM 978-1-4503-3250-7/14/11 ...$15.00.
http://dx.doi.org/10.1145/2670757.2670779.

1. INTRODUCTION

Computer science education has always been faced with the difficult challenge of meeting the rapid changes and developments in computing. Teaching needs to foster skills that meet the state-of-the-art in this rapidly evolving science. Also, examples for classroom use should be oriented at current systems. But this seems to be difficult, since a lot of knowledge and methods are rapidly being overtaken by the ongoing developments in this field.

A possible solution for handling this dilemma is to emphasize the *fundamental ideas* of computer science in schools. Schubert & Schwill [18, 19] argue that despite numerous paradigm changes, computer science has hardly been shaken in its history; obviously a basis of ideas and principles exist that remain stable despite these changes. Such ideas can also serve as guidance for computer science education: instead of imparting short-living knowledge, teachers should emphasize the long-lasting ideas that are also perceptible from a historical perspective (*criterion of time*) and that are used in different areas of the science (*horizontal criterion*). Peter Denning also developed a comparable approach, the *Great Principles of Computing* [7], which may also be used for identifying and evaluating relevant topics of computer science education.

In the field of *databases*, it seems as if a coherent and widely accepted set of such ideas and principles has already been found: next to algorithmic thinking, the topic "data modeling" plays an unchanged and prominent role in different educational contexts (e. g. curricula, educational standards). In particular, relational database management systems and various tools for accessing them play an essential role. In the German didactic community[1], the concept of relational databases proposed by Codd [6] in 1970 has been discussed as a topic for CS education relatively early (e. g. in 1987 by Borg [2]). Later discussions mainly focused on the concrete design of teaching as well as using and creating appropriate tools for database education. For example, every 14-year-old student at Bavarian high schools with a technical-scientific focus ("Naturwissenschaftlich-technologisches Gymnasium") learns the principles of data modeling as well as how to implement these models using relational database management systems: *"The students structure data on examples from their living environment. [...] For using the model and testing its usability, they implement it using a*

[1]Due to the lack of available literature in the international discussion on CS education in the 1980s and 1990s, we will focus on the situation in Germany at this point.

relational database system." [11] [2]. This point of view on using data in relational databases was introduced into schools in the early 1990s and has hardly been changed since then[3].

While in practical use, the relational model was also dominating for a long time and hence could show the ideas and principles of CS at its best, today the domination of relational databases is challenged due to the developments in this field within the last years. For example, with Big Data and NoSQL, new types of databases have evolved in order to be able to handle the increasing amount of data everyone manages today, as well as for using the advantages of distributed databases that allow parallel data processing. Also, today databases typically need to be highly available on the internet using various clients, e. g. web applications. Several ideas and principles of the newly arising non-relational, so-called NoSQL[4], databases differ from the ones of relational databases (cf. [9]). Thus, ideas in this field that were considered as essential and fundamental goals so far, lose their importance in the context of new requirements and therefore also need to be re-evaluated on their importance for teaching. For example, the relevance of redundancy and consistency clearly changes when considering distributed data storage[5].

Revisiting and re-evaluating the topic data management also opens up the chance to analyze topics and content, which were so far considered as fundamental for the field databases in terms of their underlying ideas and principles. At the moment, an often taught principle when teaching data modeling is to normalize database schemes up to the third normal form in order to prevent any potential redundancies and inconsistencies. However, a discussion of the ideas behind these concepts and their use in other parts of CS (cf. *horizontal criterion* [19]) is not known to us. This raises the question of whether current curricula and educational standards for secondary CS education are covering the state-of-the-art in computer science as well as the fundamental ideas, concepts and principles that are arising due to the advances in this field, and if not, which of the missing aspects are important in class.

2. DATA MANAGEMENT

2.1 Developments in Data Management

So far, databases are the predominant topic of data management in school contexts. Its importance for teaching was especially discussed in the early 1990s (cf. e. g. [21]). As there were only few changes in this field for several years, only a few new approaches for teaching this topic in CS education were proposed (e. g. by Antonitsch [1]). Instead, the educational research in this field mainly concentrated on tools for simplifying, modernizing and subsequently improving teaching of the topics *databases* and *SQL* (cf. [9]). In contrast to its representation in teaching, the field data management and its topic databases have changed tremendously within the last few years: although the relational database management systems (RDBMS) are dominating since they were proposed by Codd in 1970 [6], not only new types of DBMS evolved, like the much-discussed NoSQL databases, but also the kind of data being managed using these tools has changed tremendously. While in the past relatively small amounts of structured data were stored at mainly one site / in one database, nowadays often large amounts of data with varying structures are stored in distributed databases. This reflects the main characteristics of Big Data, a development which is often summarized by the three V's, namely *volume*, *velocity* and *variety* [13]. The progress in this field can also be recognized by having a look at popular textbooks on databases. The German book on database systems by Kemper & Eickler [12] was not only expanded from about 450 pages in 1996 to nearly 850 pages today, but also seven new chapters on popular topics like "data mining", "in-memory databases", "internet database connections" or "Big Data" were added. Another clear hint regarding the ongoing developments in this field within the last years is the number of articles published in the ACM digital library[6] per year related to the search term "database": this number was more than ten times higher in 2013 (about 7400 articles) than it was in 1990 (about 650).

These developments suggest that the topic data management needs to be re-evaluated in context of general secondary CS education. The new topics in this field of computer science may be important to be considered in teaching, but also common concepts may have changed since this topic was discussed by the CS education community in the 1990s. In the pages that follow, we will describe a study with the aim of identifying the state-of-the-art of data management teaching and its differences to the developments in data management research.

2.2 Research Questions

The results of this study will provide a basis for further discussions on the consequences of the described developments for CS education, e. g. on whether and which new topics should be introduced into teaching. Hence, we will focus on the following three research questions:

1. Which aspects of data management are represented in current recommendations and guidelines for general secondary computer science education?

2. What is the gap between the aspects of data management that are currently included as topics in CS education and the aspects of data management considered as being characterizing for this field from a professional point of view?

3. Could data management serve as a guideline for bringing together important aspects that are only considered marginally in current CS education?

The first two of these questions are strongly interrelated: while the first one emphasizes the topics that are taught in the state-of-the-art of CS education, the second question is

[2]Original in German, translated by the authors.

[3]When bringing object-oriented programming into classrooms, the object-oriented view on databases was also introduced in secondary education. But this has hardly changed the underlying (data) model itself.

[4]Today, NoSQL is usually interpreted as "not only SQL" [8] and is used as a general term for non-relational databases.

[5]The CAP theorem [3] (also known as Brewer's theorem) clearly shows that only two of the three properties "availability", "consistency" and "partition tolerance" can be achieved at the same time. Therefore, when using distributed databases, one of the main decisions is to deliberate over whether redundancy and consistency are required.

[6]http://dl.acm.org, last checked on October 6th 2014.

concerned with the differences between these aspects and the ones that are regarded as important by the data management community. The third question instead focuses on whether the term "data management" can bring together several school topics of CS concerned with the handling of data, such as security, privacy, and the legal aspects.

3. ANALYSIS PROCEDURE

To answer the research questions, we will analyze the major relevant educational standards as well as various curricula for general secondary CS education. This will provide us with an overview of the relevance of different data management aspects in current teaching. In general, such text analyses may be done in either a qualitative or a quantitative way. As our target is to characterize data management in the context of secondary CS education, we will use the qualitative content analysis approach as proposed by Mayring [14]. A similar study using this method was also completed by Bröker, Kastens & Magenheim [5] who analyzed higher education curricula on the considered topics using this method in order to provide a basis for developing an empirically oriented competency model for undergraduate students.

In this section, we will describe the methodological approach as well as the development of our category system, the sample of materials which were chosen for analysis and the analysis process itself.

3.1 Qualitative Content Analysis

An important task during qualitative content analysis, as described by Mayring [14], is selecting the material that will be examined in the analysis. As this material is typically a sample, we needed to decide how much and which material to choose, while keeping in mind that this sample must be a representative selection in order to ensure valid results. Thereafter, another fundamental step is defining the category system, which will be used for coding the documents afterwards. In our case, this system will characterize the topics of the field data management that might occur in the analyzed documents. Such a category system may be derived deductively from an existing theory or inductively by analyzing the texts directly and simultaneously building up the category system. These methods may also be combined, which is what we will do in this work.

After deriving the category system from an existing characterization of data management (deductive aspect), we will code the material using a software tool intended for qualitative data analysis. Therefore, we can on one hand ensure the completeness of our category system, because missing topics are added during the analysis phase (inductive aspect of deriving the category system). On the other hand, this analysis will provide an overview on the coverage of data management topics in different educational standards and curricula as well as on their characteristics.

3.2 Analyzed Material

For our content analysis, we have chosen various international material, but we will also set a focus on material coming from several German federal states. We decided for this division in order to broadly consider different orientations, characteristics and possibilities by using the international documents on one hand, but on the other hand to also go into detail as well as to reduce heterogeneity, achieve comparability and to be able to draw conclusions for the local curriculum development using several German curricula. Hence, we will for instance analyze the common educational standards by the Computer Science Teachers Association (CSTA) [20] as well as the ones by the German Informatics Society (GI) [4, 17][7]. In addition to this broad view on the field, we have chosen curricula from Austria, Israel and Canada, as well as from various German states' upper secondary schools ("Gymnasium") and the Computing at School curriculum[8]. In the following pages, you can find the full list of documents that we analyzed together with an abbreviation which we will use for later referrals:

- *EPA:* German high school examination requirements
- *GI:* German Educational Standards for Computer Science in Lower Secondary Education[7]
- *K12:* CSTA/ACM K–12 Computer Science Standards
- *BY:* Curr. for the "Gymnasium" in Bavaria, Germany
- *HE:* Curr. for the "Gymnasium" in Hessen, Germany
- *HH:* Curr. for the "Gymnasium" in Hamburg, Germany
- *NRW:* Curr. for the "Gymnasium" in North-Rhine-Westphalia, Germany
- *RLP:* Curr. for the "Gymnasium" in Rhineland-Palatinate, Germany
- *AT:* Curr. for the "Allgemeinbildende Höhere Schule" (AHS), Austria
- *CA:* Ontario CS curriculum, Canada
- *CSC:* Computing at School Curriculum[8]
- *IS:* Curr. for a High-School Program in CS, Israel

3.3 Category System

After selecting the material for our analysis, the next step was to develop the category system. We decided to derive the basis of this system deductively, because a completely inductive development would only describe the current state-of-the-art as the categories are derived from the analyzed material. In contrast, a deductive derivation can also show gaps between current teaching and the field of CS research as the categories are based on a theory or other material. Since considering the current developments is one of our main goals, we decided to mainly base our category system on this method. Because there is no analysis of the relevant aspects of data management as a topic in an educational context yet, we decided to base our categories on a professional characterization of this field in order to clarify the gap between the state-of-the-art of data management in CS and its current representation in CS education. The only such detailed characterization was found in the Data Management Book of Knowledge (DMBoK) by the Data Management Association (DAMA) [16]. This framework provides a broad overview of this field, e. g. by describing the main functions

[7]For the analysis, we used the original German version [17] as it is more detailed than the summary in English by Brinda et. al. [4]

[8]http://www.computingatschool.org.uk/?id=cacfs, last checked on October 6th 2014.

of data management as well as activities, tools, and so on from a professional point of view. To derive the categories for our analysis, we will focus on the functions described in this characterization, as they are referred to as the main concepts of data management[9] and hence meet the topics in class at its best. The functions described in the DMBoK are:

- **Data Governance**:
  Planning and controlling data management processes at an enterprise level.
- **Data Architecture Management**:
  Planning and managing complex data structures.
- **Data Development**:
  Managing structured data from design up to maintenance (mainly using databases).
- **Data Operations Management**:
  Maintenance, support and administration of data management solutions.
- **Data Security Management**:
  Aspects of data privacy and data security with respect to data management.
- **Reference & Master Data Management**:
  Ensuring a consistent use of data in various systems.
- **Data Warehousing & Business Intelligence Management**:
  Using data analysis for decision-finding, especially emphasizing data warehousing and business intelligence.
- **Document & Content Management**:
  Storing, structuring and managing unstructured data (mainly without databases).
- **Meta-data Management**:
  Using meta-data as well as managing how and where to use meta-data.
- **Data Quality Management**:
  Measures for ensuring, raising and controlling data quality.

Some of these aspects are high-level functions that are especially relevant for managing data on an enterprise level with complex data architectures and operations processes, but hardly for general secondary computing education. The function *data governance* mainly considers developing and supervising guidelines on data usage from an enterprise-wide view. *Reference & master data management* emphasizes the replication of data over multiple systems as well as common data usage through system boundaries. These aspects require high-level knowledge on data management and thus are hardly reasonable for use at a school. However, the basic principles of *data architecture management* are also contained in data modeling. Also, the basics of replicating data between multiple systems are contained in *reference & master data management*, as it has various aspects in common with data synchronization. Hence, the basic principles of these aspects are also contained in other parts of data management on a lower level that is more appropriate for use at secondary schools. In addition, *data operations management* emphasizes aspects of maintenance and support that are hardly provided in a secondary school level since projects are mostly planned only for a short lifetime. Consequently, we will not consider these functions in our category system. Another very specialized function is *data warehousing & business intelligence management*. This function covers a specialized use of data in enterprise environments: using data for decision-making and reporting. While the principles of data warehousing and business intelligence are too complex and extensive for general teaching in schools, there are other examples for using data that are strongly related to students' daily life. Therefore, in our category system, we will replace this function by *data usage* with a broader view on using data in various contexts.

Based on these considerations, we built up the base of our category system. In addition to the described top-level categories, we also included various subcategories coming from the DMBoK for a more detailed view. By adding complementary subcategories, we also allowed the category system to be expanded during the analysis phase in order to cover aspects that were not yet part of the existing categories, but only if they fit into the description of the top-level category. In the following, you can find an overview of our final category system, which also includes the inductively derived categories (marked by [+]).

1. Data Development
   - 1.1 Data Modeling
     - 1.1.1 non-relational model [+]
     - 1.1.2 object-oriented model [+]
     - 1.1.3 relational model [+]
   - 1.2 Implementation
     - 1.2.1 Database Management System
       - A. Non-relational [+]
       - B. Relational [+]
     - 1.2.2 Query Language
2. Document & Content Management
   - 2.1 Acquisition / Retrieval
   - 2.2 Storage
   - 2.3 Backup & Recovery
   - 2.4 Content Management
   - 2.5 Retention
   - 2.6 Purging
3. Data Security Management
   - 3.1 Data Security
   - 3.2 Data Privacy
   - 3.3 Access control
   - 3.4 Encryption
4. Meta Data Management
5. Data Quality Management
   - 5.1 Integrity

[9] In a draft of the next version of DMBoK, these functions are instead renamed as *knowledge areas*, which better fits the characterization as basic concepts of data management.

5.1.1 Consistency
5.1.2 Redundancy

5.2 Data Accuracy, Reliability & Completeness [+]

6. Data Usage

6.1 Data Analysis

6.2 Data Interpretation [+]

6.3 Data Sharing [+]

6.4 Large amounts of data [+]

6.5 Legal, social and ethical aspects [+]

3.4 Analysis Process

After determining the category system, we coded the material listed before using the tool MaxQDA[10]. The main goal of our text analysis was to explore topics that are commonly taught in the field of data management in order to compare them to the presentation of the field as described by DAMA. Because of this and because such documents strongly vary in how detailed they are, we considered the coding on a per-material level as Boolean values. Hence, we concentrated on whether a topic was found in the document or not. As the exact number of coding of one topic within the material did not affect the results of our analysis, no restriction on the length of the coding units and hence on whether to code single keywords or larger phrases was necessary. However, it was taken into account that some occurrences of topics in the documents may only be used for delimiting other topics. So when coding the documents, it was ensured that only real instructional use of the topics was coded. In addition, directly coding the top-level categories was avoided, except for *meta-data management* (which was not refined into subcategories), as these categories only provide a very broad overview, while a more detailed view was subsequently desired. In contrast, we considered each upper category as mentioned when at least one of its subcategories was set. If coding an occurrence was not possible because a required subcategory simply did not exist, we manually created it in a temporary section. At the end of the analysis process, we reviewed these and decided whether to add them to the category system depending on if this category was consistent with the DMBoK. This led to the (sub-)categories *data interpretation*; *data sharing*; *legal, social and ethical aspects*; *data accuracy, reliability & completeness*, as well as to splitting of the categories *data modeling* and *database management systems* into *relational* and *non-relational* accordingly. Also, *object-oriented model* was added, as this topic was mentioned in one curriculum. In all other cases, the category system matched the occurrences in the material; especially since there were no candidates for categories that were contradictory to the DMBoK.

4. RESULTS

The results are shown in table 1, while in addition we visualized the occurrence percentage of the topics in figure 1. In the following text, we will describe these results in relation to the corresponding research questions as described in section 2.2.

[10]http://www.maxqda.com, last checked on October 6th 2014.

4.1 Representation of Data Management in CS Education

The analysis results show that there is a clear overlap between the topics mentioned in various curricula and educational standards for computer science education and the characterization of data management provided by DAMA in their Book of Knowledge [16]. Especially, overall all the top categories in our category system are covered in the analyzed documents; however, not all subcategories and not in every document.

In the analyzed curricula and educational standards, there is a dominance of aspects related to managing structured data (categories 1.1, 1.2), while instances of managing unstructured data (2.1–2.6) are relatively less considered. For example, the Canadian curriculum (*CA*) mentions the competency to *"demonstrate the ability to read from, and write to, an external file (e.g., sequential file, database, XML file, relational database via SQL)"* [15], while there is no hint on topics like storing less structured data e. g. in the cloud.

4.2 Characterization of the Gap between Data Management in CS and CS Education

The results also show that the categories selected for our analysis in general meet the topics covered by the current state-of-the-art in CS education. In contrast, there are also various categories which are not or are hardly considered in current CS education and hence need to be further discussed according to their relevance to this field. All aspects that were added during the analysis were consistent to the deductively derived category system (cf. section 3.3) as they only added more details.

This shows the gap between current CS education and the field data management in CS: current teaching covers relatively well-known topics, but numerous modern topics like managing data in non-relational databases (1.1.1, 1.2.1.1) or aspects of data analysis (6.1) are often missing. In addition, table 1 also shows a clear consensus between multiple materials: the German curricula (BY, HH, HE, RLP, NRW) are matching most of the aspects mentioned in the German educational standards (GI), but there is also a large overlap with international curricula (AT, CA, CSC). In particular, *data development* (which for example includes *relational databases*) is represented in all materials, while *data security management* (including *data privacy*) was found in 10 out of 12 materials. Instead, *document & content management* was found in seven, *data quality management* in three out of 12 materials. The topic *meta-data* is only covered marginally in one curriculum: when talking about meta-tags on websites and the corresponding benefits for search engines in the curriculum of Hessen, Germany (HE). Also, various examples for the use of data are provided in almost all of the analyzed materials.

4.3 Data Management as Guideline in CS Education

During the analysis, we could only find a few topics that we considered as related to managing data, but which were not covered by the categories derived from the DMBoK. These seven differences were previously marked by [+] in section 3.3 when introducing our category system: at first, the categories *1.1 Data Modeling* as well as *1.2.1 Database*

Table 1: Overview of the coding results. Columns show the material abbreviated according to section 3.2, rows represent the categories according to section 3.3. Categories which are covered in at least 80% of the analyzed material are highlighted gray.

Category	EPA	GI	K12	BY	HE	HH	NRW	RLP	AT	CA	CSC	IS
1. Data Development	×	×	×	×	×	×	×	×	×	×	×	×
1.1 Data Modeling	×	×	×	×	×	×	×	×			×	×
1.1.1 non-relational model												
1.1.2 object-oriented model				×								
1.1.3 relational model	×	×						×			×	×
1.2 Implementation	×	×		×	×	×	×	×	×	×	×	×
1.2.1 Database Management System	×	×		×	×	×	×	×	×	×		×
1.2.1.1 non-relational DBMS					×							
1.2.1.2 relational DBMS				×				×		×		×
1.2.2 Query Language				×	×	×	×	×		×		×
2. Document & Content Management		×	×	×	×	×	×		×	×	×	
2.1 Acquisition & Retrieval			×						×	×		
2.2 Storage		×		×	×		×			×	×	
2.3 Backup & Recovery										×		
2.4 Content Management		×		×		×	×					
2.5 Retention												
2.6 Purging												
3. Data Security Management	×	×	×	×	×	×	×	×		×	×	
3.1 Data Security	×		×	×	×		×	×		×	×	
3.2 Data Privacy	×	×	×	×	×	×	×	×		×	×	
3.3 Access control			×	×		×	×	×				
3.4 Encryption		×	×			×	×	×				
4. Meta Data Management					×							
5. Data Quality Management		×		×	×		×	×			×	×
5.1 Integrity				×	×		×	×			×	×
5.1.1 Consistency				×				×			×	
5.1.2 Redundancy				×				×				×
5.2 Data Accuracy, Reliability & Completeness		×		×							×	
6. Data Usage	×	×	×	×		×	×	×	×	×	×	
6.1 Data Analysis		×	×							×		
6.2 Data Interpretation	×	×					×					
6.3 Data Sharing		×		×		×	×			×		
6.4 Large amounts of data			×						×		×	
6.5 Legal, social and ethical aspects		×	×			×	×	×	×	×		

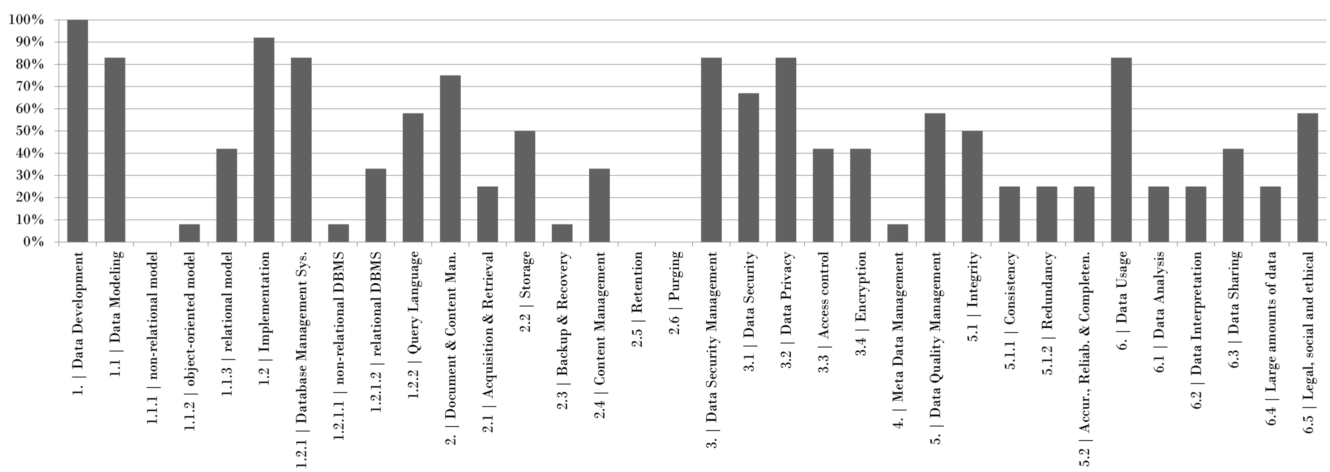

Figure 1: Percentage of material covering the analyzed topics/categories.

Management Systems were split into a *relational* and a *non-relational* category each, and the category *object-oriented model* was added in order to analyze whether or not an object-oriented view on the data model is used in current CS education. In addition, we added the category named *5.2 Data Accuracy, Reliability & Completeness*, which is consistent with the description of data quality management but not explicitly mentioned in the DMBoK, as well as the categories *6.2 Data Interpretation*, *6.3 Data Sharing*, *6.4 Large amounts of data* and *6.5 Legal, social and ethical aspects* that are only additional examples of using data.

Since all these aspects are consistent to the characterization of data management by the DAMA, our analysis shows that the topic data management can bring together various aspects considered as relevant in current CS education. This also includes important aspects like the ones summarized under the term *Informatics, Man and Society* by Brinda et. al. in the educational standards by the German Informatics Society [4]. For example, in this field students should acquire the competencies to *"be aware of being able to make decisions in the use of informatics systems and to adhere to social norms in usage"* as well as to *"be able to restrict to risks involved in using informatics systems"*.

5. DISCUSSION

Our results show that the state-of-the-art CS education especially emphasizes managing structured data with (relational) databases as well as strongly related topics. For example, the curriculum for the secondary school "Gymnasium" in Bavaria, Germany (*BY*), mentions the competency to *"implement objects, classes and relationships in a relational database management system"* [11] [11]. These topics were originally brought to CS education during the 1990s. Since then, only a few new aspects were included into CS education. Modern aspects like "*data mining*", "*internet database connections*" or "*big data*", which have been included in the aforementioned textbook by Kemper and Eickler [12] in the previous years, are generally not part of today's CS education. Since these topics also have a strong influence on our society, there is a clear gap between the requirements of handling and managing data in daily life and the current CS education: Grillenberger & Romeike [10] discuss key competencies that everyone needs today to successfully handle data and to overcome the newly arising challenges experienced in daily life. These key competencies include aspects like using data backup and recovery, encrypting data, chances and threats of meta-data as well as aspects of data privacy and security. While the first ones of these topics are missing in most of the analyzed documents, the latter ones are included.

However these aspects are often only mentioned marginally: for instance, students should *"understand ethical issues that relate to computers and networks (e.g. equity of access, security, privacy, copyright, and intellectual property)"* according to the CSTA K–12 Computer Science Standards [20]. These descriptions are much vaguer than more technical topics such as *"pupils can design and use complex data structures including relational databases"* [20], which describes concrete and assessable goals. Hence, a presumption that needs to be further examined in future work is that such topics are often taught without detailed contextualization in the field of computer science and thus, might clearly profit from being further discussed from a data management perspective.

Another example of how data management can bring together topics that are widely spread over current CS education is the single occurrence of the topic *meta-data* that we found in the curriculum for the "Gymnasium" in Hessen, Germany (*HE*). This curriculum states that students should realize meta-tags on websites as a criterion of quality and subsequently use them to control search engines. However, this is only one example of the opportunity of using meta-data, but there are various other examples and principles for using meta-data. These principles are often not considered in current teaching; hence, students often cannot realize that the same ideas may apply for handling their videos, music and documents. On the other hand, meta-data can also lead to serious threats regarding data security and privacy.

Although this analysis has shown a clear gap between computer science research and CS education in the field of data management, it was not intended to discuss the relevance of different topics/aspects of data management in comparison to others or for deciding to include aspects of data management into curricula only based on this study. Even though aspects like data privacy are mentioned in almost all analyzed documents, we cannot make statements about their relevance for teaching. In addition, since the analysis is solely based on standards and curricula, we cannot conclude how and if these topics in practice are implemented in everyday schooling. Lessons may clearly differ from the learning objectives described in the analyzed material, since other topics may be wittingly or unwittingly discussed during class, e. g. as a side effect in examples or because teachers set different priorities.

In addition, this analysis can only identify the gap between CS research and education, but not if or how closing this gap needs to be overcome: The results obtained cannot show whether the lack of these topics in school affects the daily life of the learners or if they are easily able to learn to manage these topics otherwise. Hence, analyzing the learners' attitudes, perceptions, pre-knowledge and competencies in this field is another important task to bring the promising aspects of data and data management into secondary CS education.

Summarizing, our study shows that the topics of data management and the topics concerning databases and handling data in current CS education are matching: we could not find any topics in the analyzed educational standards and curricula that do not fit into the description of data management by the Data Management Association (DAMA) [16]. However, various additional aspects of data management are not considered at present in general CS education, especially such that arose from the developments in the field data management during the last years. Hence, our future work will focus on finding the important concepts, principles and fundamental ideas of data management in order to involve the long-lasting aspects of the current developments in the field data management into CS education.

[11] Original in German, translated by the authors.

6. 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] B. Borg. Didaktisch-methodische Aspekte des Einsatzes von Datenbanksystemen [didactical-methodological aspects in the use of database systems]. *LOG IN*, 5/6, 1987.

[3] E. Brewer. CAP twelve years later: How the "rules" have changed. *Computer*, 45(2):23–29, 2012.

[4] T. Brinda, H. Puhlmann, and C. Schulte. Bridging ICT and CS: Educational Standards for Computer Science in Lower Secondary Education. In *Proceedings of the 14th Annual ACM SIGCSE Conference on Innovation and Technology in Computer Science Education*, ITiCSE '09, pages 288–292, New York, NY, USA, 2009. ACM.

[5] K. Bröker, U. Kastens, and J. Magenheim. Competences of Undergraduate Computer Science Students. In T. Brinda, N. Reynolds, and R. Romeike, editors, *KEYCIT 2014 – Key Competencies in Informatics and ICT*, Commentarii informaticae didacticae. Universitätsverlag Potsdam, 2014.

[6] E. F. Codd. A Relational Model of Data for Large Shared Data Banks. *Commun. ACM*, 13(6):377–387, 1970.

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

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

[9] A. Grillenberger and R. Romeike. Big Data – Challenges for Computer Science Education. In Y. Gülbahar and E. Karataş, editors, *Informatics in Schools. Teaching and Learning Perspectives*, volume 8730 of *Lecture Notes in Computer Science*, pages 29–40. Springer International Publishing, 2014.

[10] A. Grillenberger and R. Romeike. Teaching Data Management: Key Competencies and Opportunities. In T. Brinda, N. Reynolds, and R. Romeike, editors, *KEYCIT 2014 – Key Competencies in Informatics and ICT*, Commentarii informaticae didacticae. Universitätsverlag Potsdam, 2014.

[11] ISB (Staatsinstitut für Schulqualität und Bildungsforschung). Lehrplan für das Gymnasium in Bayern, Fach Natur und Technik [Curriculum for the Bavarian Secondary School, Subject "Informatics", Scientific & Technical Branch], 2009.

[12] A. Kemper and A. Eickler. *Datenbanksysteme: Eine Einführung [Database systems: an introduction]*. Oldenbourg Wissensch.Vlg, 2013.

[13] D. Laney. 3D Data Management: Controlling Data Volume, Velocity, and Variety. Technical report, META Group, February 2001.

[14] P. Mayring. Qualitative Content Analysis. In *A companion to qualitative research*, pages 266–269. Sage, 2004.

[15] Ministry of Education. The Ontario Curriculum, Grades 10 to 12: Computer Studies. 2008 (revised). http://www.edu.gov.on.ca/eng/curriculum/secondary/computer.html.

[16] M. Mosley, M. Brackett, and S. Earley. *The DAMA Guide to the Data Management Body of Knowledge (DAMA-DMBOK Guide)*. Technics Publications, LLC, 2010.

[17] H. Puhlmann et al. Grundsätze und Standards für die Informatik in der Schule: Bildungsstandards Informatik für die Sekundarstufe I [Principles and Standards for Computer Science in Schools: Educational Standards in Computer Science for lower secondary schools]. *LOG IN*, 150/151, 2008.

[18] S. Schubert and A. Schwill. *Didaktik Der Informatik [Didactics of Informatics]*. Springer DE, Berlin, 2011.

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

[20] D. Seehorn et al. *K–12 Computer Science Standards*. Computer Science Teachers Association, Association for Computing Machinery, 2011.

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