Teaching Data Management: Key Competencies and Opportunities

Anonymized, *anon@example.com*
Chair, University
Street Address

Abstract
Data management is a central topic in computer science as well as in computer science education. Within the last years, this topic is changing tremendously, as its impact on daily life becomes increasingly visible. Nowadays, everyone not only manages data but also generates large amounts of data continuously. In addition, Big Data and data analysis are intensively discussed in public dialogue because of their influences on society. For the understanding of such discussions and for being able to participate in them, fundamental knowledge on data management is necessary. Especially, being aware on the threats accompanying with the ability to analyze such large amounts of data in nearly real-time becomes increasingly important. This raises the question, which key competencies are necessary for daily dealings with data and data management.
In this paper, we will first point out the importance of data management and of Big Data in daily life. On this basis, we will analyze which are the key competencies everyone needs concerning data management, in order to be able to deal with data in a proper way in daily life. Afterwards, we will discuss the impact of these changes in data management on computer science education and especially database education.

Keywords
Data Management, Key Competencies, Big Data, NoSQL, Databases, Data Privacy, Data Analysis, Challenges

INTRODUCTION
Nowadays, data take a key position in nearly everyone's daily life. Enormous amounts of data are generated and organized every day, for example when storing documents, music or photos, but also during suspicious activities, e.g. when using busses or trains (by using electronic tickets), when consulting a doctor (by using electronic health insurance cards), and so on. Therefore, dealing with data in daily life has many facets: data may be captured wittingly or unwittingly, it may be stored locally or in online (cloud) stores, it may have different structures, and so on. As these data are captured everywhere in daily life, large parts of the daily routine may be reconstructed by analyzing these data. For example, this is the case when taking together data coming from public transportation with the payments with credit cards and the data captured by a smart meter used in the private household. With these information, the daily routine like working hours, shopping habits but also which devices are used at home may be reconstructed by analyzing the times one drives by public transportation, where someone goes shopping and by analyzing the power consumption information captured by the smart meter (Molina-Markham, Shenoy, Fu, Cecchet, & Irwin, 2010). While this example deals with data collected about a person by third parties, in rarer cases people also collect data on themselves. For example, participants in the "Quantified Self" movement (also known as "life logging") actively acquire and analyze data on their daily life for different purposes, sometimes for improving health, for improving well-being, or for improving own productivity, in other cases just out of curiosity.
As nowadays data have a clear influence on daily life and also, as this influence will continuously increase in future, a central task in daily life is dealing with these data

in a responsible way, as storing, modifying, deleting and using data are typical aspects concerning everyone's life. In private life, this involves protecting own data and such about other persons from being manipulated, lost or from being used abusively. Nevertheless, using the innovations and newly emerging possibilities coming from modern data analysis is important for making the data that are captured in daily life valuable for personal use and for extracting new information out of known data. Nowadays, such possibilities are opened up for everyone because of various comfortable analysis tools available for free, but also because of the Open Data movement, which targets on publishing as many data sets as possible in order to allow many-sided usage.

In computer science, the aspects mentioned before are often summarized under the term "Big Data". This is a topical subject in various contexts, as Big Data not only affects daily life and computer science, but it has also strong impact on economy, politics and security. Typically, the term Big Data is described as dealing with large amounts of data with varying structures and high velocity, which includes continuously generating data but also fast processing of such in nearly real-time.

Various topical subjects that are being discussed in public dialogue are strongly affected by Big Data and especially by data analysis. For example, this is the case for early data retention or surveillance programs of intelligence agencies. These topics are often hard to understand, as data management and data analysis are complex topics that are becoming increasingly important, while the knowledge needed for understanding such topics is not part of current (computer science) education.

Additionally, various threats are emerging when being able to store and analyze large amounts of data with relatively small effect and computational costs. These threats especially affect data privacy, as by combining data sets from different sources highly private information may be gathered. Additionally, by involving statistical aspects into analysis, even information one never shared may be guessed with high confidence, and user profiles may be generated. This is especially harmful because nowadays everyone generates data always when using for example computers and smartphones—but also when driving a car, when using a smart electricity meter in a household, or when using other smart home innovations, as they are typically based on gathering and analyzing (big) data.

In the context of these current developments, the relevance of data management for people changes fundamentally: While until now, data management and especially databases were topics which were mainly considered as relevant for educational and professional use, these topics are now affecting the whole daily life of everyone. Despite this changing impact, current teaching considers databases as central topic of data management education (cf. e.g. Brinda, Puhlmann, & Schulte (2009), Seehorn et al. (2011)). In future, as other topics like data safety and data privacy will gain importance in daily life but also in education, the purpose of data management education changes tremendously. While the current focus of teaching is set on concepts of and knowledge on databases, in future this emphasis must be changed to fostering competencies needed in the daily life of everyone. Therefore, current database education must be revised and adapted to these new requirements in order to teach sustainable concepts and aspects of data management that are not only relevant in computer science and computer science education, but especially in daily life. So, for being able to deal with the new possibilities and threats, the new key competencies coming from data management must be fostered in class.

In order to point out these new and reappraised key competencies of data management, in this paper we will first describe main aspects of data management and Big Data that are relevant for teaching but also for pupils' daily life. On this basis, we will point out major key competencies, which people need for being able to deal with the new possibilities and threats evolving in context of data management. Finally, we will outline the consequences for computer science education by

discussing main challenges computer science education will have to deal with in future, when considering the described new aspects in data management education.

DATA MANAGEMENT IN THE CONTEXT OF BIG DATA

Dealing with data is an important task in daily life. This includes planning, organizing and utilizing data, methods which are typically summarized by the term "data management" (Bodendorf, 2005). In addition to these aspects, data management also comprises for example evaluating the quality of data, acquiring data, securing access to data as well as data backup and recovery (DAMA International, Mosley, Brackett, & Earley, 2009). Especially, these aspects are clearly concerning data management in daily life, as described before.

The topic data management changes tremendously in context of current developments like Big Data: Although data management is an important task in daily life for years, people often only considered it as topic in computer science (education), because the influences of data on daily life were only hardly recognizable. Nowadays this influence becomes increasingly obvious: While in 2012 approximately a total of 2.7 Zetabytes (10^{21} Byte) of data existed, about 2.5 Exabyte (10^{18} Byte) of additional data were generated per day (IBM, 2012). However, Big Data is not only characterized by the large amount of data and this high rate of data generation, but also by the variety of data: About 80% of these data are unstructured, as they are especially coming from social media. For example, in 2012 about 100 Terabytes of data were uploaded daily on Facebook and 230 Million tweets were sent on Twitter per day (IBM, 2012). Therefore, Big Data is often characterized by (at least) three Vs: "volume", "variety" and "velocity" (Laney, 2001), additional Vs like the "veracity" of data are added in some descriptions. When dealing with Big Data, new challenges are evolving for data management: usually, the most commonly used relational database management systems are optimized especially for consistent, durable and reliable storage of data. In contrast, newly emerging databases (often summarized under the term NoSQL[1] databases), set the focus on fast processing of distributed stored data and therefore accept limitations, such as in consistency. In addition, other aspects of data management, especially data safety and privacy are strongly affected by these new developments.

KEY COMPETENCIES IN DATA MANAGEMENT

With the increasing impact of data management, it is necessary to analyze which skills people need in order to deal with this topic in daily life. As data management is a complex demand with strong influence on different fields of daily life, skills necessary for successfully dealing with this topic are described as key competencies according to the definition by Rychen & Salganik (2001).

The key competencies needed for dealing with data and data management are especially coming from all parts of data management—including its main aspects structuring, organizing and utilizing data (Bodendorf, 2005)—but also from discussing the consequences of data management and its influence on data privacy. Therefore, in the following, we will point out the main key competencies considering the topic data management. These key competencies will be illustrated by examples describing their relevance for daily life.

Storing Data

Nowadays, everybody stores and manages enormous amounts of data every day, e.g. text files, videos, music, emails and so on. For storing data, various possibilities exist that strongly differ especially in the following aspects:

[1] Nowadays, the term NoSQL is interpreted as "Not Only SQL" (Edlich, n.d.), while originally it was used as name for a database management system not supporting SQL at all (Strozzi, 1998).

- storing data offline or in the cloud,
- storing data as files in a file system or as entries in a database,
- using specialized stores like media stores or not,
- storing data in a structured or unstructured way.

In most cases, data are stored as files in file systems, locally or in the cloud, especially when dealing with documents, photos, music and so on. In contrast, there are also data stores that are specialized on only a few data types. For example, email programs are data stores for emails but also for contacts and in some cases calendar entries, while music could instead be stored in media libraries, often together with videos and pictures. By using an application specialized for such concrete use cases, dealing with data will be distinctly simplified. As each time when storing data the requirements strongly differ, it is not possible to decide in general if to use specialized stores or not, as these stores even have disadvantages: for example, they often use proprietary file formats and therefore comprise the threat of a "vendor lock-in"[2]. Also, all other aspects mentioned above must be decided from case to case. These decisions are summoned by the question: *"Which data should be stored—and where and how?"* Therefore, dealing with data implicitly involves knowledge on the different possibilities for storing data.
The decision which data store with which functionalities should be used in the concrete use case must be made as the case arises. So, for being able to deal with data, and especially large amounts of such, in a proper way, it is necessary that students *"understand and apply different ways for storing data"*.

Dealing with Meta-Data
All these large amounts of data that are generated every day bring additional information with them, the so-called meta-data. These meta-data are for example visible as attributes of a file (e.g. creation as well as last-modified date, author/creator…), as log files (e.g. in cloud storage services: "file 'example.txt' was deleted on 2014-02-01 09:07 by 'user1'") or as tracked changes in a document. Although most meta-data are actually visible to the user, they are often disregarded: While final versions of a document are typically cleaned from e.g. comments, often meta-data are not revised. These data may contain information not supposed to be contained in the final, or they may be generated from old content, like old or wrong keywords. In addition, information that may be confidential, like the concrete author of a document or the time span between creation and last modification is included usually. There are various examples for cases, where confidential data was disclosed by meta-data included in published documents. For example, in 2005, the United Nations published a report on Syria's involvement in a murder; this document not only contained the visible information, but also tracked changes. These annotations were only hidden and contained names of persons involved in this plot that were not supposed to be disclosed (Zeller, 2005). Therefore, when dealing with data in daily life, people need the key competency to "*note that additional data may be included in data sets as meta-data*".
On the other side, these meta-data also facilitate dealing with data: by adding additional information as meta-data, locating data is simplified and accelerated. Such meta-data are especially necessary when searching for information by substantial criteria, as typical search engines cannot interpret most file contents directly, so searching by content might only be possible for pure text files. For example, when dealing with photos, adding meta-data, which for example describe the place the photo was taken at, or by marking persons who are visible on it,

[2] The term "vendor lock-in" describes the dependence on a single vendor of products. This is for example the case, when data are stored in a proprietary file format, so that using them in another vendor's product is hardly possible.

locating this photo afterwards will obviously be simplified in comparison to searching without such information. As meta-data are typically considered by search engines of operating systems, but also within most of the currently used database management systems, being able to deal with meta-data simplifies daily data management a lot. When dealing with such information, also the disadvantages of using them should be kept in mind: not only the effort of assigning and maintaining such meta-data may be relevant for the decision if such information should be added, because even the usefulness of such information strongly depends on the concrete use case. Additionally, while the existence of meta-data strongly benefits reading data from the data store, typically writing operations are slowed down, because not only the data but also the meta-data must be updated in order to ensure consistency. Therefore, another key competency in data management is to *"understand the purpose of meta-data and use them in a proper way".*

Dealing with Redundancy and Consistency

When structuring data, people will always have to deal with redundancies and inconsistencies: for example, it often seems reasonable to store a copy of a file in two folders of the file system, when a file concerns two topics managed in different folders. This way of storing data leads to inconsistent data if one file is being updated while the other one is (accidentally) left untouched. Therefore, one eventual consequence of storing data redundantly is the emergence of inconsistencies between multiple copies of the data. Since this problem, needing a duplicate copy of a file in another location/folder, is not unusual when saving data, students need to understand the consequences of storing data redundantly. In addition, they have to deal with this requirement, such as by linking the data at the second location instead of saving a real copy of it. Therefore, "*understand the consequences of storing data in a redundant way*" as well as *"save data in a proper way in order to prevent inconsistencies"* are key competencies of data management.

Today, inconsistencies are also often caused by synchronizing data between multiple locations. Nowadays one person carries in average 2.9 (mobile) devices including laptops, smartphones and tablets (Truong, 2013), and the overall number of devices used by one person may be even higher. Therefore, data are often synchronized between two or more devices, and not only read but also modified on these. This leads to inconsistencies when modifying data that was earlier changed in another location, but not successfully synchronized to the other devices yet. This leads to different possible consequences, dependent on the application used for synchronization and on the type of synchronized data: while only in special cases (such as pure text files) such conflicts may be automatically resolved, in most cases duplicate data will come up or in the worst case data will be lost. So, another key competency in this topic, which is needed in order to be able to understand threats when synchronizing data, is *"understand the consequences of synchronizing data and deal with synchronization conflicts"*.

While commonly redundancies and inconsistencies should be avoided, there are also use cases where both concepts are used intentionally: for example, backups are exactly such redundant copies of the original file and will become inconsistent as soon as the original file is modified again. However, in this case redundancy occurs by design, because backups serve as fallback copies, especially for the case that data are accidentally deleted or changes must be reverted. Therefore, they need to be redundant to the original file (for restoring) and need to become inconsistent when the original file is being modified (for reversing changes). Today, as in most operating systems different backup functionalities exist, people must also be aware of the different ways for creating backups: continuous backup vs. backup at discrete points in time, incremental vs. complete backup, hot vs. cold backup. These possibilities clearly differ in used hard disk space, in the speed of the backup and restore processes, and in the typical frequency of backups. For each use case, it is

necessary to decide, which aspects are required—a decision which must be done in context of the value of the concrete data. Therefore, an additional key competency in this field is *“create redundant data sets for backup / data safety in a proper way”*.

Data Safety and Encryption

Nowadays, data are especially covering great parts of daily life—as with smartphones and other mobile devices, an increasing amount of moments is immediately captured as data, for example in form of posts in social media, photos, but also as data captured in background, like position data, probably log files of sensors and so on. Often, the data everyone manages and generates daily, are not only stored on stationary desktop PCs, but also on mobile devices as well as portable USB drives without additional security measures, and they are also often transferred via insecure communication channels. This results in privacy issues, but also enables even more problematic threats like identity theft—or when thinking on professional use, financial losses may occur. Storing data on such mobile devices or storage media enlarges the risks of unauthorized usage of these data, of manipulations and of data theft. This risk has harmful impact on daily life, but it becomes even more relevant when concerning eventual losses of data in vocational context. Therefore, it is an important task to store and transfer private or confidential data in a secure way. This may be reached by different alternative ways. Especially a typical method relevant for securing data in case of theft of the device, on which the data are stored, is to restrict access to these devices. This is especially done via password protection or similar authentication methods. Although this will increase data safety, as accessing data becomes more difficult, data safety cannot be guaranteed by this method, because data are yet accessibly stored on the device that enforces the authentication. A simple approach for surrounding such authentication methods is reading the data store: For example, the hard drive, using another device that does not enforce this authentication. Users must be aware, that usually typical authentication methods cannot suffice to secure their data, as they are only a hurdle for accessing these. Therefore, people must differentiate between restricting the access to devices and to the data itself. To (relative strictly) ensure data safety, it must be prevented that the meaning of data is recognizable without the required permissions. This is the goal of encrypting data: While encrypted data might still be read from the hard disk, they have no value for anyone until being decrypted using the right key (or enormous computational costs). So, another key competency when managing large amounts of data is to *“understand the difference between restricting access to a device or service and protecting the data stored on it*” as well as to *“encrypt data and communication”* in order to prevent prohibited access to these large amounts of data.

Another aspect concerning data safety is to decide whether to confide specific data or not. For example, the author attribute of files, emails and so on is typically not protected against changes nor is the content itself. Therefore, such data carry the risk of being manipulated. As in various use cases it is necessary to be able to trust data, everyone must be aware of methods for checking the genuineness of data. For example when reading emails, nowadays most people keep in mind that these messages may contain non-genuine content, as they may be sort of junk or phishing mails. However, with an increasing quality of such messages, it will be increasingly hard to figure out if an email is genuine or not. Especially, other data than email are often less questioned, because threats are often less obvious and less discussed in public dialogue. Therefore, methods for proving the genuineness of data will become increasingly relevant in future, especially because an increasing amount of legally relevant tasks is done via electronic communication methods. One technique for guaranteeing the authenticity of the sender information as well as the validity of the content is by adding a digital signature to the data. This enables the recipient to

check if data were manipulated. Therefore, it is necessary that people "*know methods for guaranteeing the authenticity of data and use them in a proper way*".

Using Methods of Data Analysis

Today, various sets of information and data are available for free, but in most cases only few people are able to use these data in another way than by only viewing them. For example, by combining data from various sources, interesting new use cases may be found as well as new information may be extracted from these data. This is possible for everyone nowadays, even if only few people use this chance: different simple tools for analyzing data are provided for free by the large Big Data companies like Google or IBM. In addition, there are simple tools for creating mashups, a form of integrating multiple data and especially media. An example for using such open data sets is evaluating whether to book a hotel in a concrete borough in another way than reading the opinions of former visitors. As for example the City of New York offers many data they capture daily as open data sets[3], they also publish calls to the service number 311, which include complaints on noise, street or sidewalk conditions and so on[4]. While the data direct result when analyzing these data are relatively obvious, they can also be combined with other data, such as restaurant inspection results[5] in order to gain prediction factors, for example if the noise conditions in a borough and the ratings of the restaurants in this part of a town correlate or not. Doing such data analysis is possible with simple techniques, for example included in spreadsheet applications or available as online tools. Therefore, typical data analysis methods will be used, especially grouping data ("clustering", in this case by borough), categorizing them by different characteristics ("classification", in this case e.g. the types of service calls but also the restaurant grades) or by determining interdependencies (if-then-relations) between data ("association", for example the described analysis if the restaurant grades and noise conditions correlate). So, another key competency in data management is *"use, find and combine data in order to gather new information"*.

Additionally, by analyzing data themselves, people will be enabled to recognize the threats for data privacy coming from data analysis. With the ability to combine data from different sources, it is only a small step into discovering that the same methods may be used when analyzing personal data. Therefore, even data strongly anonymized or pseudonymized according to data privacy acts may be deanonymized—so data privacy acts would be bypassed. This was for example the case, when AOL released a set of search data, which included a unique person ID, which was related to a person, but without revealing personal data of this person, as well as the user's search terms. By analyzing these data, some data analysts could rapidly recognize some persons out of these data, so they were able to find their real name as well as contact data together with their search habits at AOL's search engine (Barbaro & and Zeller Jr., 2006).

Therefore, another key competency of data management, which involves not only this topic data analysis, but also data privacy, is *"know the threats for data privacy and analyze data with keeping ethical demands in mind"*.

Being Aware of Data Traces and Data Privacy Issues

With the possibility to store and analyze huge amounts of data, different threats for data privacy are evolving. As mentioned before, meta-data may be harmful if the user does not know about them, or when dealing with them in an inappropriate way.

---

[3] NYC Open Data: https://data.cityofnewyork.us

[4] Analyzation of New York City 311 Service requests: http://opendatabits.com/new-york-city-311-service-requests-open-data

[5] Meshup of NYC 311 calls together with restaurant inspection results: http://opendatabits.com/nyc-restaurant-inspections-results-open-data

In addition, a lack of data safety and encryption strongly affect data privacy. This threat is even intensified when dealing with modern devices, applications and services, as various types of data on this usage and on the user are captured continuously. While the main aim of some services is capturing data in a relatively obvious way, such as in social networks, in other cases they are generated in a hidden way besides the intended use, for example as log files. Additionally, applications supposed to generate data, like the mentioned social networks, tend to store more data than actually needed for the service to work. Therefore, while in some cases the user is aware of this data generation and actively decided for capturing these data, such as when participating in the "Quantified Self" movement, this is usually not the case. But by combining different sources of such data, large parts of daily life may be reconstructed. As nowadays, everyone uses different kinds of data stores, but also applications using these data, one leaves digital traces everywhere.
Therefore, the question if one trusts an application/service or not becomes increasingly important for data privacy. Another example confirming typical chat applications that are offering the possibility to display "last online" times to "friends": by having a look on these times (and perhaps comparing them to the ones of other persons) other people can digitally trace persons with simplest methods. Depending on the concrete application, this tracking is even possible without prior contact to a person, only by adding them on the contact list without a need for approval of this request by the person added.
So, raising pupils' awareness on such abilities and threats is an important aspect when discussing data privacy. As such methods for collecting data are typically hard to discover and in most cases cannot be prevented, users must be aware of these possibilities in order to be able to recognize hints on such issues, for example the "last online" times in chat applications mentioned before.
Therefore these aspects of data privacy are summoned by the questions "*Who stores which data on me? Who has which information on me? Who can I confide data about me?*" So, "*note own data traces*" but also "*know the possible threats of using data storage services*" are important key competencies, which are hard to foster, because tracking such traces is mostly done in an invisible way, as well as threats when using data storage services are typically hard to discover.

Overview
As described before, several key competencies in data management could be identified. These will be summarized in order to get a complete overview:

People…
- understand and apply different ways for storing data
- note that additional data may be included in data sets as meta-data
- understand the purpose of meta-data and use them in a proper way
- understand the consequences of storing data in a redundant way
- save data in a proper way in order to prevent inconsistencies
- understand the consequences of synchronizing data and deal with synchronization conflicts
- create redundant data sets for backup / data safety in a proper way
- understand the difference between restricting access to a device or service and protecting the data stored on it
- encrypt data and communication
- know methods for guaranteeing the authenticity of data and use them in a proper way
- use, find and combine data in order to gather new information

- know the threats for data privacy and analyze data with keeping ethical demands in mind
- note own data traces
- know the possible threats of using data storage services

It is clearly visible, that these competencies face different aspects of data management. But at the same time, they are all strongly related to daily life. Additionally, most of these key competencies have one central aspect in common: they face newly occurring decisions, which are necessary in order to deal with data management in a proper way. This mirrors the current developments in computer science: while until the last years, mainly one database system was leading and used for most use cases, since the development of the NoSQL databases, a great variety of such systems evolved, what makes it necessary to decide for a concrete database according to the use case.
Therefore, also overall key competencies, like judgment / decision-making as well as analytic thinking, are clearly fostered by these competencies in data management.

CHALLENGES FOR COMPUTER SCIENCE EDUCATION

In contrast to their relevance for daily life, these key competencies do not yet receive sufficient attention in current data management education. With the increasing relevance of data management, current curricula for computer science education must be revised with keeping the new requirements and possibilities in mind. By considering these aspects in class, computer science education may change tremendously: Especially, while current data management education mainly focuses on databases, in future the relevance of various additional topics will increase clearly, while other current topics may then be less important. Therefore, the key competencies developed above, need to be considered in computer science education, since no other subjects in general educational schools can foster these, because this requires more than basic knowledge on these topics.
In contrast to these new requirements, current computer science education mainly focuses on relational database management systems when talking about data management. Since the topic databases was intensively discussed in the context of computer science education in between the years 1986–1998, only occasionally papers and articles on this topic were published. While in the earlier of these years the relevance of (relational) databases in class was the main topic of research, in the later of these years and now on, the focus of publications is set on supporting the teaching of databases. Therefore, various tools were discussed, especially for teaching SQL, e.g. by Grillenberger & Brinda (2012) and by Sadiq et. al. (2004). Since 1998, only few research results on this topic were published, especially there are no publications concerning current developments like Big Data or the increasing relevance of data management in daily life, yet. In addition, currently no compilation of key competencies concerning this field exists. Only different educational standards, like the K-12 Computer Science Standards by the Computer Science Teachers Association (Seehorn et al., 2011) or the German Educational Standards for Computer Science in Lower Secondary Education (Brinda, Puhlmann, & Schulte, 2009), mention some competencies on this topic.
In the following, we will outline some of these competencies for comparing them with the key competencies in data management described above. By having a look on the educational standards, important competencies on data management / database education are especially found in the topics “structuring data”, “data safety” and “data privacy”. The former especially includes aspects of creating and using data structures, e.g. students “know principles for structuring documents and use them in

an appropriate way"[6], they "know and use tree structures, for example directory trees"[6] or they "navigate in directory trees and manipulate directory trees in a proper way" [6] (Puhlmann et al., 2008). The latter ones—"data safety" and "data privacy"—need to be distinguished, even though they are strongly related to each other. Data safety focuses on the technical aspects, like preventing prohibited access to confidential data, encrypting such data and so on, while data privacy focuses on using data (and especially personal data) in a proper way. So, competencies needed concerning data safety are for example "*explain the principles of security by examining encryption, cryptography, and authentication techniques.*" (Seehorn et al., 2011), while a typical competency concerning data privacy is "*evaluate situations in which private data should be shared*"[6] (Puhlmann et al., 2008). Also, an important competency considering data management, mentioned by Seehorn et. al. (2011), is "*use data analysis to enhance understanding of complex natural and human systems.*".

By comparing these competencies currently considered as important in computer science education with the ones described before, a clear difference is visible: Although most competencies of current database education will remain relevant in future, various additional ones are supplemented. In addition, especially the current competencies are strongly related to computer science, while future key competencies of data management especially face dealing with data in daily life.

In addition, the topic databases will change clearly in context of Big Data, and especially in context of the newly emerging NoSQL databases. In order to meet the main requirements of storing Big Data, the management of large amounts of data with high variety and high velocity, new types of databases arose. These non-relational databases are typically summarized under the term NoSQL[7]. Various concepts of databases that were so far assumed as being fundamental to this topic are dropped by these databases, in order to speed up access to distributed stored data. For example, while consistency is a main requirement of relational databases, as it is part of the ACID[8] criteria, this concept is dropped in NoSQL databases, because they are only "eventually consistent" according to their main requirements summarized as BASE[9]. Therefore, in order to teach sustainable concepts and aspects of data management and databases, the concepts fundamental to databases—and not only for relational database management systems or for NoSQL databases—need to be analyzed.

CONCLUSIONS

As described in this paper, by considering the new aspects coming from current developments like Big Data and because of the increasing importance of data management for daily life, data management education changes tremendously. By discussing the new aspects coming from these topics in class, various key competencies of data management will be fostered. Although, these new key competencies are becoming increasingly relevant in daily life, they are mostly not yet fostered in current education on the topics data management or databases. Especially aspects coming from data privacy and data analysis will increase in importance in future data management education, but also all the other key competencies described before need to be fostered, as key competencies are "of

[6] Original in German, translation by authors

[7] NoSQL nowadays is interpreted as "not only SQL" (Fowler & Sadalage, 2012). In original, by Carlo Strozzi (1998), this term was used as name for a database not supporting SQL.

[8] ACID is the abbreviation for Atomicity, Consistency, Isolation and Durability, the four main requirements on relational databases (Elmasri & Navathe, 2011).

[9] BASE is the abbreviation for Basically Available, Soft-State, Eventually Consistent, the three main requirements on NoSQL databases (Edlich, n.d.).

prime importance for a successful life and effective participation in different fields of life" (Rychen & Salganik, 2001).

This will especially ensure a better fit between education and daily life, because today most people use applications involving Big Data analysis multiple times daily, but without being able to notice the collection of their data or its analysis—and often without even knowing about possible consequences. Additionally, as everyone deals with and generates large amounts of data continuously, also the importance of data management in daily life increases continuously: such as when storing data in different data stores (like the file system, media libraries and so on), when synchronizing data between multiple applications and/or devices or when creating backups of data.

In addition, Big Data has strong impact on current and newly emerging professions. Especially, various professions are changing when considering aspects of Big Data as well as of data analysis. Also, at the moment new professions are evolving, like the data scientist (Davenport, Patil, & others, 2012). In this profession, aspects coming from informatics, especially data analysis, are combined with mathematical ones, especially coming from statistics. Therefore, knowledge on fundamental concepts and methods of dealing with Big Data will have sustainable influence.

REFERENCES

All electronic sources were retrieved on 1th March 2014.

Barbaro, M., & and Zeller Jr., T. (2006, August). A Face Is Exposed for AOL Searcher No. 4417749.

Bodendorf, F. (2005). *Daten- und Wissensmanagement [Data and Knowledge Management]*. Springer.

Brinda, T., Puhlmann, H., & Schulte, C. (2009). 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* (pp. 288–292). New York, NY, USA: ACM. doi:10.1145/1562877.1562965

DAMA International, Mosley, M., Brackett, M. H., & Earley, S. (2009). *The Dama Guide to the Data Management Body of Knowledge Enterprise Server Edition*. Technics Publications Llc.

Davenport, T. H., Patil, D. J., & others. (2012). Data scientist: the sexiest job of the 21st century. *Harvard Business Review*, *90*(10), 70–77.

Edlich, S. (n.d.). NOSQL Databases. Retrieved from http://nosql-database.org/

Elmasri, R. A., & Navathe, S. B. (2011). *Fundamentals of Database Systems*. ADDISON WESLEY Publishing Company Incorporated.

Fowler, M., & Sadalage, P. J. (2012). *NoSQL Distilled - A Brief Guide to the Emerging World of Polyglot Persistence* (1. ed.). Amsterdam: Addison-Wesley.

Grillenberger, A., & Brinda, T. (2012). 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* (pp. 101–104). New York, NY, USA: ACM. doi:10.1145/2481449.2481474

IBM. (2012). The Flood of Big Data. *Infographic*. Retrieved from http://ibmdatamag.com/2012/04/managing-the-big-flood-of-big-data-in-digital-marketing/

Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity, and Variety*. Retrieved from http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Molina-Markham, A., Shenoy, P., Fu, K., Cecchet, E., & Irwin, D. (2010). Private Memoirs of a Smart Meter. In *Proceedings of the 2Nd ACM Workshop on*

Embedded Sensing Systems for Energy-Efficiency in Building (pp. 61–66). New York, NY, USA: ACM. doi:10.1145/1878431.1878446
Puhlmann, H., Brinda, T., Fothe, M., Friedrich, S., Koerber, B., Röhner, G., & Schulte, C. (2008). 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 for computer science lower secondary]. *Supplement to LOG IN*, *150/151*.
Rychen, D. S., & Salganik, L. H. E. (2001). *Defining and selecting key competencies.* Hogrefe & Huber Publishers.
Sadiq, S., Orlowska, M., Sadiq, W., & Lin, J. (2004). SQLator: An Online SQL Learning Workbench. In *Proceedings of the 9th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education* (pp. 223–227). New York, NY, USA: ACM. doi:10.1145/1007996.1008055
Seehorn, D., Carey, S., Fuschetto, B., Lee, I., Moix, D., O'Grady-Cunniff, D., … Verno, A. (2011). *K–12 Computer Science Standards*. Computer Science Teachers Association, Association for Computing Machinery.
Strozzi, C. (1998). NoSQL: a non-SQL RDBMS. Retrieved from http://www.strozzi.it/cgi-bin/CSA/tw7/I/en_US/nosql/Home Page
Truong, K. (2013). INFOGRAPHIC: Users weighed down by multiple gadgets - survey reveals the most carried devices. *Sophos Naked Security*. Retrieved from http://nakedsecurity.sophos.com/2013/03/14/devices-wozniak-infographic/
Zeller, T. J. (2005, November 7). Beware Your Trail of Digital Fingerprints. *The New York Times*. Retrieved from http://www.nytimes.com/2005/11/07/business/07link.html

Biography

Photo **Name** is…