PERSPECTIVES FOR AN INTERDISCIPLINARY DATA SCIENCE CURRICULUM IN GERMAN SECONDARY SCHOOLS

Rolf Biehler & Carsten Schulte
Universität Paderborn, Germany
biehler@math.upb.de, carsten.schulte@uni-paderborn.de

We present some initial guidelines and ideas for an interdisciplinary data science curriculum in German secondary schools, based on a brief discussion of educational philosophy, as well as thematically relevant approaches and traditions in teaching and learning mathematics and computer science.

INTRODUCTION

Data science and its associated buzzwords, for example *big data*, are more and more seen as relevant for education, but so far, few attempts have been made to introduce the field at the school level. Data science is not yet part of the curriculum in secondary schools in Germany and most other countries.

In this introductory chapter, we discuss the goals of the symposium to inform future curriculum development, and we analyze experiences and curricular traditions from the two main subjects, computer science education and statistics education, on which we can build and that we relate to each other. Moreover, we will discuss principles of curriculum development on which we will base our approach.

We also present a first look into our current idea of a future data science curriculum for secondary schools. As a step towards helping teachers and schools implement aspects of data science, we plan a collaboration with selected schools and teachers to implement and try out some ideas of data science in actual secondary classrooms. We subsequently plan to implement a year-long course of 3 hours per week for students who volunteer to take the course, which will in turn become an elective duty course for them that counts for the final examination. We will test ideas from data science with students, establish relations with teachers who are co-designers, and develop material for classroom teaching and professional development courses. Based on that work, we will work on a position paper describing essential components of data science for secondary students.

CHALLENGES FOR DESIGNING A DATA SCIENCE CURRICULUM FOR SCHOOLS

The first goal of the symposium is to exchange ideas, material, and experiences of ongoing projects on data science at school level, at tertiary level, and in internal training programs in companies. As a second goal, an understanding of “fundamental ideas” of data science should emerge: views of data science as a scientific discipline including its relation to statistics and computer science and its historical development, current state, and future perspectives. The notion of fundamental ideas has successfully been used in statistics education to orient curricular developments and teacher actions in the classroom (Biehler 2014a, 2014b; Burrill & Biehler 2011). Thirdly, we want to identify relevant and typical applications of data science and uses of big data in economy, industry, and society; consider ethical, political, legal, and social responsibility aspects of these applications; and reflect on their educational relevance and potential for being made accessible to school teaching. We view our future students in two different roles: working as data scientists themselves or exploring existent systems with the aim of developing simple system models; and also making the black boxes more transparent and identifying underlying assumptions, including economic, political, and cultural conditions and interests. Last but not least, societal and cultural conditions and implications have to be discussed. These aspects are already being analyzed in media education research, in digital humanities, in socio-informatics, and in many popular books (Aoun & ProQuest (Firm) 2017; Harari 2017; O´Neil 2016; Spitz 2017; Weigend 2017). For a data science curriculum, the question is how to educate students so that they can take a thoughtful position in these debates, and on a more practical level, how to integrate societal issues with formal and technical aspects of data science as scientific discipline. The societal aspects include questions about providing private data to companies, critical media competence, “News competence” (dealing with “fake

In Biehler, R., Budde, L., Frischemeier, D., Heinemann, B., Podworny, S., Schulte, C., & Wassong, T. (Eds.) (2018):
Paderborn Symposium on Data Science Education at School Level 2017: The Collected Extended Abstracts.
Paderborn: Universitätsbibliothek Paderborn. http://doi.org/10.17619/UNIPB/1-374

news”), statistical literacy (media reports including products from data journalism and scientific studies using data), and using data and/or data science for one’s own goals and in everyday situations.

In sum, various aspects have to be integrated into an overall educational philosophy that comprises all the areas we mentioned; see Figure 1.

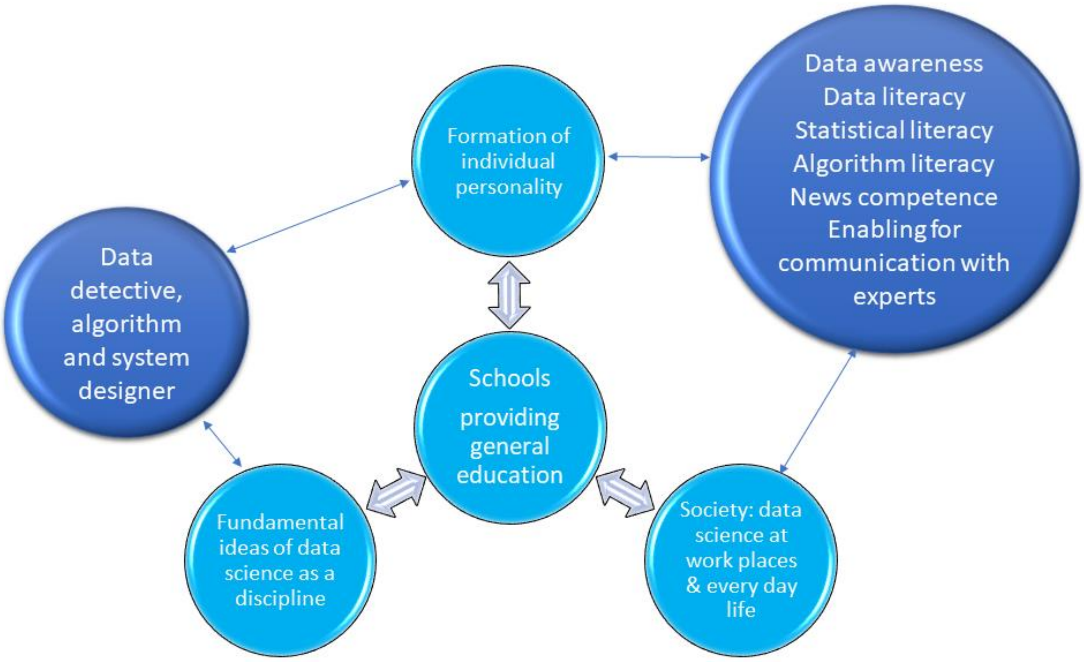

Figure 1: Facets of an educational philosophy for data science

When thinking about curricular reform for school, the question of the German tradition of *Allgemeinbildung* e.g., as expressed by W. Klafki (1996) emerges: Why teach, for whom, with what goals? The underlying so-called “rationale” of the curriculum addresses these questions.

In general, the overarching goal should adhere to self-determination, responsible actions, developing interests, and being introduced to basic ideas of the discipline. Within the context of our data science curriculum development project, we agreed upon four basic guidelines for the curriculum:

1. Develop practical educational resources
2. Figure out and teach fundamental ideas of data science
3. Ensure practical relevancy for everyday life by
    a. identifying relevant application areas
    b. reflecting whether this is of educational value for students
4. Integrate societal and cultural aspects of data science

FRAMEWORK FOR CURRICULUM DEVELOPMENT

Developing a curriculum can be difficult, because a variety of levels, aspects, and people have to be involved.

It can be surprisingly challenging to define the term *curriculum*. Thijs & van den Akker (2009) suggest a broad definition as a “plan for teaching” that can be observed or represented in different levels for various stakeholders. Usually, as in this project, the curriculum is presented as a written document describing an idealized plan for teaching. This plan serves teachers and schools as a point of reference to implement data science in their local classrooms, and probably to derive their own local school-wide curriculum model.

As a plan for teaching, a curriculum model describes several aspects, e.g.: the goals of the teaching, the content, some teaching methods, maybe some specific examples and materials, guidelines for assessment, and so forth. In the curriculum model from the SLO (Netherlands Institute for Curriculum Development), these different aspects are presented as a curricular spider web (Figure 2).

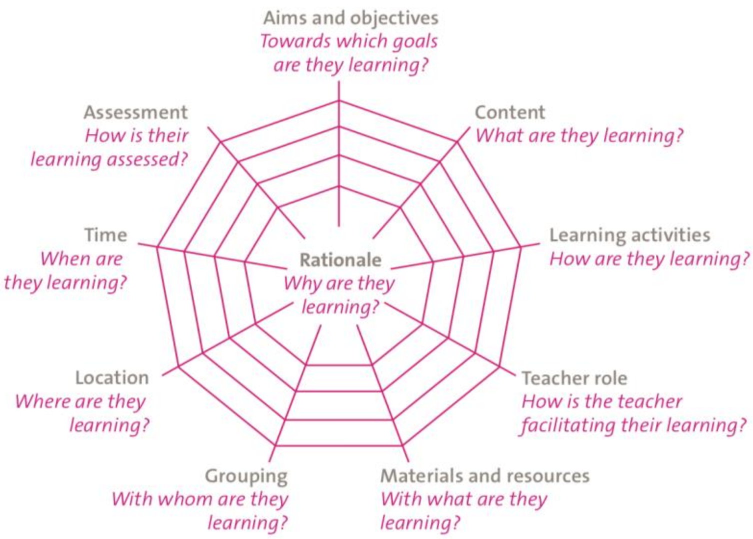

Figure 2: The curricular spider web, taken from Thijs and van den Akker (2009, p. 59)

The figure presents the different dimensions of a curriculum as a spider web to highlight the notion of interdependent and mutually connected aspects that have to be coherent in order to form a suitable curriculum model—or the web will rip apart. Secondly, the graphical presentation highlights the need for an underlying rationale: a philosophy and maybe implicit understanding behind the dimensions that ensures such coherence.

On the level of a teacher, this rationale can also be seen as shared understanding or belief in the nature of the discipline, the core aspects and goals of the subject. The data science symposium and papers in this publication can be interpreted as an attempt to develop such a shared understanding, by inviting experts from different subjects and contexts, and to watch for commonalities, especially in the implicit understanding of the "nature of data science." We thus included experts as observers who presented their view on shared themes as well as differences in the final panel.

The presentations in the symposium focused on perspectives on data science from the academy from business, and from international schools. In this paper, we highlight some important aspects of the curriculum based on curricular traditions in German schools for the two closest subjects to data science, namely statistics education and computer science education.

DATA SCIENCE EDUCATION FROM THE PERSPECTIVE OF STATISTICS EDUCATION

We see the following dimensions for rethinking important impacts of statistics education

- Work flow: Updating the PPDAC cycle
- Extending the statistical view of "data"
- Taking into account Extended and new methods for data science
- Selecting digital tools for data science that support data analysis, data management, and algorithm design
- Taking into account important insights from the statistical literacy discussion

A recent paper that discusses consequences of the data revolution to statistics education is Ridgway (2015).

Work flow

Many statistics educators base their view on the process of statistical inquiry on the so-called PPDAC cycle (Figure 3).

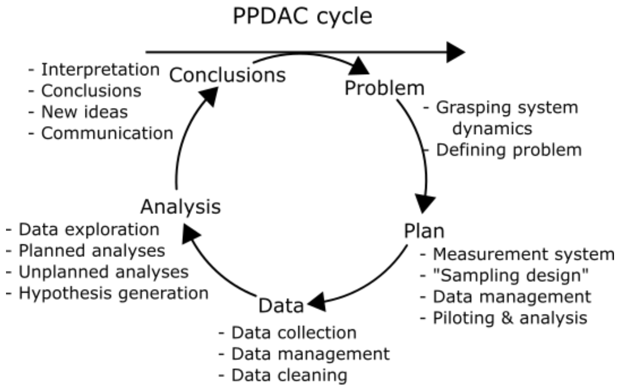

Figure 3: PPDAC cycle according to Wild and Pfannkuch (1999, p. 226)

In books on data science different work flow diagrams are used. We quote from Berthold, Borgelt, Höppner, and Klawonn (2010).

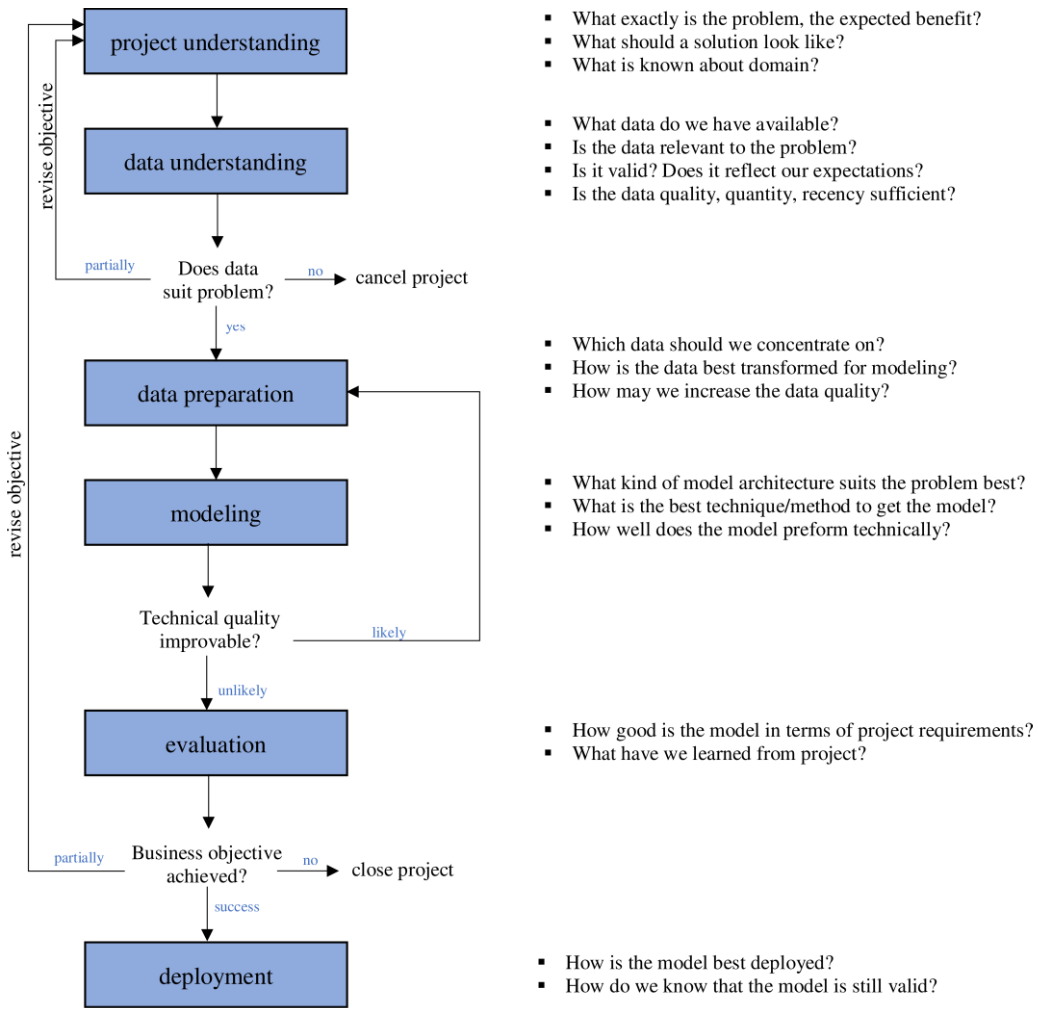

Figure 4: Data analysis cycle according to Berthold et al. (2010, p.9)

We notice several differences between the two cycles; some of them are important for future data science processes.

- Data may be “already there” and not collected according to a plan; the starting point of the cycle may be therefore different (Huber 2011; Tukey 1962)
- Data preparation and data cleaning are regarded as much more important than in statistics
- Project and data understanding are emphasized as part of the “problem” step
- Modeling as a step should be added to the PPDAC cycle
    - Classical statistical theory assumes the model to be given
    - Data science uses new types of algorithmic models (Breiman 2001)
    - Validation of the model is missing as a step (cross validation; distinguishing data for training and for testing is essential)
    - Prediction as a goal for modeling has to be emphasized
- “Conclusions” as a final process step has to be extended
    - Statistics aims at “knowledge”; but data science and computer science “deploy” models. This includes social responsibility, an important step

Extending the statistical view of “data”

The following aspects will be new given the current minor role of data in the school curriculum, where univariate numeric data dominate the statistics curriculum, and bi- and multivariate data are rarely curricular topics.

- Standard in statistics but not in school:
    - The rectangular data tables with different variable types
    - Data of moderate size, multivariate data
- New types of data for statistics
    - Data collected by sensors
    - Data collected by personal devices
    - Transactional data (traffic, supermarket buys)
    - Images and texts
    - Data scraped from webpages
    - Data with geographic information
- Big data; open data

Traditional statistics education often focusses on data with high quality, stemming from controlled randomized experiments or random samples of a well-defined population. Exploratory data analysis in the tradition of John Tukey has always been open to “dirty data,” while remaining aware that the kind of conclusion one can draw depends on the quality of the data and a deep knowledge of meta-data. This problem is exacerbated by the many available open and big data sets on the internet, whose origin and quality is often unclear.

Extended and new methods for data science

From the perspective of statistics education, methods such as machine learning, algorithmic models, decision trees, and clustering are new to the curriculum and not yet accessible to school students. We have to cope with the situation that sometimes methods known in statistics get a different name, such as *regression*, which has become one of the methods of *supervised learning*. However, this is not only a new name but a different perspective, with different uses and different generalizations of this traditional statistical method. It may also be the case that new methods can only be introduced in school on the basis of old methods in order to secure better understanding. For instance, one may have to start with bivariate linear regression as a starting point for more complex multivariate and non-linear methods. Teaching bivariate methods from a data science perspective could mean:

- Model fitting with different function classes

- Discussing algorithms for fitting, not treating them as black boxes
- Dealing with model selection, overfitting, and cross validation
- Using different “score functions” (not only least squares)
- Emphasizing validation (residual analysis)
- Using nonlinear regression (smoothing)
- Being aware of different study goals: Explanation or accurate prediction

Selecting digital tools for data science

Currently, in German schools, not much technology is used to support probability and statistics education. If we see technology at all, we see uses of spreadsheets, Geogebra, and graphic calculators, but no statistics tools. Only in experimental classrooms, tools especially designed for supporting the learning and doing statistics and probability are used, such as Fathom and Tinkerplots (https://www.stochastik-interaktiv.de, https://www.tinkerplots.com, https://fathom.concord.org), and building on Tinkerplots and Fathom, the data exploration environment CODAP (http://codap.concord.org, see also the contribution of Bill Finzer to this volume). The question of requirements for digital tools in statistics education (Biehler 1997; Biehler, Ben-Zvi, Bakker, & Makar 2013) has but recently broadened the view by including requirements from data science (McNamara 2015). Whereas the data science at school project led by Rob Gould (see his contribution to this volume) has decided to use R with an adapted set of commands, McNamara also includes Jupyter Notebooks in her review, which can be used as an environment for Python, making it an advanced, relatively easy-to-enter-into programming environment for computer scientists. A growing number of books covering data science with Python have been published (Grus 2016; Haslwanter 2016; Igual & Seguí 2017; McKinney 2017). It is an open question which tools can be adequately introduced at what age level, and whether one should aim at one single tool or use an “entrance tool” for easy data exploration such as CODAP for easy data exploration and then move on to more advanced tools such as R and Python. The latter would, in any case, require the compilation of a student-oriented library of commands, algorithms, and programs. Using Jupyter notebooks (Toomey 2017) may be supportive of this endeavor, but at its root, a curriculum has to incorporate strategies to support an adequate “instrumental genesis” for this these tools for working on data science problems (Guin & Trouche 1999; Madden 2013).

Insights from the statistical literacy discussion

Last but not least, designing a data science curriculum could profit from the lessons learned from the statistical literacy discussion, which takes cultural and societal aspects into account as well as education for critical thinking (Gal 2002, 2003).

We see the following facets:

- Problems of measurement (operationalization of variables, adequacy problem)
- Biases in sampling
- Distinguishing observational studies from experimental studies
- Random assignment and the problem of confounding variables
- Simpson’s paradox; ecological fallacy
- Confounding of conditional probabilities
- Understanding visualizations of complex data (including interactive ones)

Various new perspectives that already take data science aspects into account have recently been published (Gould 2017; Grant 2017; Prodromou & Dunne 2017; Schield 2017; Sutherland & Ridgway 2017).

DATA SCIENCE EDUCATION FROM THE PERSPECTIVE OF COMPUTER SCIENCE EDUCATION

In this section we focus on the following aspects, which we will discuss in the next paragraphs:

1. Updated view of data
2. A model of the data science process
3. Integrating societal aspects
4. Tools and resources

Updated view of data

In computer science at school a certain understanding (probably implicit) of data is already taught—and probably needs to be updated or adapted in light of data science.

This model is depicted in a widely accepted framework for educational standards for computer science in Germany, developed by the German *Gesellschaft für Informatik* (GI; see Brinda, Puhlmann, & Schulte 2009), in which a certain understanding of data is taught in computer science at school. The core aim is to emphasize the difference between data and information: Information is construed as understanding data, which happens only in a human mind. The term data is used to label the representation of data in a machine. In this model, a computational device can process, represent, and visualize only data, but not information (see Figure 5).

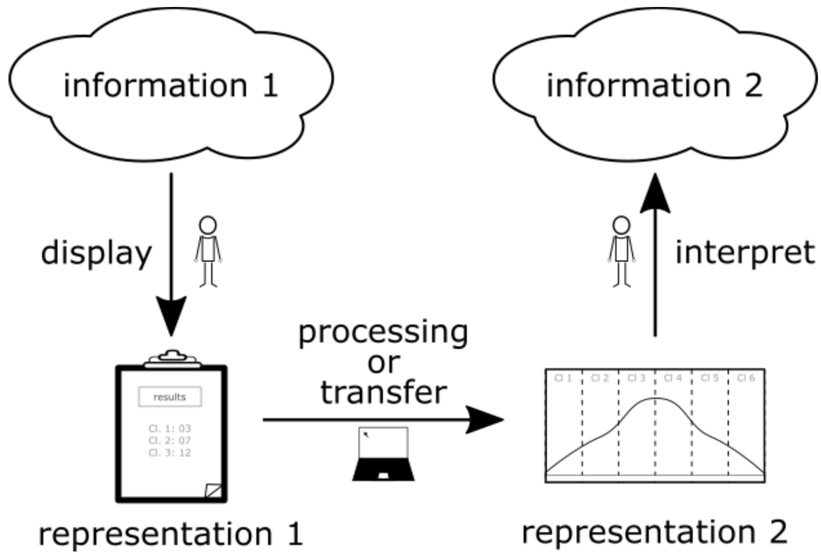

Figure 5: The GI-model of Data vs. Information

Thus, the role of computing technologies should be made clear: we see computing as syntactical operations on data, driven by algorithms working on data represented in suitable data structures. In lower secondary education, the focus is more on operations in standard applications, as depicted in Figure 5, where a data table is (e.g., by an appropriate spreadsheet visualization) transformed into a line graph. In connection with such transformations, rules for appropriate visualizations should be learned, allowing for meaningful interpretation.

In upper secondary education, the focus is more on modeling suitable data structures to store, organize, and retrieve data. Here the focus is on databases and SQL. However, the current focus probably needs to be adapted to incorporate other forms of "data management" in the curriculum (Grillenberger 2014).

The model of data vs. information serves well the intended purpose of highlighting the difference between human and technological data/information processing. In data science, however, more differentiated views on data are used, although there is no consensus on one core model. A popular one, the DIKW-pyramid, distinguishes between **d**ata, **i**nformation, **k**nowledge, and **w**isdom; in connection with this model there is a discussion on the problems of defining knowledge (in contrast to information) as an exclusively cognitive phenomenon (see Wikipedia 2017). In our data science curriculum, we probably need to develop an appropriate definition and educational model of "data" and "information," with regard to the open question whether some levels of understanding (== (information)) should be construed as an exclusively cognitive capacity of the human mind that eludes computational data processing. This is especially important with regard to artificial intelligence and how this will and should be included in the data science curriculum model.

A model of the data science process

The above outlined model of data vs. information is also based on a classical idea of computing and the general structure of algorithmic processing: First, data is represented (Input); then algorithmically processed or computed (Processing); and then the result is presented via a user interface (Output). This IPO-model is in line with classic algorithmic problem solving—and also echoes the underlying educational model of problem solving and computational thinking as an important learning goal. In this view, students learn to analyze a problem, design a solution, and then implement and test the solution. The solution is the program (The P in the IPO model). The roles of the human and the technology are strictly separated. It can be doubted whether this model is suitable as general problem solving process (with technology), see e.g., Tedre & Denning (2016). In current interactive systems, the user is not only applying pre-defined solutions but interactively designing such solutions. Hence, the strict separation between tool-building and tool-usage becomes blurred.

In terms of the IPO-model, this can be seen as a quick succession of IPO-cycles with immediate feedback. For example, using a standard tool, one can quickly produce different types of data visualization, tweak and adapt them, and decide which general type of visualization (e.g., bar plot vs. pie chart) to use, based on the result. So overall, the problem-solving process is not done prior to technology use, but interactively while using computing. Such a problem solving process is not captured by the idea of structuring data and designing and implementing algorithms, but needs to take into account the socio-technical system or hybrid system in which the human operates (Schulte, Sentance, & Barendsen 2018).

Overall, when problem solving in data science relies on computational tools, the question arises: to what extent do they need to be transparent, and to what extent can they be treated as black boxes? When thinking of using a pre-defined tool to transform some data in a basic visualization (pie chart, bar chart, line chart) it would seem that using the tool as a black box is OK. But what if machine learning is used in a data science course: Would it be appropriate to treat it as a black box, too? Probably not.

Similar questions arise with regard to the aims of project-based learning and the roles of tools: Is it about implementing, or using, or understanding? Is there a need for a new educational approach for the role and/or a new role of problem solving (Tedre & Denning 2016)? And, in summary, what is the need for a new educational process model for data science projects at school?

Data processing can be seen as a central notion in a data science curriculum, while traditional models for projects in computing education focus on software or programming projects. The question is whether or how far these models need to be adapted to data science projects. In computer science education at school, models for organizing software projects are common. Such a process model resembles models from professional software engineering, so that students in class can learn from experiencing a software project. As data science also strongly focusses on “data projects,” we will discuss some lessons learned from software projects in education.

Originally, educational projects were oriented on the traditional “waterfall” model (Royce 1970), but subsequently more and more cyclic approaches and elements from agile process models were included. Together with this methodological shift, the learning goals changed, too. Originally the idea was to enable hands-on experiences in an authentic manner, but the increasing size and complexity of real software projects revealed that these expectations were unrealistic. In response, expectations shifted towards more general goals such as learning teamwork, problem solving, and only on some aspects of software development—especially those connected to earlier phases and less with constructing real production systems. The earlier phases focus more on developing ideas for solutions. This development leads to more focus on modeling, and less focus on programming.

Sometimes a consequence is that a successful project probably doesn’t have to work, but it suffices to demonstrate (only) a promising idea for solution. See Berger (2001, p 277ff) for a discussion of teachers’ expectations. Berger interviewed teachers who teach both math and computer science; this group is likely similar to our prospective data science teachers. In his study he concludes that teachers are likely to be satisfied with promising ideas as the result of a project, and regard problems with implementing the solutions as more or less irrelevant. With regard to software projects this means that (only) prototypes are designed and built; nevertheless, a prototype necessarily shows some characteristics of a real solution. We are not sure how this will be with data science projects: Does a preliminary and in some aspects wrong data analysis help the student understand the real meaning of the data, or will it lead to wrong interpretations of the data, and in turn also to a wrong understanding of data science and data science methods?

It might be that teachers (and students alike) who are used to seeing projects as developing prototypes tend to trust their preliminary data analysis too soon and forget to systematically rule out counter explanations. In summary, a suitable and useful conceptualization of data projects is a crucial question for the data science curriculum.

There are yet more issues to projects:

First, differences between industry and education: In computer science education it became clear that the overall goals for projects in industry versus education are different (Schubert & Schwill 2011): In industry a group of highly skilled and trained experts works together in a team to produce a working solution; in education a group of untrained learners work together to learn something together. In the educational context, therefore, the division of tasks and e.g., the forming of sub-groups has to be done in a way that ensures the same learning opportunities for all.

Second, from the perspective of computer science curricula, data science projects add a new approach to problem solving. Classically, the process is roughly organized in phases like analysis of the problem, designing a solution, and implementing the solution (with probably several iterative and cyclic steps added). When machine learning on large data sets is applied, then the solution is not directly designed by a human, but “learned” by the machine, based on training data.

While there is this new approach, pragmatically, the overall approach to machine learning is still chosen or influenced by humans, but probably on another level. Such human influence or participation may take place in each of the phases of a data process.

We thus believe it makes sense to conceptualize data science education and data projects in the context of “hybrid” systems. In a report on the future of jobs, the consulting company Cognizant formulated the idea of a hybrid system in terms of a future job, named “Man-Machine Teaming Manager,” whose task is to “help combine the strengths of robots/AI software (accuracy, endurance, computation, speed, etc.) with the strengths of humans (cognition, judgment, empathy, versatility, etc.) in a joint environment for common business goals. […] The end goal is to create augmented hybrid teams that generate better business outcomes through human-machine collaboration.” (Pring, Brown, Davis, Bahl, & Cook 2017, p. 30). By replacing the focus on business with a focus on society and societal aspects in general, the impact of this view becomes more apparent for education. In Schulte et al. (2018), a first attempt has been made to further elicit this perspective.

Integrating societal aspects

Societal aspects can be viewed as emerging when data science projects are applied. This is also the traditional idea in computer science education, where societal aspects can be discussed in connection with applying software projects. However, the technical view on designing and implementing a new piece of software often overpowers discussion of societal aspects. In addition, software projects often aren’t applied in earnest, but are more or less “toy projects” which do not really raise any societal impact or implication. Therefore, unfortunately in computer science education, discussion of societal aspects is more or less decoupled from the more naturally occurring technical aspects of software projects. Similar issues will probably arise with data science education and data science projects. We hope that the notion of hybrid systems helps to integrate societal aspects.

Another way to support teachers and the implementation of the curriculum is to help them to teach ethical issues through carefully-designed curricular material / teaching examples. One specific question that arises for the data science curriculum is whether it makes more sense to integrate reflection on societal aspects into data projects, or instead to establish a learning phase or learning module that solely focuses on societal issues. The first version probably makes inclusion of societal aspects more natural for teachers, whereas when presenting such issues in isolation, teachers are probably more inclined to leave out these aspects and think that societal issues should be taught in social science subjects at school.

With regard to curriculum emphasis probably teachers differ in their approach to inclusion of societal aspects. While we do not have empirical data for data science teachers, a study with chemistry teachers showed that societal aspects are regarded as less important by German teachers (in comparison to other curriculum emphases), see (Driel, Bulte, & Verloop 2008; Markic, Eilks, van Driel, & Ralle 2009).

Tools and resources

In order to do data science, computational tools are needed. We can broadly distinguish two types: On the one hand are interactive tools like spreadsheets, which let one directly manipulate and visualize data. On the other hand are tools like RStudio or Jupyter Notebooks, in which data manipulation is done by using programming languages like R or Python. The first type of tools relies on the "What you see is what you get" and "direct manipulation" paradigms, which aim to create the impression that the user directly works with the data and gets direct feedback. Programming on the other hand is more indirect, as first a set of data manipulations is coded in a formal syntax, and then applied. Both approaches have their merits and fallbacks; e.g. WYSIWYG-tools are easier to use, but programming tools are better for checking how data was manipulated: one can change the script and run again on the original data. We think that both types of tools should be introduced and reflected on by the students.

From the computing education perspective, the intention is not only to use tools in order to learn data science (learning with tools), but also to learn about tools. Learning about tools includes understanding the role and influence different tools have on the data science process, and to understand that tools are designed and constructed for a purpose—and hence that tools can be re-designed. Often, easy-to-use interactive tools are not readily open for re-design by a user, whereas the more programming-like tools afford and inspire adaptation to one's own need.

OUR APPROACH TO CURRICULUM DEVELOPMENT

As outlined at the beginning of this chapter, we draw on the model by Thijs & van den Akker (2009). From the different approaches described there we take on a mix of the communicate and the pragmatic approach. In terms of the communicative approach we have organized a data science symposium and invited experts to discuss perspectives for data science at school from differ related perspectives (see the other chapter in this report), and we have help from four observers which at the end of the symposium reflect their impression on the results of the discussion. This is in line with the approach where the aim is to reach a consensus among experts. From this perspective more such discussion would need to take place with drafts of the curriculum. On the symposium no curriculum draft could be designed, but first promising perspectives be discussed.

In terms of the pragmatic approach we will collaborate locally, and implement a draft of the curriculum in one school: the idea is to meet the requirements of the users, the teachers at school, and get frequent feedback on the curriculum draft via its implementation at school. In terms of the pragmatic approach a cyclic development with refinement of drafts would be way to go. See Thijs & van den Akker (2009) p. 19.

According to them, sustainable curriculum development is based on the synergy with teacher development, and school organization development. For the Katter one option is to employ so-called project courses in upper secondary school, which are open for school curriculum development.

For teacher development that means we have to think about teacher education, too.

One way to go—in line with the above outlined general approaches to curriculum development—is to use design-based approaches and educational reconstruction.

In this process, (Thijs & van den Akker 2009) p, 35ff, suggest to "focus on elements that are essential for the innovation and which may, at the same time, be considered vulnerable as a result of possible complexity or lack of clarity."

CONCLUSION

Designing a curriculum for data science is a challenging task due to a number of issues: Its interdisciplinary nature, complex prerequisites, fast developments, broad application areas, and its relevance for future lives of the students, not to mention the missing teacher education in the area of data science.

There are a number of challenges: The likely heterogeneity of students; limited knowledge on students interests and prior knowledge (and misconceptions); the tension between open exploratory and project-based teaching and learning phases and the need for systematic development of competencies; the broad and complex nature of the field and limited experiences in educational reconstruction and reduction of topics for data science at school. Luckily, we can draw on some resources, as discussed above, e.g., the traditions in computing and math education that already include some aspects of data science, the experiences and inputs from the experts in the symposium, and so forth. In addition, we will focus our task on what, in our view, are some of the most important aspects of the curriculum: the rationale, aims and objectives; and the role of tools (and best practice). These dimensions were also the dimensions our panelists suggested we should especially focus on in observing and reflecting on the symposium.

We plan to design the curriculum in the following cyclic steps:

1. In order to make the planned curriculum live, teachers need to implement it in school. We thus aim to develop material for implementing the curriculum (lesson plans, assessment, …) collaboratively with teachers of pilot schools; and to reflect on the outcomes with experts.
2. We plan to enrich the developed material with description of the underlying rationale, pedagogical goals, and teacher guidance.
3. In addition, we want to develop material for teachers' professional development (**CPD**) courses based on 1. and 2.

REFERENCES

Aoun, J., & ProQuest (Firm). (2017). *Robot-proof : higher education in the age of artificial intelligence*. Cambridge, Massachusetts: The MIT Press,.

Berger, P. (2001). *Computer und Weltbild: habitualisierte Konzeptionen von der Welt der Computer*. Springer-Verlag. Retrieved from http://helianth.net/pap/cuw.pdf

Berthold, M. R., Borgelt, C., Höppner, F., & Klawonn, F. (2010). *Guide to intelligent data analysis: How to intelligently make sense of real data*. London: Springer.

Biehler, R. (1997). Software for learning and for doing statistics. *International Statistical Review, 65*(2), 167–189.

Biehler, R. (2014a). Leitidee Daten und Zufall – fundamentale Ideen aus Sicht der Statistik. In H. Linneweber-Lammerskitten (Ed.), *Fachdidaktik Mathematik - Grundbildung und Kompetenzaufbau im Unterricht der Sekundarstufe I und II* (pp. 69–92). Seelze: Klett Kallmeyer.

Biehler, R. (2014b). Leitidee Daten und Zufall – Fundamentale Ideen aus Sicht der angewandten Stochastik. In H. Linneweber-Lammerskitten (Ed.), *Fachdidaktik Mathematik - Grundbildung und Kompetenzaufbau im Unterricht der Sekundarstufe I und II* (pp. Downloadbereich). Seelze: Klett Kallmeyer.

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clements, A. J. Bishop, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Third International Handbook of Mathematics Education* (pp. 643–689). New York: Springer.

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science, 16* (3), 199–231. doi:10.1214/ss/1009213726

Brinda, T., Puhlmann, H., & Schulte, C. (2009). Bridging ICT and CS: educational standards for computer science in lower secondary education. *ACM SIGCSE Bulletin*, *41*(3), 288–292.

Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In Batanero, C., Burrill, G., & Reading, C. (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education — a joint ICMI/IASE study: The 18th ICMI study* (pp. 57–69). Dordrecht: Springer.

van Driel, J. H., Bulte, A. M. W., & Verloop, N. (2008). Using the curriculum emphasis concept to investigate teachers' curricular beliefs in the context of educational reform. *Journal of Curriculum Studies*, *40*(1), 107–122. https://doi.org/10.1080/00220270601078259

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, *70*(1), 1–25. doi:10.1111/j.1751-5823.2002.tb00336.x

Gal, I. (2003). Expanding conceptions of statistical literacy: An analysis of products from statistics agencies. *SERJ, 2*(1), 3–21.

Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal, 16*(1), 22–25.

Grant, R. (2017). Statistical literacy in the data science workplace. *Statistics Education Research Journal, 16*(1), 17–21.

Grillenberger, A. (2014). Big data and data management: A topic for secondary computing education. In *Proceedings of the tenth annual conference on international computing education research* (pp. 147–148). New York, NY, USA: ACM. https://doi.org/10.1145/2632320.2632325

Grus, J. (2016). *Einführung in data science - Grundprinzipien der Datenanalyse mit Python*: O´Reilly.

Guin, D., & Trouche, L. (1999). The complex process of converting tools into mathematical instruments: The case of calculators. *International Journal of Computers for Mathematical Learning, 3*(3), 195–227.

Harari, Y. N. (2017). *Homo Deus: A brief history of tomorrow*. New York: Harper Collins Publishers.

Haslwanter, T. (2016). *An introduction to statistics with Python with applications in the life sciences*. Cham, Switzerland: Springer.

Huber, P. J. (2011). *Data analysis: what can be learned from the past 50 years*. Hoboken, N.J.: Wiley.

Igual, L., & Seguí, S. (2017). *Introduction to data science: A Python approach to concepts, techniques and applications*. Cham, Switzerland: Springer.

Klafki, W. (1996). *Neue Studien zur Bildungstheorie und Didaktik Beitraege zur kritisch-konstruktiven Didaktik*. Weinheim u.a.: Beltz.

Madden, S. (2013). Supporting teachers' instrumental genesis with dynamic mathematical software. In D. Polly (Ed.), *Common core: Mathematics standards and implementing digital technologies* (pp. 295–318). Hershey, USA: Information Science Reference (an imprint of IGI Global).

Markic, S., Eilks, I., van Driel, J., & Ralle, B. (2009). Vorstellungen deutscher Chemielehrkräfte über die Bedeutung und Ausrichtung des Chemielernens. *CHEMKON, 16*(2), 90–95. https://doi.org/10.1002/ckon.200910091

McKinney, W. (2017). *Python for data analysis: Data wrangling with pandas, numPy, and iPython*. USA: O´Reillly.

McNamara, A. (2015). *Bridging the gap between tools for learning and for doing statistics.* (Doctor of Philosophy in Statistics), University of California, Los Angeles. Retrieved from http://escholarship.org/uc/item/1mm9303x

O´Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. New York: Crown.

Pring, B., Brown, R. H., Davis, E., Bahl, M., & Cook, M. (2017). *21 jobs of the future: A guide to getting – and staying – employed over the next 10 years*. Cognizant.

Prodromou, T., & Dunne, T. (2017). Statistical literacy in data revolution era: Building blocks and instructional dilemmas. *Statistics Education Research Journal, 16*(1), 38–43.

Ridgway, J. (2015). Implications of the data revolution for statistics education. *International Statistical Review, 84*(3), 528–549.

Royce, W. (1970), Managing the development of large software systems. *Proceedings of IEEE WESCON*, 26 (August): 1–9.

Schield, M. (2017). CAISE 2016 promotes statistical literacy. *Statistics Education Research Journal, 16*(1), 50-54.

Schubert, S., & Schwill, A. (2011). *Didaktik der Informatik*. Springer.

Schulte, C., Sentance, S., & Barendsen, E. (2018). Computer science, interaction, and the world. In *Computer science education: Perspectives on teaching and learning in school* (1st ed., pp. 57–74). Bloomsbury Academic.

Spitz, M. (2017). *Daten - das Öl des 21. Jahrhunderts? Nachhaltigkeit im digitalen Zeitalter*. Hamburg: Hoffmann and Campe.

Sutherland, S., & Ridgway, J. (2017). Interactive visualisations and statistical literacy. *Statistics Education Research Journal, 16*(1), 26-30.

Tedre, M., & Denning, P. J. (2016). The long quest for computational thinking. In *Proceedings of the 16th Koli Calling International Conference on Computing Education Research* (pp. 120–129). New York, NY, USA: ACM. https://doi.org/10.1145/2999541.2999542

Thijs, A., & van den Akker, J. (Eds.). (2009). *Curriculum in development*. Niederlande: SLO - Netherlands Institute for Curriculum Development.

Toomey, D. (2017). *Jupyter for data science: Exploratory analysis, statistical modeling, machine learning, and data visualization with Jupyter*. Birmingham: Packt Publishing.

Tukey, J. W. (1962). The future of data analysis. *Annals of Mathematical Statistics*, *33*, 1–67.

Weigend, A. S. (2017). *Data for the people : how to make our post-privacy economy work for you*. New York: Basic Books.

Wikipedia (2017) DIKW pyramid. (2017, November 18). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=DIKW_pyramid&oldid=810990870

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–265.