TEACHING PRIVACY OUTDOORS – FIRST APPROACHES IN THE FIELD IN CONNECTION WITH STEM EDUCATION

Sandra Schulz
Humboldt-Universität zu Berlin, Germany

Abstract. *This paper presents an approach to teach privacy in secondary schools. The research field is highly topical and there is a lack of educational concepts. In this approach the students analyze pictures taken with smartphones. They are taught about hidden information stored in the pictures and how to use them to find the owner of the pictures. Smartphone pictures as learning content are related to their daily life, what makes it relevant and interesting for the students. Two versions of the approach were tested in the last years in classrooms and are presented in this paper. The evaluation shows that the students were very engaged and gained privacy awareness. A third version of this approach takes this activity outdoors and connects it with STEM subjects.*

Key words: privacy, teaching, computer science education, STEM education, meta data, problem-based learning

INTRODUCTION

In the digital age the smartphones are our constant companions, what makes the life at first glance easier and smarter. However, topics like digital addiction and the effects on social life are being researched. It was found that the smartphone usage, especially WhatsApp, dominates our daily life (Montag, Błaszkiewicz, Sariyska et al. 2015). This applies equally for youngsters and adults. People are sending tones of data without reflecting what kind of data is sent. That shows on the one hand, that privacy awareness is not high enough. On the other hand, most people do not know what information is hidden in their data, for instance in pictures. In the K-12 Computer Science Framework "privacy and security" is described as a crosscutting concept (K-12 Computer Science Framework Steering Committee 2016). Thus, privacy is an important area in the digital age that needs to be brought into school. The barriers are high, because the topic is closely related to law and it seems to be tough to teach it in an interesting way. "I have nothing to hide" and other rejections are additional hurdles when topics such as data minimization are brought up.

In this paper an educational approach for computer science lessons in secondary school is presented. The analysis of meta data out of smartphone pictures is clearly settled in the area of data protection and privacy in computer science. Two approaches were realized until now. A third version connects the approach with partial outdoor activities and interdisciplinary entry points.

First, related literature regarding teaching data protection and problem-based learning is presented. Afterwards, the educational approach is described in more detail, including an overview of the evaluation of the first and second version and data out of the studies. In the third version arguments are given on why it is planned to take this approach outdoors. Then, all three versions of the approach will be compared. Finally, the implementation in school and the empirical evaluation of the three versions are discussed and an outlook is given.

Schulz, S. (2020). Teaching Privacy Outdoors – First Approaches in the Field in Connection with STEM Education. In M. Ludwig, S. Jablonski, A. Caldeira, & A. Moura (Eds.), *Research on Outdoor STEM Education in the digiTal Age. Proceedings of the ROSETA Online Conference in June 2020* (pp. 163-170). Münster: WTM. https://doi.org/10.37626/GA9783959871440.0.20

RELATED LITERATURE

Since 2018 the General Data Protection Regulation (GDPR) is a legal requirement for all European citizen. It addresses companies on how to handle specific personal data. However, it effects students as well. On the one hand, everyday personal data is uploaded by students in social networks and the students are responsible for personal data of themselves and others, for instance when they are part of an uploaded picture. On the other hand, the students need to know which data can be collected by companies and how to protect themselves, for instance through the right of restriction of processing.

Privacy education is highly topical and claimed in education standards. For instance, the K-12 Computer Science Standards (2016) embeds privacy at the end of grade 2, going up to grade 12. For grade 12 it is explained, what should be understood by the students in the area of data collection:

> "Data can be collected and aggregated across millions of people, even when they are not actively engaging with or physically near the data collection devices. This automated and nonevident collection can raise privacy concerns, such as social media sites mining an account even when the user is not online […]" (p. 117).

These contents are highly related to privacy teaching. The declaration as a crosscutting concept also shows the relevance. After the GDPR became effective in 2018, the relevance of privacy teaching has been increasing. However, there are not many concepts on how to teach privacy in school. "If you are not paying for it, you are the product" is one approach presented by Berendt and Dettmar (2018). It is a lesson series, in which the effects of internet tracking and data analysis are uncovered. The tracking of data is visualized and the students learn about biased data by using an example case and how algorithms work in the background.

"Email for you (only?)" is another approach, focussing on how to send e-mails securely. The encryption of e-mails and the importance of encrypted communication are part of the lesson series (Gramm, Hornung & Witten 2012).

Most approaches lack an empirical evaluation or are not relevant for students. The context of encrypted e-mails might work much better for adults than for students, which rarely send e-mails. Information about the underlying methods are often missing. For this approach the problem-based learning seems to be the best fit. In that, the students get a real-world problem and need to figure out how to solve it appropriately. This is quite common in computer science education, for instance to foster critical thinking and problem-solving skills (Kay et al. 2000). For an interdisciplinary approach it seems to be appropriate, because there exist a variety of working techniques to improve problem-solving skills in STEM education as well. Problem-solving, particularly scientific inquiry, is addressed since decades in science curricula (American Association for the Advancement of Science et al. 1993).

EVOLUTION OF AN EDUCATIONAL APPROACH

The kind of data, that is used in this approach is called EXIF data. It is an acronym for "Exchangeable Image File". This data is stored in every digital picture (hidden) and can contain information about camera type, operating system, smartphone type, date and time of the image, GPS information and much more. The GPS information is stored, depending

on the settings of the smartphone, for instance if the GPS location is activated. With only a couple of pictures a profile of the owner can be constructed. Thus, the GPS coordinates contain the most important meta data.

In this paper a couple of pictures is called a "scenario". Having approximately 5 pictures, the included EXIF data can be enough to find the owner of the pictures. If that is not enough, more data can be collected from other sources. For instance, it is possible to search on the internet to find companies or sports clubs around the found GPS coordinates. Afterwards, employees or a training plan on the related websites can be detected. If there are no GPS coordinates included, it is also helpful to have a closer look at the picture to find out about the place it was taken and perhaps an occurred event, using the internet. Street signs, tourist attractions and registration plates can be helpful to figure out a location.

Figure 1 shows a screenshot of the program GIMP 2.10 (GNU Image Manipulator Program), which is a freely-available software to analyze and edit pictures. On the picture (left) a Tic Tac Toe game is shown. The content of the picture is visible for whomever it is sent to. On the right side the EXIF data is shown, which is accessible with programs like GIMP. Into this case, the GPS coordinates, which are pointing to a secondary school in Berlin, can be found. It seems plausible, that the author of the picture is a teacher or a student from this school. Playing games like Tic Tac Toe is more likely done by students than by teachers. The timestamp of the picture is included as well. Thus, it is possible to draw conclusions about when this student is in school and about the timetable (for example a free hour).

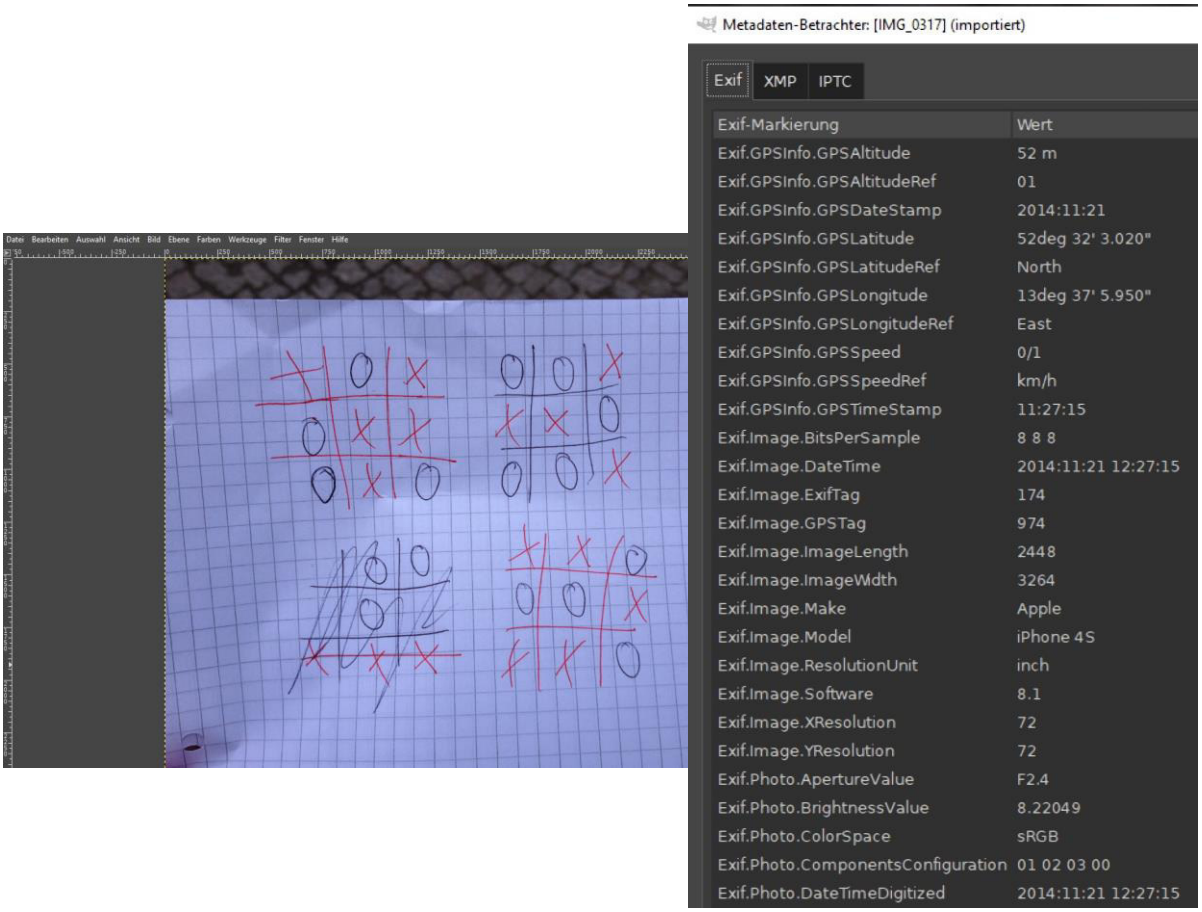

Figure 1: Picture of Tic Tac Toe game (left) and its EXIF data (right) in GIMP software.

First Version (tested in 2014)

In a project to enhance the awareness of data protection and privacy a study with students in secondary school was conducted in 2014 (Schulz and Strickroth 2019). It was tested in a computer science course in grade 11 with 15-17 aged students. In total 9 students participated in this course and the study. The whole project consists of 6 lessons, 45 minutes each, held in 3 blocks. The students got information about privacy laws in Germany and newspaper articles showing risks and chances of the analysis of meta data. The analysis of the meta data was the main part of the project. The students received information on how to insert GPS coordinates into Google Maps. This is necessary, because there exists different metrics about GPS coordinates, which are similar, but will lead to different locations (for instance the sexagesimal system or decimal system). Understanding the differences is important for the success in further lessons. Then, they got some pictures from the teacher to practice the analysis of EXIF data. The students received 5 pictures (generating a scenario), which were taken by the same person and had to build a profile of that person. The students needed to act like detectives and figure out who took the pictures. To analyze the pictures' EXIF data, the students needed the program GIMP, which is a standard freely-available software to view pictures. In the menu bar the students had to choose "picture" and afterwards "meta data" to uncover the EXIF data.

For the evaluation of the project, students' data was collected using questionnaires (pre- and post-test) and the teacher was interviewed. Both showed that the students gained privacy awareness and became more sensitive towards their personal data. The part of meta data analysis was described as "impressive for the students" by the teacher. In the questionnaires the students said that they were surprised by the hidden data and that they would also tell their friends about it. In general, it was an eye-opener, even though the project was not using the students' personal data. This could have been even more impressive.

Second Version (tested in 2019)

In 2019 a revised approach was realized in a 10th grade computer science course consisting of 16 students, aged between 15 and 16. Before this data protection topic was introduced, the students were asked to send 5 pictures each (without people on them) to the teacher. They were given instructions, for instance they had to pick pictures everybody is allowed to see, but not the purpose of the task. The students were introduced to privacy law and meta data. They received the same scenario of 5 pictures to analyze, like the group in 2014. In addition, the students were introduced in the general functionality of geolocation and its different notations. The detailed information about it turned out to be valuable for the successful analysis of EXIF data. As next exercise, the students were given pictures from other students in the course to analyze. They worked in groups of 4 to 5 students and analyzed the pictures of one other student in the course. All groups figured out the owners of the pictures by connecting data and their knowledge about the classmates. Some gave arguments like: "I know Lara was in the zoo on the 20th of November. We had a field trip with the biology course we are sharing. The picture must be from her […]" (having a picture with an animal in the zoo). Other students were using Google Street View to find out about the surroundings of a specific place on the picture to get more information.

After every lesson three random volunteers were asked with the flashlight method to name positive and negative aspects of the lesson. Directly after the data analysis lessons all students mentioned the problem-based approach as a positive aspect. They said: "It was really fun to act like a detective and to find out from whom the pictures are" and "I like to analyze the meta data of the pictures and to search for further data to connect them." The problem-based approach seems to be very motivating for students following the students' answers and the observations of the teacher. Particularly the data analysis from the students' social surroundings raises the awareness and credibility of privacy and uncovers potential problems.

Third Version (planned)

This phase is not yet tested but will combine the actions of version 1 and 2 with outdoor learning and interdisciplinary content. Therefore, the students' tasks will be the same as in the second version. But instead of the analysis of data from the classmates, the students will have to construct own scenarios for their classmates, before the different scenarios will be distributed to all students. It is necessary for the construction of scenarios that the pictures are gathered outside and for the students to travel through the city to make it more interesting, with different places included. The scenario is more interesting when different kinds of data are provided or need to be found. It is also possible to deactivate GPS during the process of taking pictures, when the students are supposed to look at the picture in more detail. The students are allowed to manipulate EXIF data to create contradictions as long as they give clear hints of the manipulation. The EXIF data of a pictures might for instance show that it was taken by night, even though it shows bright daylight. Consequently, the complexity of the scenario may be adjusted to fit different levels of difficulty. Students that solve their tasks quicker than others might be given additional tasks as developing scaffolds for their scenarios. The main task for the analysis is always to find the author of the pictures (if possible) and to gather a lot of information about that person.

Bringing the third version outdoors has the following advantages for the students and the complexity of the project:

- **Variety of GPS coordinates**: If data is just gathered inside, for instance the school or the students' homes, the complexity of the scenarios is reduced and the variety of GPS coordinates is more limited. If the students want to point at something known by everyone, classrooms would be their only option.
- **Protection of students' data**: In this project it is necessary to protect the students' data and the students should feel secure. Information like their place of residence should not be figured out by other students if the owner of the pictures does not want them to.
- **Authentic data**: Taking pictures from the internet and change the EXIF data could be an alternative, but it is hard to find pictures on the internet, in which EXIF data is still included. Furthermore, during the analysis with software like GIMP, the changes of the EXIF data would be traceable. To increase the authenticity of data and of the fact, that many information is hidden, it seems to be logic to provide "real" data of their daily life and not do manipulate meta data of the pictures.

- **Differentiation in complexity**: Making pictures from tourist attractions or street signs make it easier to find a location, even if the GPS coordinates are not provided.

EDUCATIONAL PERSPECTIVES FOR COMPUTER SCIENCE AND STEM

All three versions contain the EXIF data analysis as the main part (see table 1). The first version is the shortest and can be integrated in usual school lessons in the classroom. The second version encompasses more lessons, because the students are analyzing two scenarios and are more engaged in understanding the function of GPS location. This leads to deeper learning because the students get more practice in the analysis of data and they get a scenario out of the real-world. Having pictures from the classmates can show, that the teacher's scenario is from the real-world and their pictures contain the same data. Finally, the third version needs at least 12 lessons, because the students are going outdoors to construct their own scenario. Therefore, it is necessary to get guidelines from the teacher on how to construct interesting scenarios with an adequate level of difficulty. The construction of own scenarios should be prepared in the classroom to plan which places to visit and what data to collect. Gathering the data will need 2–4 lessons, depending on restrictions by the teacher, for instance how far to travel via train. Afterwards the students need to process their scenarios in classroom. At that stage, interdisciplinary problems may be indicated (see below). Depending on the implementation of this version in school, it can be valuable to give the interdisciplinary tasks before the outdoor activities.

	First Version	Second Version	Third Version
Number of lessons (45 minutes)	6	10 – 12	> 12
Further subjects to connect with	-	Geography	Geography, Mathematics, Physics
Environment	Classroom	Classroom	Classroom, outdoor
Students' activities	Analysis of foreign scenarios (teacher)	Analysis of foreign scenarios (teacher and students)	Constructing an own scenario; analysis of foreign scenarios (teacher)

Table 1: Comparison of the three versions.

Especially the third version of this approach seems to be perfectly suitable for problem-based STEM learning. GPS navigation is a topic addressing multiple school subjects and is connectable to real-world problems. For instance:

> **How is GPS navigation working?** It is necessary to explain the functionality and purpose of satellites and the geocentric coordinate system (Geography). By using GPS it is also possible to measure the speed of an object. This is applicable because of the Doppler-effect and differential calculation (Physics and Mathematics).
>
> **What factors make GPS navigation imprecise?** Students might find differences between their actual location and the location that is shown in an app like Google

Maps. This can happen when they are surrounded by high buildings, because of reflection (Physics) or different weather conditions (Geography). General discussions on how signals can be sent and disturbed could be connected, too.

What is the difference between GPS location and Wi-Fi location? This question adds an ethical perspective to this topic. Is it ethically correct to use private Wi-Fi data to locate foreign devices? When the Wi-Fi location is used to improve an outdoor game, is the information stored by the manufacturers of the game? What are they doing with this information?

These questions are just a few examples of a multitude of real-world problems. Constructing meaningful and interdisciplinary problems for students can improve their motivation. Furthermore, the subject of EXIF data can be the hook to address other disciplines in computer science as ethics, for instance, which is oftentimes neglected. Nowadays many data analyzing algorithms can be implemented, but the question needs to be addressed, if the implementation is ethically correct. Using EXIF data is one example, that needs to be discussed in the field of ethics. On the one hand, EXIF data is an effective way to organize data and to provide more information about data. On the other hand, EXIF data cause damage, if the user is not aware of what personal information is being revealed. Questions to discuss in class can be: Is it ethically right to use EXIF data information without the knowledge of the owner? Can I upload pictures of others on social media platforms without asking them? What about strangers being in the picture, when taking pictures outdoors? Building the connections to different disciplines seems to be challenging for many teachers. Teachers should not insist in tackling the project by themselves. Privacy officers and teachers from related subjects can easily get involved to make this project even more interesting and diverse.

OUTLOOK

Bringing these approaches into school seems to be valuable. The first and second version showed, that the students were engaged with privacy and gave very positive feedback concerning the project. For an interdisciplinary implementation it is suitable to use project weeks to bring the related subjects together and to have enough time to address problem-based questions. Working together with STEM teachers seems to be appropriate to accurately explain interdisciplinary phenomena and to make visible, that problems are not limited by the boundaries of single subjects. After the implementation in STEM, other connections should be made, for instance to Art or History. This can be achieved by focusing on specific themes for the pictures, like street art or statues. This can improve media competences in different subjects, which are fundamental today.

For an empirical evaluation a pre- and post-test of privacy awareness might be a good choice. Additionally, an interview with the teacher should be conducted to validate the students' answers. A qualitative analysis of a portfolio is also an appropriate way to implement a phase of reflection for the students and to find out about the students' view, if they are planning to change their data handling or to delete the EXIF data before sending pictures. It is necessary to repeat questionnaires and interviews after a couple of weeks to validate, if the project has brought a sustainable change to the students' privacy awareness.

References

American Association for the Advancement of Science et al. Benchmarks for science literacy. Oxford University Press, 1993.

Berendt, B., & Dettmar, G. (2018). If you're not paying for it you are the product: A lesson series on data, profiles, and democracy. In Paderborn Symposium on Data Science Education 2017: The Collected Extended Abstracts. (pp. 55-58). University of Paderborn.

Gramm, A., Hornung, M., & Witten, H. (2012). Email for you (only?) design and implementation of a context-based learning process on internetworking and cryptography. In Proceedings of the 7th Workshop in Primary and Secondary Computing Education (pp. 116-124).

Intersoft consulting services AG (2018). General Data Protection Regulation (GDPR) https://gdpr-info.eu (last checked 04.03.2020).

Kay, J., Barg, M., Fekete, A., Greening, T., Hollands, O., Kingston, J. H., & Crawford, K. (2000). Problem-based learning for foundation computer science courses. Computer Science Education, 10(2), 109-128.

K-12 Computer Science Framework Steering Committee. (2016). K-12 Computer Science Framework (2016). Retrieved August, 28, 2017. https://k12cs.org/wp-content/uploads/2016/09/K–12-Computer-Science-Framework.pdf

Montag, C., Błaszkiewicz, K., Sariyska, R., Lachmann, B., Andone, I., Trendafilov, B., Eibes, M. & Markowetz, A. (2015). Smartphone usage in the 21st century: who is active on WhatsApp? *BMC Res Notes* 8, 331. https://doi.org/10.1186/s13104-015-1280-z

Schulz, S. & Strickroth, S. (2019). Versteckte Meta-Daten und Datenschutz - Die Förderung des Datenschutzbewusstseins durch das praxisorientierte Auslesen von Meta-Daten. *LOGIN Verlag*, 191/192, 67-73.