GDPR in the Small: a field study of privacy and security challenges in schools

Francesco Ciclosi
University of Trento, Trento, Italy
francesco.ciclosi@unitn.it

Giovanna Varni
University of Trento, Trento, Italy
giovanna.varni@unitn.it

Fabio Massacci
University of Trento, Trento, Italy
Vrije Universiteit Amsterdam, The Netherlands
fabio.massacci@unitn.it

***Abstract*—The GDPR was enacted to reign in the mighty corporations of the internet. Then, it was unleashed on all organizations, large and small alike. We report the results of a multi-site field study on Italian schools, and the challenges they face to implement the GDPR while running activities full of sensitive issues without an army of legal and compliance officers. The sample study consisted of one kindergarten, ten primary schools, two junior secondary schools, and two secondary schools. We did not find evidence of the privacy paradox (spotless on paper but careless on the field). In contrast, school staff mostly crumble when by-the-book procedures cannot be implemented with the resources that they actually have. We discuss what happen on the field, from critical privacy incidents with potential impact on pupils security and safety, to 'formal' privacy incidents for which life is too short to bother- and how a risk-based approach could address them.**

1. Introduction

Small organizations often struggle to comply with data protection rules. The ideal situation is, of course, building a corporate culture of data protection [1], but this is hard to do when tight budgets make it impossible to pay privacy trainings for regular staff or to allocate dedicated staff to privacy compliance [2], [3]. Even the Data Protection Officer (DPO) is often not the 'representative' of an office (as the DPO of Google or Facebook, or even of a University might be) but an individual who end up playing at once advisor and guardian, executor and auditor [4], [5]. Limited knowledge of actual processing might also mislead staff into thinking that privacy management is sometimes unnecessary [6].

Among small organizations, schools are an ideal case study [7], [8], [9], [10], [11], [12]. To make ends meet, they focus on formal compliance: information sheets on processing, manuals, or registers of processing activities. This phenomenon also happens for small operators in charge of securing critical infrastructures [13]. On paper, they tick all the boxes. In practice, they might fall into privacy incidents as we also show in this paper, GDPRxiv [14], an open-source archive collecting all official GDPR rulings from 2018 to 2023, reports that 123 rulings in EU member states were doled to educational institutions. Compare them to 34 rulings to the four Internet giants, 62 to telecom operators, and 138 to banks. A huge fine is often immaterial for banks, telcos or the big four. For a school, whose budget is mostly incompressible (staff salaries), a small fine might wipe out its discretionary budget.

Security technologies could help. Yet, when we deploy them we must think to field use. Consider a simple and apparently obvious question: *Does any device in the school need a password for individual staff members tailoring access to individual class school registers?* It is RBAC 101 and ensures GDPR security measures (Art. 32). Consider a Principal's challenge: the solution has to work when a teacher calls sick at 7:30am and at 8:00am a replacement needs to step in a class of 10 years old. At 7:55am you have finally found the person. Second challenge: a divorced parent is not allowed by a judge to see the kid. Make sure the replacement teacher (as all teachers of the class) knows they can't hand Bob to the father. If availability trumps confidentiality, you could just print the list of requirements and post it on the door. Third challenge: how do you make sure that a third party doesn't figure from the posted list that Alice is authorized to walk home alone as both her parents work long hours? Now, depending on the technological solution (tablet with teachers shared password in locked drawer or a top-notch RBAC system), you need an IT administrator to change access control in 5 minutes to a sensitive list of minors without making mistakes. Pity the only IT admin is fixing PCs for the 8am Physics class.

It is pointless to preach small organizations to a standard that they cannot achieve, as they lack the resources to do so. This is a recipe for privacy incidents to occur.

Our goal is to support schools to make sure we identify and mitigate structural and high risk incidents and balance the resources available in the true spirit of the law[1]. The contribution of this work are manyfold:

1) We carried out a field observational study to explore the role of human factors in the socio-technical Italian school system by analyzing the practice on the field of educational institutions including different levels of schooling (i.e., kindergarten, primary, middle, and high school);

1. Recital 4 GDPR: [...] The right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality.

2) Based on the observation, we propose a coding taxonomy of factual observations and *sequences of actions leading to incidents*;
3) We propose a classification of incidents in two types, *structural* incidents related to a poorly designed procedures, and *occasional* incidents, due to an incorrect and episodic behavior of staff
4) Finally, building on the analysis of the requirements and fines given to schools we classify incidents ranging from *high risk* privacy incidents (possibly resulting in serious damages to the rights and freedom of people and *low* risk privacy incidents.

In summary, our finding is that there is no privacy paradox (an inconsistency between people's privacy attitude and their behavior), but just a conflict between competing needs of staff who is 'called elsewhere' (because there is no one else to answer the call) and in doing so they may unfortunately generate data leaks. Thus, the solution to these privacy problems is the design of organizational protocols robust to interruptions and amenable to mitigations which balance costs with risk.

An important disclaimer applies here. The opinion whether a privacy incident belongs to the categories 'high risk' or 'life is too short' is entirely due to the authors and it should be in no way attributed to the people working at the institutes under investigation. Unfortunately, the GDPR applies identically to the large and to the small, to the tragedy and to the mockery.

1.1. Artifact Availability Statement

We make available through Zenodo (https:\doi.org/10.5281/zenodo.15133671) the following material: the templates of the information and consent forms, survey questionnaires and survey results, coding guidelines with selected examples, codes' occurrences and correlations. The field observations will be not publicly released to preserve the privacy of people participating in the study. Full text of the field annotations cannot be released as it could allow teachers and staff of the surveyed schools to re-identify most individuals responsible for the privacy incidents. Researchers willing to analyze the raw material can contact the authors to view it on our premises.

2. The Italian School System and its GDPR Requirements

2.1. The Italian School System

Since school systems are different from country to country, we provide here a brief introduction to the Italian school system. By the term 'school' one typically denotes a single type of school, such as an elementary school, with a principal, the teachers, and then the pupils distributed into classes. For example an elementary school typically includes pupils of the age bracket 5-10. In Italy, the schools are grouped into larger *Educational Institutions (Istituto Comprensivo)* (EIs for short) grouping early childhood, primary, junior secondary and secondary schools, or some secondary educational institutions of different type (for example scientific vs. vocational education). Figure 1 summarizes this structure. As should be clear from the above explanation, the Italian Istituto Comprensivo is *not* a "*comprehensive school*", term used in some countries to designate post-primary schools that accommodate pupils between the ages of 11 and 18.

The Italian School System in a nutshell. A public *Istituto Comprensivo* is an institution composed by N primary schools (usually at different sites) whose pupils continue in M middle schools, then in K high level school types (e.g. science and classics vs accounting), all under the same principal. Each level comprises few schools of each type. A principal is in charge of the entire organization and some deputies supervise the different sites. The level of inequality in the Italian school system is limited in comparison to the USA, so the difference between public or private schools is mostly immaterial.

Figure 1. The Italian Istituto Comprensivo

Each EI is a part of the national school system while maintaining its own administrative, educational, and organizational autonomy. EIs can be spread over several sites. The main site might include the whole educational pipeline from primary up to high-school, while a satellite site in a remote part of the county might only offer a primary school.

A *Principal* heads the EI and is responsible for both administrative and educational management. The school principal of each EI is selected from among tenured teaching staff in an open competition (so not necessarily among the teachers of the very schools of the EI). Moreover, each EI has a dedicated administrative office (secretary) for administrative purposes and to maintain relations with third parties. The EI principal decides all activities after consulting with the *School Council*, an internal body of the EI including as its members the parents' representatives. If the EI is multi-site, a deputy is appointed for each site.

2.2. The GDPR Requirements

The handbook of the Italian Data Protection authority reports all privacy requirements peculiar to schools [15]. Schools can only process strictly necessary data, regardless of the nature of processing (electronic or paper). They are allowed to process personal data only for institutional purposes (e.g., education and training) defined by law. For other purposes, some types of data (e.g., those required from theater courses not included in the school curriculum) may be processed after obtaining the consent of the data owners. Special categories of personal data relating to pupils (such as ethnic or racial origins, religious beliefs, health status, and so on) may be processed only for specific purposes established by law. Schools must also make sure that data

shared on online platforms is kept secure and protected. Documents with sensitive data should be stored and transported in closed folders for protection.

The *Data Controller* is the school's legal representative, i.e., the principal, who defines the activities to be undertaken and how to implement them. The school's administrative and teaching staff are authorized to process personal data in their respective areas of competence. The data controller must (i) prepare a processing register to have an up-to-date picture essential for the assessment and the analysis of the risk associated to such processing; and (ii) appoint a *Data Protection Officer* (DPO). In addition, a set of technical and organizational *minimal mandatory security measures* must be taken to prevent both circulation of data among unauthorized colleagues and its unwarranted disclosure to third parties (e.g., other students and families). Staff have a duty of confidentiality and professional secrecy regarding the data they process .

Table 8 in the appendix lists the main issues of data processing the Italian legislation on the the protection of personal data in the school setting.

2.3. Data Protection Authority Measures

Private operators are typically fined because they took some person's data for some purpose and then started using for untold, own commercial purposes. In contrast, schools are typically fined because some staff did something in good faith that disclosed data, sometimes even of a single pupil whose parents complained.

The Italian Data Protection Authority (Garante della Privacy) uses three instruments:

- *Rules/manuals* apply to a category of organizations in terms of recommended practices and minimal security measures. There is a specific manual (Vademecum) on schools [15].
- *Enforcements* includes fines and obligations of the affected parties to stop (or start) using some process or technology. They can be found in GDPRxiv.
- *Warnings* just notify the offending party that they have to fulfil a certain condition before is it does not want to incur stronger enforcement measures.

Since the full implementation of the GDPR, over the past seven years, 29 schools received either sanctions or warnings and 58 out of 2561 published published issues) in the database of the Italian Data Protection Authority refer specifically to educational institutions.

On GDPRXiv out of 388 enforcement actions concerning GDPR in Italy, 27 involved schools receiving sanctions for posting excessive personal information online (16 cases), in restricted areas of the electronic register (3 cases), or on bulletin boards (1 case).

Mere *warnings* are not in GDPRxiv, but they exist (14 out of 58 measures): the authority recognized that the violations were determined by small scales, lack of resources, or mere material error and only issued a warning. Analyzing the case history of rulings provides only a partial view of real-world occurrences. The scarcity of ruling case histories can be attributed to various factors, such as the absence of filed complaints with the Data Protection, arbitration of specific cases (e.g., when data subjects and data controllers reach their own agreements), the lack of publication of some measure on the Authority's website (it is a secondary sanction), and the Authority's inspection activities focused on other organizations.

Field observations are therefore crucial for understanding the extent and prevalence of privacy incidents.

2.4. Non-Goals (According to the Principals)

After talking with the principals, a clear non-goal emerged: the management of digital data by third-parties. This was surprising as in 2019 there was a significant controversy in Germany on the use of Microsoft O365 online services in school, following a report of the German Data Protection Conference (DSK) – which consists of the German Federal Data Protection Authority and 16 state regulators. Office365 was then banned across all schools[2].

Obviously, a part of the data processing happens when schools use third-party applications to administer, for example, emails, assignments, parent-teacher communication, and privacy violation can occur in such system. For example, some EIs made agreements with Google to offer emails to staff and students. Also financial data is often processed by software providers responding to procurement tenders at county level across multiple EIs.

While the responsibility for privacy violations always stay with the data controller according to the GDPR (in this case the EI's principal), if the contract with the third party is well drafted (and most are standard), then the EI has several legal protections. Loosely speaking, if the data of pupils is lost by the software in charge of processing grades, before the buck arrives at the door of the EI's principal, it would have made several other stops. The first target of the privacy authority would be the ministry or the county council (which procured the contract and therefore the EI could not refuse), then the third party itself for having failed the duty of care.

The major concern of the EI is the management of the school's own IT and the mishap of their own making. They are overwhelmingly fined for something their staff did, and not for something their suppliers did. For example, Zoom platform recordings were used for disciplinary purposes, video surveillance systems were used during school activities, and pupil health data was even communicated to other pupil families.

3. Related Work

We reviewed literature on privacy management in small users' communities and small organizations taking into account the last five years of papers published in the proceedings of the following international conferences: *ACM*

2. https://datenschutzkonferenz-online.de/media/dskb/2022_24_11_festlegung_MS365_zusammenfassung.pdf

CCS, *ACM CHI*, *IEEE SSP*, *PoPETs*, *USENIX Security Symposium*, and *USENIX SOUPS*.

Some studies investigate how specific communities perceive and manage privacy and security issues. For example, some of them analyze how elderly people communities collaborate to cope with these issues (e.g. [16]), others investigate the privacy perceptions of bystanders (e.g. [17], [18]) or users [19] of smart homes or smart speakers [20], [21]. Finally, some studies report how software developers shape their responses to the security and privacy requirements they receive in the development process (e.g. [22]).

Qualitative studies on organizational issues are uncommon. Kokulu et al. [23] analyze interviews with security workers on organizational issues in Security Operation Centers (SOCs). As another example, Hielscher et al. [24] studied how the Chief Information Security Officers (CISOs) perceive Human-Centered Security (HCS). A further example is related to investigating the role of regular security meetings of the software development team for achieving continuous software security [25].

More recently, several ethnographic studies exploring privacy issues were carried out. Chattopadhyay et al. [26] examine at which extent and how cognitive biases affect software development, and the practices and tools that developers adopt to cope with such biases. In their research, Palombo et al. [27] focus on secure software development processes in a software company, whereas Dalela et al. [28] conducted an ethnographic study of security and privacy practices in Danish companies, highlighting the outcomes differences between SMEs and large companies.

To our knowledge, however, there is only a few works targeting privacy in educational settings. Most authors address, indeed, the privacy implications of specific educational technologies (EdTech) at schools, such as WiFi service delivery (e.g. [29], [30]), VPN networks [31], proctoring systems [32], remote educational technologies [33], online conferencing platforms [34], and clouds [35]. Chanenson et al. [9] survey school staff to investigate which risks EdTech can produce for students' privacy and security. On the same note, some technical contributions were conceived to alleviate students' privacy risks in EdTech. Hasan and Fritz [36] propose an approach exploiting feature selection and adversarial censoring techniques to build privacy-preserving versions of learning analytics datasets. Hasan [37] implements a pipeline to automatically detect posts related to EdTech's privacy and security issues.

Other studies focus on how teaching staff approach to privacy and security issues. Mayer et al. [38] qualitatively study the use of password managers. Tu et al. [39] performed a series of telephone phishing experiments on university staff and faculty to explore why scam works and how to defend against it. Kumar et al. [40] investigate how educators take into account privacy and security analyzing their curricular and classroom management goals. In parallel, other works target the final stakeholders of education institutions, that is students and their parents. Zhong et al. [41] study the parents' awareness of privacy and security risks at schools, whereas Balash et al. [42] investigate students' concerns about sharing their data with third parties (proctoring platforms). McDonald et al. [43] focus on how librarians interpret or define their own privacy rules to protect users.

Our multi-site study aims at filling the current lack of research on the assessment of the implementation of organizational procedures on the field.

4. Study Design and Data Collection

The methodology of our field study takes inspiration by the previous work described in [44], [45] and [46], and grounds on the simultaneous usage of field observations and self-assessments of privacy attitude of educational staffs in their daily activities. This dual approach was chosen to tackle the *privacy paradox* stating that people's attitude and behavior concerning privacy can be inconsistent [47]. Field observations were analyzed using a thematic analysis [48] and a grounded theory methodology [49] following recommendations suggested in the Appendix A of [50] to generate concepts potentially transferable to other scenarios.

The key steps of our workflow are inspired by the work of Burrows et al. [51].

4.1. Ethical Approval

The research goals and the procedures for recruiting participants and data protection and storage were submitted to and approved by the Ethics Institutional Review Board (IRB) of the University of Trento.

The informed consents signed by participants were retained by the main researcher and secured in a safe. Data collected through the preliminary survey were anonymous, whereas data collected through observations were pseudonymized to avoid potential identification. For research integrity purposes, the tables with the original paper observations that could enable re-identification were securely stored by the principal investigator of the study into a physical safe (hence our artifact availability statement that to see the raw data you have to travel to our site).

4.2. Setting and sample

The schools involved in the research were chosen using a purposive sampling [52], leveraging researchers' direct connections with principals. The target was three Italian EIs including kindergarten, primary, junior secondary and secondary schools in two regions. The researchers contacted the EIs' principals to illustrate the research goals. After obtaining the institute's formal endorsement of the project, research began with the recruitment of participants.

We recruited participants among the EIs' staff, that is teaching / administrative / technical / auxiliary personnel as well as service directors, data protection officers and physicians. Students attending the schools and their parents were not included in the study. The researcher responsible for field observation met the candidate participants in a preliminary meeting organized at schools by the principals.

During such a meeting the research goals were clearly explained. Sometimes the meeting followed a data protection course taught the school's staff by the researcher.

To protect the privacy of the participants, principals did not attend the meeting and did not have access to information about who did or did not join the research. At the conclusion of the meeting, participants who wished to contribute to the study were given the opportunity to provide written consent. They were notified that they could withdraw from the research at any time. A few weeks later, we began our field observations for people who provided consent. Section 5 shows the case study's details.

4.3. Preliminary Survey

Some days after the preliminary meeting, participants received an email from the school office containing a QR code and a link to access an anonymous online survey aimed at determining their understanding of the institution's practices on personal data protection. To fill up the survey, participants must read the research information sheet, the privacy policies, and express their consent by ticking the appropriate check box in the welcome page of the survey. The survey consisted of 22 multiple choice questions covering key data protection issues, such as: the role and function of data protection officer (DPO), the privacy notice, the data protection policy, the processing register, the data protection impact analysis (DPIA), and so on.

The questions were elaborated during several meetings with researchers expert on privacy that already carried out similar studies. All questions were written in Italian, the native language of participants, and formulated as much as possible with a neutral valence to avoid to elicit negative feelings. The survey also included some demographic questions to determine the size and type of schools the staff works in, their role, and their years of experience.

Responses to the preliminary survey were analyzed in an aggregate form with the aim of: (i) verifying whether the declared school staff's knowledge on the privacy organizational structure reflects the actual situation (e.g., staff state that the school has appointed a DPO, we verified whether really there is a DPO at the school); and (ii) making sure that at large, staff members were aware of privacy policies and the need of compliance even if they where not aware of where the lasted form was. The data on the survey is available on Zenodo (See Section 1.1)

4.4. Field Observations

Procedure. A researcher with more than 10 years of experience in data protection carried out field observations and reported events at each EIin two distinct periods. As a design choice, half the time was spent in classrooms/laboratories, and another half in administrative offices. Behaviors were observed at different timings to cover, on a sample basis, the total work activity carried out by institutes' staffs throughout their daily and weekly duty hours. The researcher decided on the field the number of observations to make according to what happened in the different areas. More specifically, the researcher followed the personal data paths discovered in the field by choosing where to observe and what to observe in real-time. In their observation activities, the researcher moved independently between the areas. According to the grounding theory, there was no a priori decision before entering the field about what to look for (e.g., vulnerabilities and how to classify them). The researcher repeated the field observation procedure until data saturation was reached.

- In the first period, the observations covered data processing in administrative offices, such as the head office, the teaching secretary's office, and the administrative secretary's office.
- In the second period, observations concerned data processing carried out by teachers in the context of activities in classrooms and laboratories. If the school principal allows the observation of a small sample of teaching staff, the second observation period may be shorter, as fewer hours may be needed to reach data saturation.

Annotation schema. An annotation schema was defined following suggestions in Corbin and Strauss [49], Hoegl et al. [53], and Anderson et al. [54] that provide guidelines on assessing the quality of group work, and define a system of behavioral markers usable in the reference domain.

To describe how personal data travels and transforms all along the activities, the annotation schema reported the location where the observation took place (e.g. classrooms), the people present at that location (e.g. staff, collaborators, third parties), the ongoing activities and their pace, the relevant incidents, the interactions between people, and all information that the researcher in charge of the procedure considered relevant at that time. Annotations were performed as much as possible in real-time in a written form without recording or videotaping tools. A fine tuning of annotations was carried out right after each field observations to add contextual data useful for the research goals and further details [44]. Information acquired on the field study was kept confidential, and the observed dynamics did not impact staff members' performance evaluation. As reported on the experimental protocol, no data of student, their parents or members of the school's staff who did not provide their consent was collected.

5. Case Study

Three EIs located in different regions of Italy participated in the research. The sample study consisted of one kindergarten, ten primary schools, two junior secondary schools, and two secondary schools. They are divided into 17sites. corresponding to a total of 355 teachers, more than 57 people of technical and administrative staff, 1918 students over 112 classes (excluding the number of kindergarten's pupils at EI1 in Table 1). Table 1 summarizes the figures of each EI (we did not include janitors of EI 2 and EI 3 in Table 1 as well as the three principals).

TABLE 1. FIGURES OF THE EIS INVOLVED IN THE STUDY

Number of	EI1	EI2	EI3	Total
Students	763	501	654	1918
Classes	41	30	41	112
Teachers	159	73	123	355
Other staff members	39	7(*)	11(*)	57

TABLE 2. OBSERVATIONS (%) MADE AT EACH EI

Percentage values refer to the number of visited rooms and they are rounded to the nearest whole number. The *faculty rooms* wording includes the faculty room, server room, interview room, and lecture hall. The *offices* are the administrative ones.

EI	Class-rooms	Computer labs	Other labs	Faculty rooms	Offices	Total
EI1	100%	100%	100%	100%	100%	100%
EI2	100%	100%	100%	100%	100%	100%
EI3 (all sites)	55%	80%	10%	100%	100%	80%
EI3 (main site)	92%	100%	20%	100%	100%	100%

All teaching and administrative staff's members participating in the study were asked to fill in the preliminary survey. We collected answers from 112 members. Only 98 surveys were complete and retained for the analysis.

Field observations lasted in total 81 hours (30h at EI1, 29h at EI2, and 22h at EI3). The annotation took 29 days (12 days at EI1, 10 days at EI2, and 7 days at EI3. Overall, 125 staff members were observed (68 at EI1, 38 at EI2, and 19 at EI3). Observations were carried out in administrative offices for 54.3% of the time and classrooms/laboratories for the remaining 45.7%. Table 2 shows the percentages of annotations made on the premises of each EIs. In two out of three cases, the percentage of observation coverage is 100%. In the third case, this is 50% overall but rises to 76.19% when considering the primary site of the EI, where most of the annotations were made. In addition, when analyzing the detail of the type of rooms observed in the primary site, the percentage of coverage is almost total for the main types (91.67% for classrooms, 100% for computer labs, and 100% for the other significant rooms). It is reduced to 20% for the other labs with computer equipment for teachers. The researcher chose the number of observations to make based on the dynamics in the field and the need to observe different circumstances (e.g. during class or recess). If a did not give the consent to participate in the research, observations were made only after the lectures. Multiple observations were made in all administrative offices because multiple staff's members were in each room. In rare cases where a person did not consent to participate, the researcher did not observe them, focusing on the other staff members.

Thus, the number of observations' sessions made in classrooms or laboratories ranged from 1 to 3; that one of observations made in the faculty and server room varied from 2 to 7 and from 4 to 5, respectively. The number of observations in the administrative offices was higher because as the most significant flow of data was there, the researcher following the data path spent more time there.The Zenodo repository shows the codes detected day by day at the EIs.

6. Data Pre-processing and Analysis

6.1. Pseudonymization

The following elements of annotations were pseudonymized: (i) the name of the EI; (ii) the date of observation; (iii) the place where the observation took place; and (iv) the first and last names of the EI's staff involved in the study. To further reduce the re-identifiability of the EI's staff members, gender references were removed. The correlation tables with the pseudonyms were stored in a paper format in a safe and they will be permanently destroyed at the end of the storage period specified in the experimental protocol. Analysis was carried out on pseudonymized data to minimize the risks of linkage attacks on data privacy [55].

6.2. Unitizing and filtering

Unitizing is the process to segment data stream in meaningful data units, here identified by *applying a semantic criterion* of following the personal data paths observed in the field. A boundary of a unit is set when:

- the interactions between the EI staff's members processing data end;
- the link between the data and the EI staff members involved in data processing is broken;
- data is safely stored (or disclosed);
- the action(s) on data are concluded.

Thus, the units can have different temporal length. We manually inspected the segmented units to filter off those not relevant for our research.

6.3. Coding

Starting from the annotations, a set of 21 unique codes was defined after several discussion among the researchers to summarize the paths of personal data observed in the field. The researchers discussed the annotations using the Atlas.ti software and arbitrate their conflicts as recommended by Guest et al. [48]. In the rest of the section a *Quote Identifier* is denoted by QID.

The codes are centered on the notion of data processing, they describe the *types of data processed; people accessing and manipulating it; how people interact around it; different data processing; and the types of failure in data processing*. The rationale behind the definition of such codes is to concatenate them into sequences that end into a failure.

People and Data. We identified three *types of data*: physical data (*PH*), digital data (*DI*), and audio data (*AU*). Understanding what is physical or digital data is trivial. Audio data can be the transient content heard by a person during a discussion or walk on a corridor; a WhatsApp voice message, instead, belongs to the family of digital data. The codebook defines four different *types of people accessing and manipulating data*: the teachers (*T*), the administrative

staff (*A*), the head staff (*H*), and third parties (e.g. students, parents, other people alien to the educational institutions) to whom data is exposed in some way (*TP*).

Data Processing. Creating a new data is marked with the code *C*, the pairs of codes *C-DI C-PH* indicating the creation of a digital document (e.g. a file) and a physical document (e.g. a paper form), respectively. Consulting a physical or digital document is identified by the code *Q*. During processing data can undergo digital (*2D*) as well as physical (*2P*) transformations. Summarizing, we can have: i) physical-to-digital transformation (*PH-2D-DI*), ii) duplication of digital data (*DI-2D*), iii) digital-to-physical transformation (*DI-2P-PH*), and iv) duplication of physical data (*PH-2P*). Storing physical data is marked as *SP*. We did not use any specific code for the storing of digital data due to the difficulty to determine precisely, just through observations, when it is definitely stored. Transferring data between people is another typical processing. *DTR* is used when there is one or more senders and one or more receivers, and at the end of transfer the sender(s) cannot longer act on data; *DTT* is adopted when the senders(s) can continue to process data. Finally, *M* is employed to report when people stops to process data (e.g. going away). If nobody is present in the observation area but something relevant happens, the unit is coded from the perspective of actions done or not done by the person (e.g., not locking a computer or leaving a document on the table). For instance, the unit *"There is an unlocked computer in the room"* will be interpreted and coded as *"The staff' member left data unattended by not locking the computer."*.

Interactions around data. Mid-way the coding process, a consistent pattern emerged. Failures tended to happen in presence of interactions (or failed interactions) among people and data. We identified thus four types of interactions: individual actions (*IA*); interaction (*I*); indirect interactions (*II*), and missing interactions (*MI*). The individual actions consist of actions performed on data by a single person; interactions are deliberate interactions among two or more people around data; indirect interactions occur when data processing is done not voluntarily in presence of third parties; missing interactions are interactions during which some of the people involved in data processing do not behave as they should (e.g. going away and so on).

Failures The last two codes of the codebook are about lack of compliance (*F*) or lack of resources (*F-R*), respectively. A compliance failure occurs when something within the processing procedure does not work as expected, so an error or failure happens. Such kind of failure is about actions or omissions of people involved in personal data management acting in a non-compliant way with respect to the law or the organization's rules. Conversely, a resource failure occurs if a procedure or process fails due to a lack of resources. In this case, the failure is under the organization's responsibility, because it is precisely by complying with the procedures that people commit failures. Table 3 illustrates the types of failures. An example of compliance failure is a teacher lecturing by connecting their personal laptop to the school network; or a staff's member leaving the server room unmanned without locking the door. An example of resource failure is a janitor asked to perform simultaneously two tasks such as operating at the reception desk, and going in a classroom to drop off a document. Another example is not having rooms for private communications with parents, so that teachers share the available rooms with each other.

Sequences of Actions Eventually, observations of sequences of actions are coded that ends into a failure. They make it then possible to understand what caused, or at least preceded a failure. Table 4 contains the text of the quotes.

Similar units can be coded in a different way according to the content and the examples. Table 4 reports some annotations and the corresponding strings of codes about staff's members leaving a room without locking a computer. Although all the units share some codes (e.g., A (administrative staff), DI (digital data), and F (compliance failure)), some specific codes were used to specify the context in which observations about data were performed. Thus, for example, the TP (third parties) code was used if (i) third parties were present in the staffed room (Table 4, QIDs: 129 and 982), or (ii) the room was unmanned (Table 4, QIDs: 402 and 1557), with the possibility of third parties entering undisturbed, as our investigator did.

6.4. Privacy incidents Categories

A careful inspection of the units allowed us to define three privacy risk categories resulting by the combination of their impact on people and their likelihood to be detected. In particular, we focused on the technical means and the chances of attackers being discovered along the methodology integrating safety and security by Eurocontrol. The risk of an incident is determined according not only to the type of information disclosed, but also the location, and who can access information.

High Risk incidents happen if data can be directly used to impact the well being of a pupil / teacher without particular technical means and with low chances of being discovered while accessing the information; *Medium Risk* is present when data can be used for discrimination against a pupil / teacher or can be used as an escalating vector to acquire additional information by parties without major skills that would have otherwise high chance of being detected while accessing the information; Finally *Low Risk* incidents happen when data can only be used by parties who already know the information or can easily acquire it. Significant technical skills would be required to further escalate it, and the chances of being detected while accessing it would be extremely high. Each unit reporting an incident was marked with the code *High*, *Medium*, or *Low*.

The causes of privacy incidents are analyzed using a 2x2 matrix, which correlates the cause with the type of lack in data processing. Table 5 illustrates the causes of incidents. It distinguishes the causes into structural, if a change in organizational processes can directly mitigate them, and occasional if they are both due to individual staff behavior and can be mitigated by increased behavior compliance

TABLE 3. CODEBOOK: TYPES OF FAILURE IN PROCESSING

Code	Long code	Description	When to use it	When not to use it
F	Failure	Something within the processing procedure does not work as expected and generates an error or failure (lack of compliance).	A printer does not print a document because of a paper jam. A person cannot access an online service because of an application error. A person leaves a room without locking the computer's operating system. A person exits a room, leaving open documents on the desk. A person cannot consult a document because of an error generated by the document management system. Two+ people discuss processing in a public place like a hallway. A person leaves a document unattended in the printer.	A person exits a room and leaves open documents on the desk because another person is processing them. An operator leaves a document on their desk in an access-controlled room where only another operator authorized to process the data in the document is present.
R	Resources failure	A procedure or process fails due to a lack of resources. Used also when two or more procedures have conflicting requirements that fail. **Jointly use:** This code must be used with the *F* code, as it specifies a particular failure circumstance.	In a room, an unauthorized person can listen to talks about data because of the physical layout of the premises and compliance with the organization's procedures. A single janitor has to be at the reception desk and simultaneously to perform activities elsewhere. An unauthorized person can hear the conversation in another room because of the unsound-proofed walls.	Because of non-compliance with the organization's procedures in a room, there are several people inside, not all authorized, who can listen to talks about data.

TABLE 4. EXCERPT OF QUOTES' TEXT ON IT PHYSICAL SECURITY, BULLETIN BOARDS, AND STAFF HABITS

QID	Topic	Annotations	Coding
794	Doors left open	The door to the server room and the rack cabinets inside (containing the institute's server and network equipment) were not locked. It happened that someone entered the room without anyone noticing the presence.	A-TP-IA-DI-M-F
1589	Doors left open	A teacher consulted with parents in one classroom by holding the door open. While pausing in the hallway, someone could hear what the parents and the teacher were saying to each other.	T-TP-I-II-AU-DTT-F
1596	Doors left open	Before moving to another premise, janitor failed to lock the guardhouse door with video surveillance monitors.	A-TP-IA-II-DI-M-F
234	Unmanned class	During the lunch break, classroom were all open and deserted.	T-TP-IA-MI-DI-PH-M-F-R
236	Unlocked PCs	Unlocked PCs and documents opened on the screen.	T-TP-IA-II-DI-M-F
880	Unlocked PCs	The computer in faculty room 2 is always turned on and unlocked with the Gmail screen open.	T-TP-IA-II-DI-M-F
1597	Unlocked PCs	The laptop computer in the classroom is turned on and unlocked. A teacher does not man the classroom.	T-TP-IA-II-DI-M-F
68	Staff Habits	Post-its with e-mail addresses or phone numbers (e.g., from technical support or suppliers) present on monitors.	A-IA-PH-SP-F
962	Bulletin board	Inside the classroom, the children's names are written on a document on the door, along with their ways of leaving school and transportation (e.g., alone on foot or by school bus).	T-IA-PH
970	Bulletin board	Behind the faculty room door, a sheet containing all pupils' "sensitive" data is hanging. The data include whether they do or do not attend religious classes, their first and last name, the class attended, whether they eat in the cafeteria, and whether they use bus transportation. The room is accessible by third parties, and data are exposed.	T-TP-IA-II-PH-F-R
946	Bulletin board	The annual duties sheet with the children's names is on the classroom wall.	T-IA-PH
908	Bulletin Board	On the laboratory wall bulletin boards are sheets with the computers' accounts and passwords.	A-TP-IA-II-DI-PH-M-F
19	Staff Habits	OP02 exits the room, leaving documents on the desk and the computer unlocked, showing some content on the screen. [...] Only OP01 remained in the room. [...] Data is exposed for 4 minutes. [...] No third parties are present in the room.	A-IA-II-DI-PH-M-F
91	Staff Habits	OP02 exits the manned room, leaving the computer unlocked and the monitor turned on. [...] Data is exposed for 8 minutes. No third parties are present in the room [...].	A-IA-II-DI-M-F
129	Staff Habits	OP04's work is often interrupted by parental visits [...]. OP04 performs many tasks quickly and often has to get up and leave; in doing so, they leave the computer unlocked and the monitor turned on. During their absence, however, the room remained staffed.	A-TP-IA-II-DI-M-F
982	Staff Habits	OP03's computer is turned on and unlocked. There is an application (a spreadsheet) opened on the screen from which I can read data. The workstation remained unattended even while OP03 is not in the room, which is never left empty. [...] Some third parties routinely enter the room to perform certain activities.	A-TP-IA-II-DI-M-F
402	Staff Habits	OP01 exits the room, leaving the door and the computer unlocked (with applications opened).The room is unmanned. Data is exposed. [...] No third parties are in the room.	A-TP-IA-II-DI-M-F
1557	Staff Habits	In Lab 01 [...], a technical assistant's computer is turned on and unlocked. The lab has been unattended for a long time.	A-TP-IA-II-DI-M-F

with organizational process requirements. Processing failures are compliance failures in the absence of a procedure, in the presence of a poorly designed procedure that does not comply with regulations, or even in the presence of a well-designed procedure not executed by staff. In contrast, resource failures are characterized by the existence of a procedure or practice that is typically well designed on paper but fails due to insufficient resources for its completion.

7. Results

7.1. Staff's knowledge about data protection

A DPO was appointed at each EI and only some member of EI1 and EI2 were either unaware of its existence at school (23.5%). In EI3, which performed a specific training, most participants stated that the schools have a DPO and that they know how to contact them (83.3%). During the observation periods, DPOs visited the schools and met staff's members QID775: "*In the room of Manager A, HED2 and MT02*

TABLE 5. MATRIX OF THE CAUSES OF PRIVACY INCIDENTS IN DATA PROTECTION

Cause	Structural cause	Occasional cause
Lack of compliance	Either there is no business procedure or a procedure (or an established practice) that, being poorly designed, fails to comply with current regulations. The organizational process has regulatory non-conformities that cause the processing to fail precisely because the staff complies with the procedure's requirements.	The staff could perform a well-designed procedure, but it is not. Individual staff behavior may be determined by i) particular circumstances and ii) autonomous and personal adoption of practices not authorized by the educational institution.
Lack of resources	A procedure (or established practice) that is poorly designed fails because there are insufficient resources to complete it. The organizational process has conflicting requirements that cause the processing to fail precisely because the staff meets the procedure requirements.	Although well designed, there is a procedure that fails because of insufficient resources to complete it. Individual staff behavior tries to compensate for the lack of resources by deviations: i) with practices not authorized by the educational institution; ii) with specific actions determined by particular circumstances.

meet with the DPO to take stock of the situation, discuss issues and acquire opinions"). All EIs have a corporate data protection policy, and only 19.3% of staff members do not know its content.

The major knowledge gap for rank and file members of staff arise in the *specifics of the GDPR*: all EIs have a register of processing activities but 54.1% of participants do not know what this document is and 16.3% knows what it is but think the school does not have it. Similarly, staff's member ignore the meaning of data protection impact assessment (66.3%), and formal procedure for data subjects' rights management (60.2%). We believe this is not only understandable but even legitimate. As a rank-and-file staff you are supposed to know what a policy is and follow it, but not necessarily how it is stored.

7.2. Physical secure IT equipment in edu spaces

The researcher made 150 observations in classrooms and laboratories in which at least one there was a computer: 84 of 150 observations (56%) were unique. In 11 out of 150 cases (7.3%), it was not possible to detect whether the operating system was locked or not, while in 37 cases (24.6%), the computer was turned off. In the remaining 102 cases (68%), the computer was turned on. More specifically, in 38 out of 102 cases (37.2%), the computer was locked or in use by a staff's member, while in 64 out of 102 cases (62.8%), it was unlocked and not in use. Analyzing in deep these last 64 cases, we found that the room was empty in 22 occurrences (34.4%), whereas a member of the staff was in the room in the other ones.

We found laptops in 57 out of the 150 annotations concerning rooms with electronic equipment inside: 35 (61.4%) were without an anti-theft cable, while 22 (38.6%) had an anti-theft cable. All the laptops equipped with anti-theft cables were at the same EI. Checking for a double incident is helpful to assess the real risk level for data breach, so that we analyze for example how many times staff's members left the room unlocked with a laptop inside without an anti-theft cable. We found that the room was empty in 19 cases (54.3%), while the room remained handled by a staff member in 16 cases (45.7%). When laptops were without the anti-theft cable, we found that most occurrences were recorded in classrooms or laboratories (17 out of 19 of the empty room cases) and at the specific times of the day detected for the case of the unlocked computer. Hence, the mobility of the staff's member across classrooms and the possibility to access shared computers in such locations is a critical factor for privacy incidents involving digital data leakage. Incidents are almost all related to doors left open (QIDs 794, 1589, and 1596), unmanned classrooms (QIDs 234), or the presence of unlocked computers (QIDs 236, 880, and 1597).

Moreover, we encountered further incidents related to the failure to close doors in the server room (QID 794),in classrooms during parent-teacher meeting (QID 1589),and in rooms where there are video surveillance system monitors (QID 1596).We detected examples of unmanned classrooms during lunch break (QID 234),or when teachers did not use the classroom for classes. Finally, we detected situations in which a computer was turned on, but not in use and without a lock screen both in the faculty room (QID 880) and classrooms (QID 1597 for portable systems or QID 236 for the fixed ones). The absence of an anti-theft cable connected to the notebooks is a separate issue since the EI's specific organizational choice determines it. If two or more of these incidents co-occur, a multiplicative effect significantly increases the effects of individual incidents. If, for example, one considers an unmanned room and a computer unlocked, each incident amplifies the consequences of the other.

7.3. Physical secure IT equipment in admin spaces

In assessing the physical security of computer equipment used by administrative staffs, we did not observe any cases of laptops not being connected to anti-theft cables. At each EI, we observed a number of staff members varying from two to six, making the case of computers being left unlocked in unsupervised rooms marginal. We detected, however, a few cases of such an incident in each EIespecially when people were on duty according to a roster (e.g., on Saturdays or specific time slots). Overall, we detected 74 occurrences of computers left unattended and unlocked. Table 6 displays the count of individuals engaged in data processing and the instances of unlocked computers detected at each EI. An

TABLE 6. UNLOCKED COMPUTERS IN THE ADMINISTRATIVE OFFICES

Institution	No. People	No. Cases	% Cases	% People
EI1	7	17	22,9%	28%
EI2	6	9	12,2%	24%
EI3	12	48	64,9%	48%
Total	25	74	100%	100%

increase in the number of people involved in the processing often results in an increased number of occurrences detected.

Since the number of people observed is limited, the effects of individual behaviors are significant. For example, at EI3, the person who left the computer unlocked the most times determined 33.3% of the occurrences. Similarly, at the same institution, the number of occurrences determined by the first four person who do not lock their computers is 70.8% of all occurrences. By structuring the annotations on a sample basis to cover the entire work week and hourly schedule, we can limit distorting effects related to specific staff activities on a particular day.

7.4. Folklore and staff habits

During the field study, some cases deserving a specific discussion occurred. Although these cases are part of the folklore and staff habits, they are not interesting from the perspective of engineering good organizational privacy practices. In our qualitative analysis of the results, we report only major privacy incidents that can be ranked as high risk privacy incidents (see 6-6.3).

There are obviously cases that result from the habits of a specific staff member. For example , a person left an ID card unattended on the desk while the room was empty. In another example , a person used to leave post-it notes containing personal data stuck on their desk or monitor. Further, some staff's member used EI's IT resources for personal purposes, for example, participating in a video conference using a school'scomputer outside working hours.

Our qualitative results highlight the most critical implication for small organizations: the ubiquitous lack of resources. In this respect, the most critical instances to assess a possible incident and corresponding structural mitigations are those in which *the staff member moves on*, leaving the room unattended, fails to lock it, and leaves an unlocked computer inside or some physical private data.

7.5. Not all bullettin boards are equal

Let's consider a common practice at schools: posting information on bulletin boards, walls or classroom's door. The kind and the nature of such posted information broadly change and, sometimes, personal information are disclosed causing high risk privacy incidents. Table 7 shows some examples of field observations that display privacy incidents.

Consider row (QID 962) and (QID 970b). The information is essentially identical. It is a sheet reporting how children of that class go home (e.g. alone, by bus and so on). Knowing how children go home is definitely a sensitive information exposing children to several risks, such as abduction by a malicious party. In (QID 962) the information disclosure, however, happens in a location accessible only to pupils and school staff of the class who *already* know the information and in the case of the teacher have a *need to know* to make sure that a pupils should be accompanied to the bus. For a replacement teacher this is information that can be quickly accessed while escorting the pupils out of the class. For that reason, the incident is marked as low risk. An entirely different set-up is the posting in (QID 970). The particular faculty room is open to third parties, in particular parents of pupils wishing to talk to the teacher(s). However, the information extend to all classes, so a parent might get information about children of classes different from the classes of ones' own. Knowing whether a child attend (or do not) religious classes can be used for discrimination but this is at a difference level than knowing that the child is authorized to walk home alone.

Knowing a teacher's first and last name and hearing or committee attendance plan is not a major privacy breach because it concerns their teaching role (see QID 973). One may argue that such information should be public.

Finally, another incident consists in posting on the lab bulletin board personal login credentials to lab computers (QID 908). The traditional IT approach would scream for severe security incidents. We instead marked it as covering the entire spectrum. The key issue is what can be done on these computers: if the room is normally locked, there is no unregulated access to the Internet, and students can only access it to run education programs, the biggest risk is that they would copy each other's assignment. While this might be an academic violation, it is *not* a privacy incident. If the computers allow access to the personal emails of the students as provided the school this might classify as a medium incident as some official communication between the school and a student might be captured by other students logging as him or her. Finally, this might be classified as high risk if the computers allow unfettered access to the Internet as one student could impersonate another student with potentially severe consequences.

7.6. Staff incorrect use of tools

Some incidents involved the security requirements induced by the GDPR concerns how people use tools such as computers possibly containing sensitive information.

For example, having an unlocked computer (or not connected to an anti-theft cable) in an unattended room may become a more critical issue during lunch breaks or in the interval between two lectures. It's extremely unlikely that an attacker would enter a room with a staff member present, sit at a computer that isn't theirs, and start accessing data without being questioned about their were doing.

TABLE 7. EXAMPLES OF HIGH RISK/LOW RISK INCIDENTS OBSERVED IN THE FIELD.

ID	Room	Where	Type of data	Visible by 3rd Parties	Risk
962	Classroom	Inside the classroom	The children's names and means of transport for leaving school (e.g., alone on foot, by bus).	No	Low
970a	Faculty room	Paper behind the door	Children's personal data (e.g. first and last names, whether they take religion hours or not).	Yes	Medium
970b	Faculty room	Paper behind the door	Children's first and last names, the class they attend, whether they eat in the canteen, and means of transport for leaving school	Yes	High
946	Classroom	Paper on walls	Children's names.	No	Low
973	Faculty room	On the bulletin board	Documents reporting teachers' first and last names, such as the substitution sheet, hearing sheet, and committee attendance sheet.	Yes	Medium
908	Laboratory	Paper on walls	Computers' accounts and passwords.	Yes	investigate (?)

7.7. I'm called elsewhere effect

Sometimes a staff's member is solicited with requests for activities to be carried out simultaneously. This situations force the staff's member to interrupt a task in progress that then remains unfinished, subsequently leading to a failure. For example, a staff's member who was filing folders in a locked cabinet was forced to rush out of the room, resulting in forgetting the open folders on their desk. Due to an inadequate service coverage, the same staff's member often faced with conflicting tasks simultaneously. Another example is the following: the janitor's office is empty because the janitor is busy on another floor where nobody else is available. In five cases, we observed the faculty room with unlocked computers. Such observations were randomly made during the morning while teachers lectured in classrooms. In further 17 cases, we observed this dual circumstance in classrooms and laboratories, but at specific times: during school recess, lunch break, time and teacher change, or when the classroom/laboratory was not used for classes. We observed that people in secretarial offices have the tendency to leave their computers unlocked when they have to be absent from their workstations. The most common reasons attributable to this behavior are (i) going to get print out from the network printer (which is almost always outside the room); (ii) going to make photocopies; (iii) going to deliver physical documents to someone; (iv) attending in-person meetings. Some people were observed turning off the monitor instead of locking their computer, leaving it unlocked. Interestingly, this behavior was detected among operators from the same office but only at EI1.

7.8. Resources don't live up to expectations

We found mistakes in the organizational procedures adopted by the analyzed EIs due to lack of physical resources. For example, dialogues with parents were held in places or premises unsuitable to protect the content of confidential conversations between teachers and parents. An unsuitable place is the hallway, where *"it is possible to hear and understand what some teachers (who do the dialogues down the hallway) say to parents"* (QID 1590). Another example is that one of a room with poorly insulated walls where *"it is possible to hear clearly and loudly everything that is being said in the room next door [...] where a teacher is talking to parents about educational evaluations given to some students"* (QID 467).

Lack of economic resources also encourages faulty design of procedures which are amplified by the move-on effect. For example, we observed the malfunction of an automatic locking mechanism of a main entrance door, which could not be repaired due to lack of resources. To remedy the problem, the institution designed a procedure requiring a janitor to manually close the door after visitors entered. Unfortunately, we observed one case in which the janitor could not provide the door closure because he was busy elsewhere, resulting in the burden of closure being entrusted to third parties visiting the institution.

7.9. Who watches the watchers?

Regarding video surveillance monitoring systems, we found them in EI1 and EI2, and we identified four types of problems: (i) unauthorized third parties who may watch video surveillance system monitors; (ii) erroneous definition of access criteria for monitoring video surveillance systems; (iii) failure to guard the room where there is a video surveillance monitoring system; (iv) failure to lock the rack of the video surveillance system.

A problem we detected concerns a staff's member leaving the room where the monitoring system is located without locking the door (coded using the string *A-IA-M*). This behavior could produce a possible indirect interaction with video surveillance data (coded as *TP-II-DI*). Such kind of failure is marked as a compliance failure (code *F*).

Another example is about the placement of the monitoring system in a staffed location (e.g., an administrative office) accessible to third parties. As in the previous circumstances, a possible indirect interaction could occur. In this case, however, the failure relates to a need for more resources (lack of suitable rooms) that prevented the proper design of the business procedure (codes *F-R*). In another case, unauthorized individuals (e.g. janitors) monitored the video surveillance system (coded as A-IA-Q-DTT-DI). Thus, a transient digital data transfer occurred due to a lack of human resources caused by an incorrect business procedure design (codes F-R).

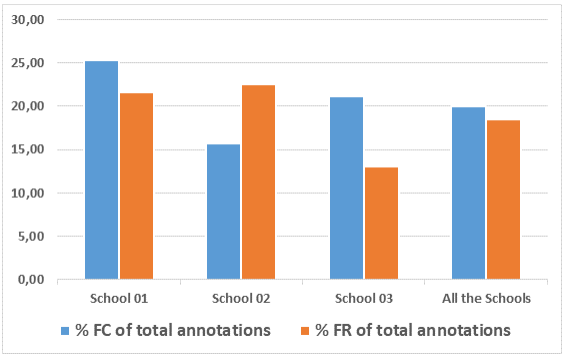

Figure 2. Number of compliance or resource failures on all annotations.

We encountered several privacy incidents in which missing interactions occur with people in charge of manning the box where the video surveillance monitoring system is placed. For example, staff's members had to leave because they were engaged elsewhere in other activities (we coded it as A-IA-M to indicate the staff's members leaving and as *A-MI-TP* to indicate the failure to interact with third parties). In such a case, third parties could interact indirectly with data (*TP-II-DI*). Here, the problem lies in the lack of human resources. The procedure, indeed, requires that a staff's member performs two contradictory tasks simultaneously; hence, we a resource failure happens (codes *F-R*).

7.10. The shadow IT

In classrooms and laboratories, we observed many IT solutions introduced by the teaching staff: the shadow IT. Haag et al. [56] define it as an IT entity (hardware, software, or services) that a worker builds, introduces, or uses for the job without prior informing the organization or acquiring its approval. Some authors [57] show how shadow IT is ingrained in the IT environments of higher education institutions and reveal that they can create new attack vectors. Specifically, we found that some teachers used their laptops for lectures by connecting them to the institution's network, acting like 'unmanaged PCs.' For example, a teacher used a personal laptop in the classroom or in the teaching room. According to van Acken et al. [57], self-acquired devices are the more frequently shadowed IT category.

8. Summary of failures

We detected overall 613 failures: 294 (48%) resource failures, and 319 (52%) compliance failures. This result denotes that, in almost half of the annotated cases, a staff's member fail to protect data while acting according to company rules or practices. Figure 2 shows (overall and by EIs) the percentage of units coded as compliance failure and resource failure, respectively. From the pattern, it is evident that about two-fifths of the annotated units of analysis result in a failure (compliance 20%, resource 18.5%). In addition, while resource failures prevail in EI2, compliance failures are the majority in EI1 and EI3. Interestingly, failure of compliance is less prevalent in EI2, where 83.3% of respondents say the institute has a corporate data protection policy. In EI2, a 40% of respondents say they know the content of the policy (in line with data from other institutes), and another 40% say they do not know the policy because the corporate data protection rules are based on practice (which is a peculiarity of this institute). Thus, EI2 is characterized by a very practice-focused organizational structure, evident in interactions with the administrative staff. The scarcity of economic resources also results in high staff turnover. e.g. "*HED1 told that it is difficult to share procedures in the working group because this changes often*" (QID: 419).

We used the codes of Section 6 to understand how to classify the results. The sequences of codes (i.e., events) are the keys to our findings, the threads linking all the anecdotes of Section 7. Figures 3 and 4 show the codes that preceded a failure: event X took place before event Y, or event Z was absent. For instance, we are concerned about unlocked PCs because the teacher doesn't lock the room when leaving (T+IA+DI+M+F sequence).

Our annotations show a predominance of compliance failure in units in which digital data are involved (52.7%), followed by those involving physical data (42.6 %) and transient audio data (21.3%). The first data is probably higher considering the difficulty of determining precisely the actions performed by staff's members on digital data by observing their activities (see Subsection 6-6.3). In the case of resource failures, physical data involvement is slightly predominant (46.6%), followed by digital data (40.1%) and transient audio data (15.6%). However, the latter case is explained by the many annotations determined by inadequate logistical equipment or practices established at the corporate level in storing data in physical form (see Section 5). Similarly, Marjanov et al. [58] pointed out that most security of processing violations (ex Art. 32) of the GDPR involve data in digital format.

8.1. Missing Interactions as a Root Cause

Figure 3 shows the number of failures occurred with missing interactions (*MI*) and/or moving on (*M*). Figure 4 shows the number of failures due to the co-occurrence of a *MI* or a *M* with respect to the total number of failures.

The analysis of the data shown in the figures clearly shows that most privacy incidents were *preceded by* interactions in which people 'left the ball drop midway' (*MI*) or they were interrupted during the processing (*M*). The highest probability of failure was indeed observed when such events co-occurred (93.1%): compliance failures occurred in 16.4% of cases. whereas resources failures in the remaining 76.7%. When *MI* was not followed by *M*, we had 28.6% failures of compliance and 35.7% failures of resources. Conversely, when *M* was not followed by *MI*, we detected 56% failures of compliance and 29.3% failures of resources.

The co-occurrences of M or MI and a resource failure happened with unattended rooms (classrooms, laboratories, faculty room, administrative and janitors office) with open

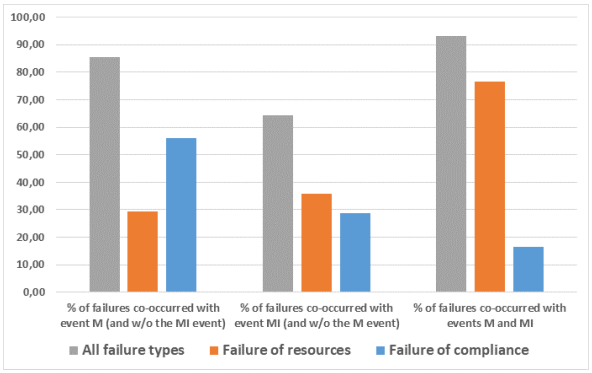

Figure 3. Number of failures related to missing interaction (*MI*) and/or move on (*M*)

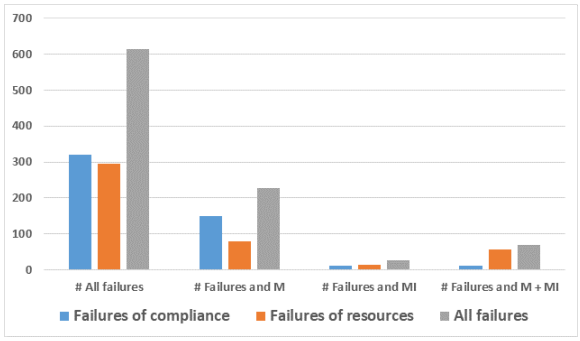

Figure 4. Number of failure co-occurred with *MI* or *M* with respect to the total number of failures

doors due to an established business practice. There are other cases in which there was a lack of manning at the EI's site entrance or in the hallways because janitors were called to perform other duties. Two remarkable cases were observed.An example: the members of a workgroup changed unexpectedly near a regulatory deadline. This change led to mishandling, as it was unclear who was in charge of what (QIDs: (568 - 577). Another example: a corporate server and removable hard drives containing data backup were located in an unattended office with open door. A design error in the processing procedures is evident in both these examples.

The co-occurrences of M or MI and a compliance failure are associated with unattended rooms or open doors due to a personal mistake by staff's members who chose not to comply with school procedures. Examples include the temporary absence of staff from administrative offices or classrooms to get a paper document from the network printer or to deliver it to someone. More critical cases occurred. For example, a staff's member left by mistake unattended and with an open door the room with surveillance system monitors. Mistakes are the sources of such failures.

8.2. Structural Violations

Incorrect design choices of actual security procedures are an incentive factor for the occurrence of security violations. Often the seriousness and erroneousness of some of these design choices are trivial, as in the case of the WiFi access password written on the whiteboard or the list of student access credentials hanging on the lab bulletin board Sometimes the fallacy of a design choice is unclear, such as using a shared account and operating system autologin mechanism with a subsequent application login phase for users. This configuration is particularly critical in laboratories where we found that it is possible to read the e-mails of students who had not logged out of the application.

9. Discussion

Incorrect personal data processing procedures design exposes schools to GDPR noncompliance with consequent fines [58]. Insufficient organizational measures may be enough to be subject to a data protection authority (DPA) fine even without a personal data breach. Analyzing DPAs' fine in the EU Member State [59], [60], [61], Marjanov et al. [58] show that organizational mistakes are the most costly for a non-compliant Data Controller.

Although our results highlight a gap between the school staff's claims to care about privacy and their actual behaviors (i.e. the privacy paradox), already studied by Colnago et al. using an online sample of US participants [62], this is not always the case.

We observed that a lack of resources pushes people to adopt organizational practices aimed at optimizing work that accidentally violates privacy requirements. For example, the governance of the EIis forced to choose a suboptimal logistical arrangement of staff due to a shortage of rooms, consequently triggering inappropriate practices on the part of staff trying to remedy the problem. We observed the presence of multiple persons involved in processing activities in a single publicly accessible room in which they talk about data while third parties being there Such a promiscuity of authorized and unauthorized people can create confidentiality problems that staff's members tackle by implementing behaviors that can result in violations. For example, to talk with a colleague about an urgent and confidential matter, a staff member working in a room with third parties present "*takes a document with him and leaves the room leaving the computer unlocked*" (QID: 1494)." In other cases, people are overwhelmed by the number of tasks they have to perform, and to optimize time and resources, they adopt inappropriate practices. For example, "*while OP05 is making photocopies using the hallway copier, OP02 comes out of secretary room A and brings him signed documents. Then, because OP05 is engaged in another activity, he tells OP02 to place the documents on his desk. So, OP02 enters secretarial room B and does as he was told*" (QIDs: 1512 and 1513). In another case, multiple simultaneous tasks are assigned to the same person, so a colleague helps them by making their scans. "*Then OP05 tells OP01 to log in using his (OP05's) badge so that the scans are directly sent to him*" (QID 1519). In this sense, *there is no privacy paradox*, the one we observed is more of a concrete need to balance privacy with work leading to *accidental violations due to the limited resources available*.

Our analysis reveals a dimensional problem and an organizational difference among the EIs. The resources available for data protection management are sometimes different and their number decrease in educational institutions' secondary sites especially when the secondary sites are in different buildings or located far from the primary sites (e.g. in nearby town or suburb). Thus, sub-optimal data protection practices occur especially in secondary sites .

Different incidents were observed in educational institutions at the main and secondary sites. For example, incidents about the video surveillance system or the institutional server system cannot occur in the secondary sites because such systems are installed / managed at the primary sites only. Incidentss related to the missing identification of unknown people are instead more common in the primary sites. Such sites, indeed, are visited everyday by a large amount of people including third parties.

10. Threats to Validity

We took several methodological expedients to ensure the study's internal validity and data quality. However, our findings should be interpreted with the following limitations.

We conducted this study in Italy where GDPR is enforced. Hence, the results may only be partially replicable in contexts with a different legislation. Some of the findings might be due to particular beliefs and characteristics of the country [63] and might not generalize to other countries. To mitigate this risk, we developed a flexible methodology to try to capture general phenomena occurring in a variety of context [50]. A fine-tuning of the codebook indeed could be needed. We acknowledge that a finite and limited codebook does not allow for a fine-grained description of what occurred, but our goal was to develop a valuable and quickly extendable tool for performing a quick macro-analysis of personal data flow in small privacy scenarios such as educational settings.

Another limitation is that some aspects are not directly observable in the field. For example, it is tough to observe some digital data processing happening when EIs use third-party applications to manage emails, assignments, or parent-teacher communication. As we mentioned, EIs do not have a direct control over this type of software, which sometimes is chosen at a national level. However, it can be interesting to understand the perception of the EI's staff about this third party software. Future work could address this issue using semi-structured interviews with selected staff's members.

The observations were performed by a researcher without using any digital recording instruments. Although audio-video recordings could have provide more information about the flow of personal data, their use would have make more complex data collection and processing. Indeed, either obtaining ethical authorizations and informed consents by all parents whose children *could* have been captured, or editing videos displaying children whose parents have not given consent for filming, would have made the setup of the study particularly time-consuming and challenging. Moreover, the adoption of recording devices would have potentially affect the natural behavior of the staff and raise severe, and justified concerns from parents. The expertise in data protection matter of the researcher who carried out the field study should have limited the bias in the observations.

Audio/video would have also collected data that was privacy protected (the very case of the conversation in the hallway about pupils) without actually contributing to the research goal of identifying *structural* failures.

A limit in the observations is that the experimental protocol approved by the University's IRB asserts that researchers must not record data about parents, students or school staff who do not provide their consent. For example, the researchers will act as if the staff member who did not give consent is not present in the room. So it might well be that there were even more than two people in a room where an incident took place. Only when we write that a room was left empty it was really empty (of observable or unobservable people).

11. Conclusions

Our field study's qualitative results offer valuable insights for governing schools and small organizations, helping them identify and correct mistakes in personal data processing procedures.

The key takeaway of our research is that privacy incidents due to lack of resources when implementing the GDPR in the small are ubiquitous and often insuperable. We argue that it is possible to better protect the personal data with less secure but practically feasible procedures than perfectly secure procedures that will be forcibly poorly implemented. A concrete approach should adopt an implementable paradigm oriented to risk mitigation. This approach may not guarantee a full compliance with all the tiny provisions and legal minutiae of the GDPR. However, in the presence of non-waivable processing, such as those of EIs, compliance with the principles of the law is still ensured in the sense of continuous improvement.

A possible solution to address this concern is staff training: one-third of the surveyed staff would have appreciated more training. However, rather than training them in the terminology of the law, making school staff more aware and attentive to data protection processing principles [24] would allow them to react better to procedures that fail and to avoid designing processes that are great on paper but doomed to fail on the field.

Acknowledgments

We would like to thank the principals of the EIs for trusting and supporting us in performing this study and all members of the EIs for their help and support in this work. This work has been partly supported by the European Union under H2020 grant. 830929 (CyberSec4Europe), HE grant n. 101120393 (Sec4AI4Sec), by the Italian Ministry of University and Research (MUR), under the P.N.R.R. – NextGenerationEU grant n. PE00000014 (SERICS), and by

the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) under grant n. KIC1.VE01.20.004 (HEWSTI).

CRediT Author Statement

Conceptualization: FC, FM, GV; Methodology: FC, FM, GV; Software: FC; Validation: FM, GV;Formal analysis: na; Investigation: FC; Resources: na; Data Curation: FC; Writing - Original Draft: FC; Writing - Review & Editing: GV, FM; Visualization: FC; Supervision: FM, GV; Project administration: FM; Funding acquisition: FM;

References

[1] R. Layton and S. Elaluf-Calderwood, "A social economic analysis of the impact of GDPR on security and privacy practices," in *2019 12th CMI Conference on Cybersecurity and Privacy (CMI)*, 2019, pp. 1–6.

[2] S. Sirur *et al.*, "Are We There Yet? Understanding the Challenges Faced in Complying with the General Data Protection Regulation (GDPR)," in *Proceedings of the 2nd International Workshop on Multimedia Privacy and Security*. ACM, 2018, pp. 88–95.

[3] M. da C Freitas and M. da Silva, "GDPR Compliance in SMEs: There is much to be done," *Journal of Information Systems Engineering & Management*, vol. 3, p. 30, 2018.

[4] N. Casutt and N. Ebert, "Data protection officers: Figureheads of privacy or merely decoration," in *16th European Conference on Management, Leadership and Governance*. Academic Conferences International limited, 2020, p. 39.

[5] F. Ciclosi and F. Massacci, "The data protection officer: A ubiquitous role that no one really knows," *IEEE Security & Privacy*, vol. 21, no. 1, pp. 66–77, 2023.

[6] I. Hadar *et al.*, "Privacy by designers: software developers' privacy mindset," *Empirical Software Engineering*, vol. 23, no. 1, pp. 259–289, 2018.

[7] E. S. Tatet, "Teaching in under-resourced schools: The teach for America example," *Theory Into Practice*, vol. 38, no. 1, pp. 37–45, 1999.

[8] J. L. Styron, "Critical issues facing school principals," *Journal of College Teaching & Learning (TLC)*, vol. 8, pp. 1–10, 2011.

[9] J. Chanenson *et al.*, "Uncovering Privacy and Security Challenges In K-12 Schools", in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* ACM, 2023, pp. 1–28.

[10] K. Schildkamp *et al.*, "How school leaders can build effective data teams: Five building blocks for a new wave of data-informed decision making," *Journal of Educational Change*, vol. 20, pp. 283–325, 2019.

[11] C. K. Hudson and W. Shen, "Understaffing: An under-researched phenomenon," *Organizational Psychology Review*, vol. 5, no. 3, pp. 244–263, 2015.

[12] N. Lever *et al.*, "School Mental Health Is Not Just for Students: Why Teacher and School Staff Wellness Matters," *Report on emotional & behavioral disorders in youth*, vol. 17, no. 1, pp. 6–12, 2017.

[13] F. Massacci *et al.*, "Economic impacts of rules-versus risk-based cybersecurity regulations for critical infrastructure providers," *IEEE Security & Privacy*, vol. 14, no. 3, pp. 52–60, 2016.

[14] C. Sun *et al.*, "GDPRxiv: Establishing the state of the art in GDPR enforcement," *Proc. Priv. Enhancing Technol.*, vol. 2023, pp. 484–499, 10 2023.

[15] Garante per la protezione dei dati personali, "La scuola a prova di privacy - vademecum ed. 2023 (doc. web. 9886884)," 6 2023.

[16] J. Kropczynski *et al.*, "Towards Building Community Collective Efficacy for Managing Digital Privacy and Security within Older Adult Communities," *Proc. ACM Hum.-Comput. Interact.*, vol. 4, 1 2021.

[17] Y. Yao *et al.*, "Privacy Perceptions and Designs of Bystanders in Smart Homes," *Proc. ACM Hum.-Comput. Interact.*, vol. 3, 11 2019.

[18] C. Cobb *et al.*, ""I would have to evaluate their objections": Privacy tensions between smart home device owners and incidental users," *Proc. Priv. Enhancing Technol.*, vol. 2021, pp. 54–75, 2021.

[19] E. Zeng *et al.*, "End user security and privacy concerns with smart homes," *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, 7 2017, pp. 65–80.

[20] J. S. Edu *et al.*, "Smart home personal assistants: A security and privacy review," *ACM Comput. Surv.*, vol. 53, 12 2020.

[21] J. Lau *et al.*, "Alexa, are you listening? privacy perceptions, concerns and privacy-seeking behaviors with smart speakers," *Proc. ACM Hum.-Comput. Interact.*, vol. 2, 11 2018.

[22] T. Lopez *et al.*, "Security responses in software development," *ACM Trans. Softw. Eng. Methodol.*, vol. 32, no. 3, 4 2023.

[23] F. B. Kokulu *et al.*, "Matched and Mismatched SOCs: A Qualitative Study on Security Operations Center Issues," *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2019, pp. 1955–1970,

[24] J. Hielscher *et al.*, "Employees Who Don't Accept the Time Security Takes Are Not Aware Enough: The CISO View of Human-Centred Security," *32nd USENIX Security Symposium*, 8 2023, pp. 2311–2328.

[25] I. A. Tøndel and D. S. Cruzes, "Continuous software security through security prioritisation meetings," *Journal of Systems and Software*, vol. 194, p. 111477, 12 2022.

[26] S. Chattopadhyay *et al.*, "A tale from the trenches: cognitive biases and software development," in *Proc. of the ACM/IEEE 42nd Int. Conf. on Softw. Eng.*, 2020, pp. 654–665.

[27] H. Palombo *et al.*, "An Ethnographic Understanding of Software (In)Security and a Co-Creation Model to Improve Secure Software Development," *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*, 8 2020, pp. 205–220.

[28] A. Dalela *et al.*, "A Mixed-method Study on Security and Privacy Practices in Danish Companies," 2021.

[29] M. H. Hue *et al.*, "All your Credentials are Belong to Us: On Insecure WPA2-Enterprise Configurations," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2021, pp. 1100–1117.

[30] D. G. Balash *et al.*, "Educators Perspectives of Using (or Not Using) Online Exam Proctoring," *32nd USENIX Security Symposium*, 8 2023, pp. 5091–5108.

[31] K. L. Wu *et al.*, "Back to School: On the (In)Security of Academic VPNs," *32nd USENIX Security Symposium*, 8 2023, pp. 5737–5754.

[32] B. Burgess *et al.*, "Watching the watchers: bias and vulnerability in remote proctoring software," *31st USENIX Security Symposium*, 8 2022, pp. 571–588.

[33] S. Cohney *et al.*, "Virtual Classrooms and Real Harms: Remote Learning at U.S. Universities," *Seventeenth Symposium on Usable Privacy and Security*, 8 2021, pp. 653–674.

[34] Y. Sun *et al.*, "ZoomP3: Privacy-Preserving Publishing of Online Video Conference Recordings," *Proc. Priv. Enhancing Technol.*, vol. 2022, pp. 630–649, 7 2022.

[35] M. Gruber *et al.*, ""We may share the number of diaper changes": A Privacy and Security Analysis of Mobile Child Care Applications," *Proc. Priv. Enhancing Technol.*, vol. 2022, pp. 394–414, 2022.

[36] R. Hasan and M. Fritz, "nderstanding Utility and Privacy of Demographic Data in Education Technology by Causal Analysis and Adversarial-Censoring," *Proc. Priv. Enhancing Technol.*, vol. 2022, pp. 245–262, 4 2022.

[37] R. Hasan, "Understanding EdTech's Privacy and Security Issues: Understanding the Perception and Awareness of Education Technologies' Privacy and Security Issues," *Proc. Priv. Enhancing Technol.*, vol. 2023, pp. 269–286, 10 2023.

[38] P. Mayer *et al.*, "Why Users (Don't) Use Password Managers at a Large Educational Institution," *31st USENIX Security Symposium*, 8 2022, pp. 1849–1866.

[39] H. Tu *et al.*, "Users Really Do Answer Telephone Scams," *28th USENIX Security Symposium*, 8 2019, pp. 1327–1340.

[40] P. C. Kumar *et al.*, "Privacy and Security Considerations For Digital Technology Use in Elementary Schools," in *Proc. of the CHI Conf. on Hum. Factors Comput. Syst.*, 5 2019, pp. 1–13.

[41] V. Zhong *et al.*, ""I'm going to trust this until it burns me" Parents' Privacy Concerns and Delegation of Trust in K-8 Educational Technology," *32nd USENIX Security Symposium*, 8 2023, pp. 5073–5090.

[42] D. G. Balash *et al.*, "Examining the Examiners: Students' Privacy and Security Perceptions of Online Proctoring Services," *Seventeenth Symposium on Usable Privacy and Security*, 8 2021, pp. 633–652.

[43] N. McDonald *et al.*, "Intersectional Thinking about PETs: A Study of Library Privacy," *Proc. Priv. Enhancing Technol.*, vol. 2023, pp. 480–495, 4 2023.

[44] R. Yin, *Case Study Research and Applications: Design and Methods Sixth Edition*. SAGE Publications, 9 2018.

[45] P. Runeson and M. Höst, "Guidelines for conducting and reporting case study research in software engineering," *Empirical Software Engineering*, vol. 14, pp. 131–164, 2009.

[46] N. Saarimäki, "Methodological Issues in Observational Studies," *SIGSOFT Softw. Eng. Notes*, vol. 44, p. 24, 10 2020.

[47] S. Preibusch, "Guide to measuring privacy concern: Review of survey and observational instruments," *Int. J. of Human-Computer Studies*, vol. 71, pp. 1133–1143, 12 2013.

[48] G. Guest *et al.*, *Applied thematic analysis*. SAGE Publications, 2011.

[49] J. Corbin and A. Strauss, *Basics of Qualitative Research. Techniques and Procedures for Developing Grounded Theory*, fourth edition ed. SAGE Publications, Inc., 3 2015.

[50] D. A. Gioia *et al.*, "Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology," *Organizational Research Methods*, vol. 16, no. 1, pp. 15–31, 2013.

[51] A. Burrows *et al.*, "Privacy, boundaries and smart homes for health: An ethnographic study," *Health & Place*, vol. 50, pp. 112–118, 3 2018.

[52] J. A. Maxwell *et al.*, *Designing a qualitative study*. The SAGE handbook of applied social research methods, 2008, vol. 2, no. 7.

[53] M. Hoegl and H. G. Gemuenden, "Teamwork Quality and the Success of Innovative Projects: A Theoretical Concept and Empirical Evidence," *Organization Science*, vol. 12, pp. 435–449, 8 2001.

[54] N. R. Anderson and M. A. West, "Measuring climate for work group innovation: development and validation of the team climate inventory," *J. Organiz. Behav.*, vol. 19, pp. 235–258, 5 1998.

[55] J. Powar and A. R. Beresford, "Sok: Managing risks of linkage attacks on data privacy," *Proc. Priv. Enhancing Technol.*, vol. 2023, pp. 97–116, 4 2023.

[56] S. Haag and A. Eckhardt, "Shadow it," *Business & Information Systems Engineering*, vol. 59, pp. 469–473, 12 2017.

[57] J.-P. van Acken *et al.*, "Poster: The Unknown Unknown: Cybersecurity Threats of Shadow IT in Higher Education," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2023, pp. 3633–3635.

[58] T. Marjanov *et al.*, "Data Security on the Ground: Investigating Technical and Legal Requirements under the GDPR," *Proc. Priv. Enhancing Technol.*, vol. 2023, pp. 405–417, 7 2023.

[59] NOYB - European Center for Digital Rights, "GDPRhub."

[60] CMS Legal Services EEIG, "GDPR enforcement tracker - list of GDPR fines."

[61] Privacy Affairs, "GDPR fines list: Find all GDPR fines & detailed statistics."

[62] J. Colnago *et al.*, "Is There a Reverse Privacy Paradox? An Exploratory Analysis of Gaps Between Privacy Perspectives and Privacy-Seeking Behaviors," *Proc. Priv. Enhancing Technol.*, vol. 2023, pp. 455–476, 1 2023.

[63] M. W. Vail *et al.*, "An Empirical Study of Consumer Perceptions and Comprehension of Web Site Privacy Policies," *IEEE Transactions on Engineering Management*, vol. 55, no. 3, pp. 442–454, 2008.

Appendix A.
Additional Information

TABLE 8. TYPES OF DATA AND DATA PROCESSING CARRIED OUT AT SCHOOLS

Type of data	Processing description	Data owners	Allowed (Yes / No)
Grades	Online publication	Students	No
Grades	Posting in the restricted area of the electronic registry with access limited to the student or family	Students	Yes
Personal data of pupils	Preparation of school communications not addressed to specific recipients / services related to the school activities (e.g. school bus, canteen).	Students	No
Health data	Dissemination of pupil health data (e.g., including the names of students with disabilities in a circular published online).	Students	No
Data on educational, intermediate, and final outcomes	Communication to third parties to facilitate career guidance, training, and job placement.	Students	Yes, by informing the data owners in advance.
Audio, photos, videos recorded during school related activities	Dissemination on the Internet (website / social networks and so on)	Students, staff, parents	Yes, only with the explicit consent of the data owners.
Personal data	Use of distance education systems within institutional purposes.	Students, parents, teachers	Yes, by informing the data owners in advance.
Audio/video recordings of lectures	Record the lecture for personal purposes only (e.g., for self-study).	Students, teachers	Yes, if it is allowed by school rules
Audio/video shooting	Dissemination on the web and social networks of audio/video footage of a previously recorded lecture for personal purposes.	Students, teachers	No, unless parents and other persons present give explicit informed consent
Video Shooting	Video recording of the lesson in which class dynamics are manifested.	Students, teachers	No, even if distance learning platforms are used.
Class composition data	Publication of data on the institutional website.	Students	No
Class composition data	Posting data in the restricted area of the electronic school register with access limited to the student or family.	Students	Yes (only first and last names).
Rankings of Staff	Online publication of rankings for selection procedures	Teachers, ATA staff	Yes, but only data necessary to identify the candidate.
Video recordings	A video surveillance system inside the building protects the building and school property.	All	Yes, but only at the end of class time and extracurricular activities.

TABLE 9. CODEBOOK: INTERACTIONS BETWEEN PEOPLE PROCESSING DATA

Code	Long code	Description	When to use it (examples)	When not to use it (examples)
IA	Individual action	a person takes an individual action on data	A person leaves a document in a room without interacting with others (excluding, for example, simple courtesy greetings).	A person leaves a document in a room, discussing its content with others there.
I	Interaction	there is a deliberate interaction between two or more people processing data.	Two or more people talk about data.	One person knocks on the door of a room where there is another person and then, without interacting, leaves a document on the table. Two or more people greet each other without talking about or manipulating data.
II	Indirect interaction	interactions between two or more people occurs at the presence of a third party not directly involved in data processing. **Jointly use:** If the indirect interaction occurs due to a compliance failure, Code *II* should be used jointly with Code *F*. If the indirect interaction occurs due to a resource failure, code *II* should be used in conjunction with codes *F* and *R*.	A third party overhears a confidential conversation in another room.	Two persons talk to each other about data or data processing.
MI	Missing interaction	It indicates a lack of interaction between two or more people processing data, i.e., interactions during which some people involved in data processing do not behave as they should.	The *MI* code will be used for interactions between people only. A person asks another one a data processing-related question without having a response. A persons call another one to ask something, but the other person is missing. A person can access a protected area freely because the front desk staff is absent.	It must not used for interactions between humans and machines. A person wants to print a document, but the printer jams, causing the action to end because the person moves on, giving up printing. A person looks for another person to talk to them but does not find them

Appendix B.
Meta-Review

B.1. Summary of Paper

The paper presents a multi-site field observational study of schools to understand personal data processing procedures. The detailed investigation highlights different scenarios and observations that can lead to privacy incidents in a real-world scenario. The paper identifies the problems with resources and accidental failures that can lead to privacy incidents in small organizations, encouraging staff training to mitigate such risks.

B.2. Scientific Contributions

- Provides a Valuable Step Forward in an Established Field
- Independent Confirmation of Important Results with Limited Prior Research

B.3. Reasons for Acceptance

1) The paper provides valuable insights on security and privacy in the context of GDPR regulation through a large-scale field investigation in schools.
2) The methodology is sound and the takeaways highlighted by the study presents important challenges and opportunities to improve privacy in schools and small organizations.