Children’s understanding of personal data and privacy online – A systematic evidence mapping

Mariya Stoilova,a Rishita Nandagirib and Sonia Livingstonea

a Department of Media and Communications; b Department of Social Policy, London School of Economics and Political Science, London, UK

\*m.stoilova@lse.ac.uk @Livingstone\_S

Privacy is both a protective and an enabling right. This article identifies the available research on how children understand, value and negotiate their personal data and privacy online. The results are interpreted according to a framework that distinguishes different privacy contexts (interpersonal, institutional and commercial) and data types (given, traces and inferred). Our systematic evidence mapping of research published by multiple disciplines in the English language from 2007 to 2019 reveals substantial gaps in current knowledge. First, even though children now use the internet from infancy onwards, very few studies include younger children, impeding regulatory or educational policy and practice tailored to socio-developmental differences among children. Second, the majority of research concerns privacy in interpersonal contexts, with less attention paid to commercial or institutional contexts. Relatedly, most research conceives of data that is deliberately and knowingly provided, with fewer studies recognising that data traces and inferred (or metadata) also matter for children’s privacy. Finally, most empirical studies concern children’s behaviours and practices but much less common are studies of children’s media literacy, especially relating to their capacity to consent to data protection practices. Notwithstanding rising concern over the datafication of childhood, few researchers have investigated the possible harms associated with infringements of children’s privacy online, or the value of any social support children receive. We conclude by calling for a more comprehensive and policy-relevant approach to researching children’s practices and understanding of their online data and the privacy implications of engaging in today’s increasingly commercial digital environment.

Keywords: privacy online, personal data, media literacy, children, systematic evidence mapping

# Introduction

Privacy is under scrutiny as the technologies that increasingly mediate communication and information of all kinds become more sophisticated, globally networked and commercially valuable. The conditions under which privacy can be sustained are shifting, accompanied by a rising tide of concern about people’s control over their personal information, contestation over what is public or private in digital environments, and a host of privacy infringements resulting from both the deliberate and unintended actions of individuals and organisations (public and private sector, legal and criminal). Among the many research and policy questions that arise, some concern the position of children growing up in a digital age, whether from a research, policy or regulatory perspective. What do children understand about personal data and privacy in the digital environment? How should it be managed or regulated in their interest? Is digital innovation changing the role of parents, educators and other organisations in supporting the child’s right to privacy?

The policy context regarding children’s privacy online is proving controversial. It includes urgent calls for legislation over the major platforms such as Facebook, Instagram, Google and Snapchat, which host children’s everyday interactions, to mitigate a multitude of risks of harm including content, contact, conduct and commercial risks (Hasebrink, Livingstone, & Haddon, 2008). While policy on online safety is now well-developed, attention to the data economy underpinning platform business models is more recent, as is data protection regulation specifically addressing children’s privacy and wellbeing (DefendDigitalMe, 2018; Kidron, Evans, & Afia, 2018; Montgomery, Chester, & Milosevic, 2017). There are increasing regulatory efforts to ensure businesses explain their data practices more clearly to users, operate more transparently, provide accessible means of redress and conduct risk impact assessments to calibrate their approach to personal data and the risk of harm (Information Commissioner's Office, 2018a, 2018b). But for the most part, such efforts have concentrated on the general public rather than the needs or rights of children in particular (although in the USA, the Child Online Privacy Protection Act, COPPA, offers some specific protections to children; Montgomery et al., 2017). Further, child rights and child welfare organisations are only belatedly recognising the transformative significance of the digital environment for their traditional concerns regarding children’s wellbeing.

Ideally, a judicious apportioning of responsibilities could be reached among the responsibilities of regulation, business, education and parenting. But when public policy interests seem to clash with market interests in relation to digital matters, policy makers often hope that more media literacy education can solve the problem, fearing that new regulation could impede digital innovation. Media literacy education raises questions about what children understand about their personal data and privacy online, and what they can and should be taught. Traditionally such teaching has been led by media educators in school, although sometimes computer science teachers teach ‘digital literacy’ or ‘data literacy’, and there are also calls for such education to be provided by the technology industry, to spare public funds. However it is delivered, such teaching is not easy – many experts are themselves struggling to grasp and explain the legal, technical and organisational processes (involving algorithm-driven marketing, third party data brokers and profiling services, advertising auctions, machine learning and artificial intelligence, and more) that underpin today’s data economy, and, it is increasingly argued, the ‘datafication’ of our lives (Lupton & Williamson, 2017).

In this article we ask what research is available on how children understand about, value and negotiate their personal data and privacy online, and what the significant evidence gaps are as well as next steps for research. Our work is prompted by the adoption of the General Data Protection Regulation (GDPR) across Europe and the UK Information Commissioner’s development of a code of practice for age-appropriate design. We conducted a systematic mapping of the evidence (EPPI-Centre, 2018; Gough, Thomas, & Oliver, 2012; Grant & Booth, 2009), paying particular attention to children’s age, and using a comprehensive multidisciplinary search strategy to recognise possible contributions from childhood studies, developmental psychology, education, computer science and technology studies, among others. The purpose is to ensure that research with children informs policy debates over whether and how their privacy can be protected in the digital environment.

# Conceptual framing and research questions

Privacy is both a fundamental human right vital to personal autonomy and dignity and the means of realising other rights. In relation to children, privacy is more often recognised as a protective than an enabling right, with policy and welfare organisations seeking to protect children’s safety, family relations and emotional development from privacy infringements (Lievens, Livingstone, McLaughlin, O’Neill, & Verdoodt, 2018). But it is also important for children’s freedom of expression, exploration of self, information-seeking and civic rights, as specified in the United Nations Convention on the Rights of the Child (1989). Yet society does not always fully acknowledge children’s positive rights to provision and participation, especially online. Nor does it always recognise that, in certain circumstances, the child’s right to protection from harm or to participate in society may necessitate their privacy from parents or caregivers, school, business or the state.

In the digital age, children’s privacy is increasingly threatened by new forms of data collection and surveillance enacted by both parents and the state (including schools, health and welfare systems and law enforcement) even as the internet provides new opportunities for children’s interactions, exploration and information and help-seeking – often valued precisely because they can be conducted privately. Children’s privacy is also threatened in ways still poorly understood by the commercialisation of the digital environment and the commodification of children’s online speech and actions as data (Livingstone, Stoilova and Nandagiri, 2019; Lupton & Williamson, 2017). The creation of data – which can be recorded, tracked, aggregated, analysed (via algorithms and increasingly, via artificial intelligence) and ‘monetised’ – from the myriad forms of human communication and activity which throughout history have gone largely unrecorded, is generating a new form of economic value and, thereby, a new set of market actors, data processes and emergent consequences.

In this article, we note Westin’s (1967, p. 7) classic conceptualisation of privacy as ‘the claim of individuals, groups, or institutions to determine themselves when, how and to what extent information about them is communicated to others’ resonates with everyday definitions. This prioritises a conception of information as that which is knowingly given or shared with others. This tends to downplay data ‘given off’, to use Goffman’s (1971) term, although in the digital environment data given off is more the norm than the exception, with online activities leaving data traces (or providing meta data – data about the data; see van der Hof, 2016) that are routinely viewed, shared, collected and retained by others. Although Westin does not ignore the inferences drawn by others from data given or given off, his emphasis on individual control seems at odds with the extent to which modern data analytics based on inferred or profiled personal data now constitute the core business model of many digital companies, as well as being increasingly valuable to public institutions deploying learning or health analytics. In short, the evolving data ecology includes multiple types of data whose processing goes beyond questions of individual control (van der Hof, 2016).

To approach this complexity, we draw on Nissenbaum’s analysis of privacy as contextual integrity, meaning ‘neither a right to secrecy nor a right to control, but a right to appropriate flow of personal information’ (Nissenbaum, 2010, p. 3) in order to grasp the relational nature of privacy generated through interactions among people and/or organisations in diverse and often complex digital contexts. In the emphasis on ‘appropriate’ (as judged by whom and negotiated how?) and ‘flow’ (from whom to whom or what?), this definition emphasises culturally shaped negotiations among participants in particular contexts. Thus, norms and regulations regarding expectations, appropriateness, inclusion, purpose, modality and violation should be taken into account when making judgements about privacy and privacy-related actions (O'Hara, 2016).

What kinds of context are pertinent to children’s privacy in the digital age? We distinguish three – interpersonal, institutional and commercial. Together these contexts capture distinct kinds of relationship in which people are embedded: with (i) other individuals or groups, (ii) public or third sector (not-for-profit) organisations, and (iii) commercial organisations. These contexts vary in the cultural norms, power relations and regulatory mechanisms that apply, resulting in different privacy-related problems experienced by children and to be addressed by policy makers. It is important to consider what children understand and expect regarding their privacy in each context, and how they anticipate which problems may arise in each.

These contexts routinely co-occur: for instance, when a child chats to friends on a popular app in the school playground, questions of privacy arise regarding (i) what they share with their friends and what they in turn may share, (ii) the business that owns the app and, thereby, the digital record of interactions (and what they may share with business partners) and (iii) the school, which may have safeguarding reasons to monitor (manually or technologically) the child’s ‘private’ communication (and possibly, share this with parents, social workers or the police) (Livingstone, Stoilova and Nandagiri, forthcoming; Shade & Singh, 2016). These contexts become further interconnected in the digital environment, resulting in ‘networked privacy’ (the geographically extensive, public-by-default nature of personal communications in the digital era) (boyd & Marwick, 2011). Privacy online is shaped by four key affordances (or socially designed features) of the digital environment – persistence, replicability, scalability and searchability – and can result in what Marwick and boyd (2014) term context collision, uncertain audiences and context collapse. In short, in the digital age, people must contend with new digital dynamics over and above the established demands of social interaction. This results in emergent tactics designed to evade or manage online visibility, secrecy or surveillance.

How can children grasp such complexities of context and consequence in a fast-changing digital environment? Can this be taught? The UK national curriculum pays only partial attention to an operational and critical understanding of online data and privacy.[[1]](#footnote-1) From 2020 the curriculum will include compulsory teaching about relationships and consent from the start of primary school and sexual relationships and consent from the start of secondary school (Department for Education [DfE], 2018). This may include interpersonal privacy online, together with the rights to delete and redress. The computing curriculum, also compulsory from primary school, includes an introduction to algorithms and using technology safely and responsibly (KS1–2),[[2]](#footnote-2) while the KS3–4 curriculum explores the safe, responsible and secure use of technology including the implications of technology for online identity and privacy (DfE, 2013). The citizenship curriculum (compulsory from KS3) prioritises an understanding of democracy, government and the rights and responsibilities of citizens. The (optional) computing GCSE curriculum adds material on the ethical, legal and environmental impacts of digital technology on wider society, including issues of privacy and cyber security (DfE, 2016a, 2016b). The GCSE media studies curriculum offers in-depth critical understanding of the digital media ecology, but is chosen by few students (DfE, 2016a, 2016b).

Insofar as children are not taught or are too young to learn enough about data and privacy online to protect their rights, what is the responsibility of regulators? The classic social policy answer in the UK refers to the so-called Gillick principle (NSPCC, 2018; Taylor, Dove, Laurie, & Townend, 2018), established in 1982 when Lord Scarman adjudicated on when contraception could be given to a child without informing their parents. He stated that the ‘parental right yields to the child’s right to make his own decisions when he reaches a sufficient understanding and intelligence to be capable of making up his own mind on the matter requiring decision’. Since sexual activity with a child under 13 is a criminal offence, this in effect set a lower age limit for children’s independent decisions. Related considerations about children's cognitive development appear to have informed the USA’s 1998 COPPA, designed to protect children under 13 from marketing messages without parental permission (Haddon & Livingstone, 2017). Although developed originally for television advertising, COPPA became the de facto age regulation for services offered by (mostly USA-headquartered) digital platforms worldwide (hence Facebook, Snapchat, etc. require users to be 13+). In 2018 the European GDPR set the age of 16 (with member states permitted to derogate by choosing any age between 13 and 16) as the so-called ‘digital age of consent’ at which children could decide about uses of their digital data without parental permission. This has been much misunderstood, since the restriction applies only to personal data processing on the legal basis of consent – resorted to less by platforms than expected (Livingstone, 2018). More recently, policy makers are exploring the possibilities for privacy-by-design solutions, including those which take into account the age and maturity of child users (Kidron & Rudkin, 2017; Livingstone, 2014; Macenaite, 2017; UNICEF, 2018).

This evidence review aims to identify and evaluate the existing evidence on children’s data and privacy online, focusing on empirical evidence regarding children’s understanding, experiences and practices. The research questions are:

RQ1. How much and what kind of research exists on how children understand, experience and act regarding their data and privacy online?

RQ2. Does the available research encompass children of different ages and maturity?

RQ3. What are the significant gaps and needed next steps in research on children’s data and privacy online?

# Methodology

Systematic evidence mapping refers to a review process that systematically identifies and describes the research that exists within the boundaries of the review (EPPI-Centre, 2018). It does not offer the advantages of a systematic evidence review but rather, depicts the nature of the research field in order to inform the conduct and interpretation of a subsequent synthesis of findings (EPPI-Centre, 2018; Gough et al., 2012; Grant & Booth, 2009). As the volume and breadth of available research was initially unknown, a mapping approach seemed advisable as the first step.

Three fields were identified as relevant to the scope of the review – research on children (generally taking a social, psychological or cultural approach), legal and regulatory studies focused on privacy and rights, and technological/computer sciences of data protection, privacy-by-design and the larger data economy. As shown in Figure 1, the intersections of these fields are of particular interest – for instance, media literacy, digital skills and media education fall in the segment that intersects children and the digital.

Figure 1. Concept mapping of three disciplines and their overlaps

A close up of text on a white background

Description generated with high confidence

In consultation with our university’s specialist librarian, we selected 19 databases for their suitability to the review’s scope and aims (see Figure 2; for details, see Livingstone et al., 2019). These cover the social sciences, legal studies, computer science studies, government publications, legal documents and grey literature (policy reports, conference papers, advocacy tools and case studies). The decision to include grey literature was taken because this is a fast-moving field in which a range of privacy-focused, regulatory and child welfare organisations are conducting research that does not always get published in peer-reviewed journals. The inclusion criteria were set to capture studies from any country but published in English, and published since 2007 to ensure relevance for current contexts and technological advances (older research was judged less pertinent given the many changes in both technological contexts and user practices in the past decade).

The database searches were conducted using three groups of search terms: (i) child terms, (ii) digital technology terms and (iii) privacy terms. Search testing was conducted to ensure validity, optimal coverage and efficiency. The terms were discussed, conceptually mapped and reviewed by the team for reliability before fine-tuning them to produce the final search combination:

* Child terms: child\* OR youth OR teen\* OR adolescen\* OR minor OR kid OR girl OR boy OR student OR pupil
* Digital terms: digital\* OR mobile\* OR internet OR online
* Privacy terms: priva\*

The search included title AND abstract AND keywords (where keyword search was available). For some databases, search options restricted us to abstracts, metadata, keywords or title only. In the initial search testing, Group (ii) included the term ‘data’ but this returned many extraneous results. In Group (iii) we attempted including ‘data’ and ‘secur\*’ but this produced many irrelevant sources and were, therefore, removed.

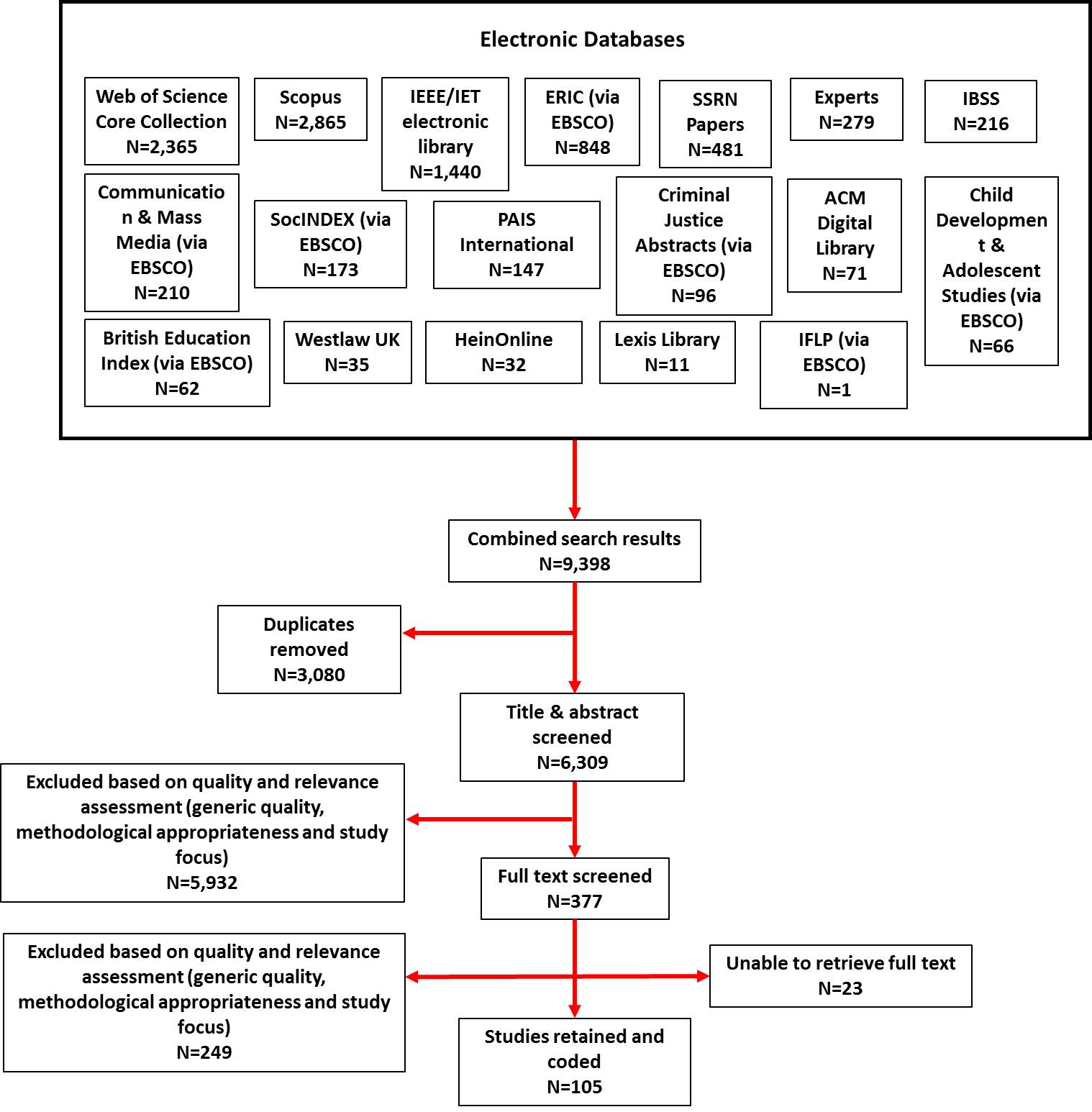
The search produced 9,119 sources. To capture the complexity and insights available, the team requested recommendations from experts on research literature. This brought the combined search results to 9,398 sources (see Figure 2). Removal of duplicates reduced the search results to 6,309. At this point, it was recognised that available research was considerable in volume but diffuse in its boundaries, spanning multiple disciplines and of the variable quality to be expected of a relatively new field where definitions and measures are still in flux. It was not practical to retain the weaker studies for detailed analysis, as is often done in a systematic mapping (to identify areas of stronger or weaker evidence).

It was decided to conduct an assessment of quality, appropriateness and relevance first during the screening of the 6,309 studies based on their title and abstract resulting in 377 studies, and then again, during the screening of the full text of the 377 studies. This screening was informed by EPPI Centre’s weight of evidence framework (Gough, 2007). A global evaluation of each study was derived, based on the three main criteria in this framework:

* The generic quality of execution of the study (EPPI 1): to ensure the comprehensiveness of the review in this emerging field, the initial scoping included peer-reviewed journals as well as policy or advocacy-related publications from non-governmental organisations (NGOs), government reports, industry sources and other relevant grey literature. Only studies that were assessed to be of medium or high quality (in relation to clarity of purpose, accuracy of claims and conclusions, accessibility and quality of methodology) were included. Some exclusions related to unconvincing description of the methodology or unclear terminology.
* The appropriateness of the methodology to the review (EPPI 2): to answer the research questions, we only included studies that involved primary empirical research with children (under the age of 18). The studies assessed to be of low relevance in relation to this criterion were excluded. These used the term ‘child’ in relation to adults (e.g., child custody; childbirth; adult children), or used adult ‘student’ samples.
* The focus of the study (EPPI 3): this was judged for its relevance and only studies assessed as of medium or high relevance were included – these focused on children’s online privacy. The studies judged to be of low relevance included studies of the private sector (captured by the search term ‘priva\*’), technological developments that mentioned privacy only incidentally (e.g., development of apps for children); online purchase choices (persuasive online design); ICT education and digital skills without a specific focus on privacy; or technical aspects such as engineering solutions to privacy hazards. Some studies assessed as of low relevance did not focus on the digital environment (from the databases that did not allow the cross-over of all three search groups). The search also produced databases of conference proceedings (rather than individual papers), which were removed due to low relevance of the individual sources.

In effect, while we had planned to assess the generic quality of all the studies identified (EPPI 1), given the large number of studies only medium or high-quality studies were included in the final results. As the requirements of primary empirical research with children (EPPI 2) and a focus on children’s privacy online (EPPI 3) were introduced after the initial database search, the resulting methodology is classified as an ‘emergent criteria’ approach (Gough, 2007). Where the full text was not available results were removed, as were similar outputs by the same author (e.g., conference paper and a journal article – in such cases, the most recent or reliable source was retained). This produced a final set of 105 studies to be summarised and coded via a coding template developed for the purpose.

Figure 2. The screening process



The coding involved reading the 105 articles in full and coding for key information on the type of methodology used, age of the children in the study, main topics covered, privacy contexts and data types discussed (see Table 1). This was conducted by two of the authors: they independently coded the same 10 studies before discussing discrepancies and reaching a consensus about the application of the coding framework. The remaining studies were coded by one researcher and checked by the other to ensure coding reliability. The coding framework proved effective, except that age had to be coded in multiple and not mutually exclusive categories because studies used different, overlapping or even unclear age groupings (e.g., only reporting the average age). The year of publication was noted for each study to examine changes in this field over time.

Our systematic evidence mapping was restricted by language, focusing only on English publications. Some level of subjectivity in assessing quality, appropriateness and relevance of articles is inevitable, even with the specific guidelines established during the screening phase.

# Results

The coded results of the 105 studies identified by the systematic evidence mapping are shown in Table 1.

Table 1: Evidence mapping of 105 studies by coding categories

|  |  |  |
| --- | --- | --- |
| Category | Type | Number of studies |
| Type of study  (N=105) | Primary research | 98 |
| Secondary data analysis | 7 |
| Methodology  (mutually exclusive categories) | Survey | 36 |
| Interview (13 individual, 4 focus group discussions) | 17 |
| Mixed qualitative methods | 12 |
| Mixed qualitative/quantitative methods | 10 |
| Experimental or quasi-experimental | 9 |
| Participatory | 9 |
| Secondary analysis | 7 |
| Digital research | 5 |
| Age  (coded in multiple categories) | Age 0–3 | 2 |
| Age 4–7 | 11 |
| Age 8–11 | 45 |
| Age 12–15 | 84 |
| Age 16–19 | 71 |
| Could not categorise | 3 |
| Topic  (coded in multiple categories) | Behaviours and practices | 70 |
| Values and beliefs about the internet | 53 |
| Media skills and literacies | 45 |
| Privacy strategies and tactics of the child | 36 |
| Support and guidance from others | 10 |
| Design and interface (affordances) | 9 |
| Risks and harms | 4 |
| Consent | 2 |
| Privacy contexts  (mutually exclusive categories) | Interpersonal | 63 |
| Commercial | 10 |
| Institutional | 3 |
| Multiple | 29 |
| Data type  (mutually exclusive categories) | Data given | 76 |
| Data traces | 3 |
| Inferred data | 1 |
| Multiple | 25 |

## Volume of research by method and year

Children’s privacy in the digital environment represents a recent and growing area of research, with publications rising steeply from around 2013 (Figure 3). Most research is primary (93%), although a few studies involve secondary data analysis (see Table 1). One-third of the studies (34%) rely on survey methods, with interviews used half as much. Mixed methods (qualitative, or qualitative and quantitative) are moderately popular, with less use of experimental, participatory or specifically digital methods (grouped as ‘Other methods’ in Figure 3). The diversity of methods used in research rose from around 2013.

Figure 3. Mapping evidence by year and methodology (N=105)

## Volume of research by age

The youngest age groups 0–3 and 4–7 are barely researched (only 11 studies, of which two focused on children younger than 4; Table 1). Bearing in mind that most studies’ sample populations cross or combine the age categories used here, older age groups are fairly evenly represented: 43% studies included children aged 8–11, 80% included those aged 12–15 and 68% included those aged 16–19.

## Privacy contexts researched

Most (62%) studies focus only on interpersonal privacy, with a further 26% addressing multiple privacy types. The latter generally includes both interpersonal and commercial privacy (24 studies), although the findings do not always distinguish the two. Although only 10 (10%) studies focus exclusively on commercial privacy, in all, 38 (37%) address it. Few studies address institutional privacy (8 in all, 3 as an exclusive focus – these are diverse in nature, including children’s privacy in relation to school, law enforcement, state welfare and civic authorities, regulators and employers). The recent rise in privacy-related studies is largely due to studies of interpersonal privacy, with the other two privacy contexts still barely researched (Figure 4).

Figure 4. Evidence mapping of privacy types by year of publication (N=105)

## Data types researched

Overwhelmingly, 76 studies focus on data given (72% of all studies). The remaining studies generally focus on multiple data types, of which 20 of the 25 covering multiple data types focus on data given and data traces. Inferred data has received scant attention in the literature, notwithstanding the rise in profiling and the value of the associated data economy. The recent rise in studies on children’s data and privacy online has not led to a diversification in types of data researched (Figure 5). Studies exploring multiple privacy contexts tend also to focus on multiple data types. Even where studies include a focus on institutional privacy, this tends to concern data given over other data types.

Figure 5. Evidence mapping of data types by year of publication (N=105)

## Topics researched in relation to children’s data and privacy online

Most empirical studies concern children’s behaviours and practices (67%). Bearing in mind that most encompassed multiple topics (Table 1), next most commonly studied was children’s values and beliefs about the internet (50%), their media skills and literacies (45%), and their privacy strategies and tactics (34%).

Much less common are studies of the support and guidance that children receive from others, the design and interface of the sites and services that children engage with online or children’s capacity to consent to data protection practices online – surprisingly, given the public and policy attention devoted to this. Also barely researched is the nature of the risk of harm or actual harm associated with infringements of children’s privacy online, although much policy and regulation is designed to promote children’s understanding and regulation of digital services so as to mitigate harm.

Broadly speaking, children’s privacy-related behaviours and practices, and their values and beliefs, are addressed in relation to the different privacy contexts and cover multiple data types (Table 3). Given that the interpersonal privacy context is the most often researched, evidence for the different topics is concentrated here. Some of the gaps are worrying: we identified no research on children’s understanding of consent in commercial contexts online, and studies of children’s online privacy strategies and tactics never address institutional contexts (such as school or state), although 11 address commercial contexts (2 studies just on commercial privacy, 9 combine interpersonal and commercial privacy). Similarly, research on risks and harms is not only sparse (in the context of a mapping of privacy research), but also largely limited to data given in interpersonal contexts.

Interestingly, there is a modest body of research on children’s media skills and literacies that encompasses not only interpersonal privacy but also multiple privacy contexts (where 19/20 studies cover commercial contexts), making for 23 studies that include children’s understanding of commercial contexts.

Table 2: Evidence mapping of privacy contexts and topics covered

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Topics  (multiply coded) | Interpersonal privacy | Commercial privacy | Institutional privacy | Multiple privacy contexts |
| Behaviours and practices | 49 | 5 | 2 | 14 |
| Values and beliefs about the internet | 34 | 5 | 2 | 12 |
| Media skills and literacies | 20 | 4 | 1 | 20 |
| Privacy strategies and tactics | 25 | 2 |  | 9 |
| Support and guidance | 6 |  |  | 4 |
| Design interface | 4 | 1 | 2 | 2 |
| Consent | 1 |  | 1 |  |
| Risks and harms | 3 |  |  | 1 |

We have not tabulated the findings here for data types, given that only three studies uniquely addressed data traces and there was just one on inferred data.

# Discussion and conclusions

This systematic evidence mapping examined research on children’s understandings, experiences and behaviours regarding their data and privacy online. Adopting an exploratory approach to this already-large yet diffuse field spanning multiple disciplines, we observed a recent and rapid growth in the available research, likely in response to increasing public and policy interest. Studies reviewed were found to use a range of methodologies, doubtless reflecting the challenge of researching complexities of privacy, technology and social context with children. However, this breadth of methodologies is focused narrowly in terms of populations (children’s age) and topics studied.

Most research concerns children over eight years old, with the preponderance of studies on adolescents. This may reflect the past demographics of internet users, but as children now use the internet from infancy onwards, including for interpersonal communication and for sites and devices that collect and share personal data, it is timely to include younger children in future studies. Around eight years old is commonly taken to be the age at which children begin to understand the intentions of others in direct and mediated interactional contexts (Livingstone, 2009) and arguably, parents should be responsible for children’s data and privacy online in the early years. But in practice, these responsibilities may be only partially fulfilled. While developmental research shows that children of different ages have different understandings and needs, it is also clear that privacy is vital for wellbeing across the age ranges (James, Jenks, & Prout, 1998; Kidron & Rudkin, 2017). Until more research is undertaken, it remains unclear just what young children understand and how they act in relation to the digital environment (Mascheroni & Holloway, 2019). It may be that research with parents can shed light on young children’s understanding, although it would still be valuable to include children’s own voices where possible. This matters because policy and practice that respects the child’s best interests is likely to require a tailored approach that acknowledges developmental and individual differences among children.

Of those studies that included children younger than eight, most used a small sample size. They suggest that very young children have low risk awareness (Bakó, 2016) and little understanding that sharing information online can create privacy risks. Rather, they rely on adults to advise them and create rules. Between 8 and 11 years old, children still have gaps in their ability to decide about trustworthiness of the sources or identify commercial content, but they are beginning to understand that sharing personal information online is associated with risks. They manage their privacy based on rules they have been taught (mostly relating to ‘stranger danger’) but have not yet internalised these behaviours and can be unsure how to apply their knowledge to practical situations. From age 12 onwards, children are more aware of privacy risks and often engage in careful consideration of information disclosure. They begin to grasp aspects of institutionalised and commercial privacy – for example, in relation to school monitoring of online activities and exposure to advertising content based on browsing history. Even the oldest children (aged 17) understand little of data flows and digital infrastructure – they mostly see data as static and fractured.

However, for older children and adolescents, the review identified a considerable imbalance in the available research, finding it highly concentrated on specific topics in interpersonal contexts, with less conducted on commercial contexts and very little on institutional contexts. An interest in interpersonal privacy is to be expected given the rise of social media, the increasing mediation of personal relationships, the privacy policies of powerful and profitable platforms and the associated risks of harm. One persistent focus is on the gap between a claimed concern for privacy online and actual behaviour. This ‘privacy paradox’ has prompted researchers to explore whether raising children’s awareness of risks and the importance of privacy could support better self-protective behaviour. While some studies find support for this (Chai, Bagchi-Sen, Morrell, Rao, & Upadhyaya, 2009; Chi, Jeng, Acker, & Bowler, 2018; Madden et al., 2013; Moscardelli & Divine, 2007; Youn, 2008), others demonstrate the importance of contextual factors, such as whether the child feels in control of their information (Davis & James, 2013), or how they think their parents and peers would judge their actions (van Gool, van Ouytsel, Ponnet, & Walrave, 2015). Another possible explanation concerns whether children have the media literacy to perform certain functions (managing privacy settings, deleting contacts, reporting content) (Bowler, Acker, Jeng, & Chi, 2017; Livingstone, Mascheroni, Ólafsson, & Haddon, 2014; Selwyn & Pangrazio, 2018).

Relevant to growing interest in the datafication of childhood, we found that one-third of available studies address children’s understanding of the privacy implications of engaging in the commercial digital environment; few of these were conducted recently, an added concern given companies’ fast-evolving practices of data harvesting and profiling. Even fewer studies examine children’s understanding of other institutions that collect their personal data – school, medical, youth or health services, transport systems, and so forth. It may be supposed that responsible adults – parents, teachers, policy makers – have effectively protected their interests in this regard, but there is little evidence that this is the case for either planned data uses or to prevent unanticipated data breaches or illicit uses of data (Fontaine & Dave, 2018; Leetaru, 2015). A child rights approach asserts that children’s own agency and self-determination should be respected as regards their privacy, including by their family and school (Council of Europe, 2018).

Although in principle the three privacy contexts on which this review has focused may deploy multiple types of data, most research conceives of data in terms of that which is deliberately and knowingly provided by the child. A quarter of the studies reviewed recognise that data traces and metadata also matter for children’s privacy. But the fact that others – individuals, institutions, businesses – make inferences based on data collected is little studied in relation to children’s understanding. As a result, there is little evidence from children to inform policy deliberations regarding the risks of profiling (e.g. for advertising and marketing) or, indeed, its benefits (e.g. for learning or health analytics). Yet although children develop their privacy-related awareness, literacy and needs as they grow older, even the oldest children struggle to comprehend the complexity of internet data flows and data commercialisation to protect their privacy online.

Finally, the range of topics researched includes strengths and weaknesses. Most studies concern children’s behaviours and practices, followed by their values and beliefs, media literacy and privacy tactics. This implicitly conceives of the young internet user as an isolated individual faced with key choices in the digital environment. We know little of the social support that children receive – whether before making key privacy decisions or afterwards, perhaps when something has gone wrong. Few studies follow up on children’s choices to identify the consequences or their responses. Nor do they relate children’s choices to the design of the digital interface, notwithstanding research on user-centred design and lively policy debates on the design of child-friendly privacy settings or other data protection features.

In addition to broadening the range of research questions, future research should develop greater definitional clarity. We have distinguished privacy contexts and data types in ways that, if carried into future research designs, will ensure new evidence directly informs unfolding policy. It is problematic that studies of children’s privacy online concern the data they knowingly give, not the digital traces they leave or the profiling that may result.

Second, notwithstanding the importance of age as a proxy for children’s social development (or in child rights terms, their evolving capacity), the analysis was impeded not only because studies grouped children differently by age – for example, by year of birth, school year or school type (primary, middle, secondary) – but also because not all studies were clear about the ages included in the sample. This makes it difficult to compare across studies, as well as to relate findings to current policy. For instance, European debates over the so-called digital age of consent (at which children can access services without parental consent) need evidence in choosing 13, 14, 15 or 16 (Lievens et al., 2018; Livingstone, 2018). Relatedly, calls for revisions to the school curriculum need evidence about what children know, and need to know, at different ages. While we acknowledge critiques of ‘ages and stages’ approaches to child development (Buckingham, 2015; James et al., 1998), and recognise that children develop in different ways and on different timescales, it is unhelpful to report research for ‘children’ (or, even, ‘adolescents’) without granular differentiation.

What should the next steps be for research on children’s data and privacy online? Given the concentration of studies on teenagers’ interpersonal privacy online, a substantial gap exists around younger children’s understandings of privacy online. Equally clear is the need for further research on the institutional and commercial contexts within which children’s personal data are used. In addition to the topics identified earlier as attracting more or less research, the interrelations among these merits further investigation (and theorisation). Studies included in the present mapping suggest that a combination of internet use, media literacy, trust and socio-cultural norms influences how children understand, experience and act regarding their data and privacy online. More research is needed to explain how these factors interact across the age range, and to address such under-researched topics as how social support and guidance, on the one hand, and better technological design, on the other, may help protect children’s privacy.

By far the most research on harms to children linked to the digital environment concerns their interpersonal safety. The potential for other harms (discrimination, exclusion, political manipulation, commercial exploitation, radicalisation, etc.) has barely been scoped, let alone researched, but such potential is likely in future to be founded on institutional and commercial uses of personal data. To raise such issues up the policy and public agenda, evidence is needed on whether and how the risk of harm is linked to privacy infringements. Research should consider the relevance and consequences of privacy-related risks (disclosure of sensitive information, receipt of targeted advertising, confusion over privacy settings, etc.) and harms (personal upset, reputational damage, exploitation via scams or data breaches, etc.). Questions also arise regarding the timescale – should consequences be limited to the immediate situation or researched over months or even years, given that personal data risk in childhood may have consequences in adulthood (Children's Commissioner for England, 2018)?

At present, the digital environment typically presents children with a dilemma in which they find themselves forced to weigh the desire to protect their privacy against the opportunities of online participation. Often their decisions favour immediate benefits of participation over uncertain privacy risks in the future. Often, too, children are themselves criticised for such decisions. Future research might examine how educational, regulatory or design interventions could alter the nature of this dilemma and, therefore, the decisions children face. To prevent children in the future being presented with such invidious dilemmas, to reach a more effective balance of responsibilities among industry, government, educators and parents, and to empower children as actors without unduly burdening them as individuals in a complex digital environment largely not of their own making, we urge more multidimensional research on children’s data and privacy online.

**Acknowledgements.** We thank the Information Commissioner’s Office for funding the research on which this article is based. We also acknowledge the valuable guidance of our expert advisory group and university subject librarians. This article draws on material that originally appeared in Livingstone et al. (2019).

**References**

Bakó, R. K. (2016). Digital transition: Children in a multimodal world. *Acta Universitatis Sapientiae, Social Analysis, 6*(1), 145–154.

Bowler, L., Acker, A., Jeng, W., & Chi, Y. (2017). 'It lives all around us': Aspects of data literacy in teen's lives. Paper presented at the 80th Annual Meeting of the Association for Information Science & Technology, Washington DC, USA.

boyd, d., & Marwick, A. E. (2011). Social privacy in networked publics: Teens’ attitudes, practices, and strategies. Paper presented at A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, Oxford, UK.

Buckingham, D. (2015). Defining digital literacy – What do young people need to know about digital media? *Nordic Journal of Digital Literacy, 10*, 21–35.

Chai, S., Bagchi-Sen, S., Morrell, C., Rao, H. R., & Upadhyaya, S. J. (2009). Internet and online information privacy: An exploratory study of preteens and early teens. *Transactions on Professional Communication, 52*(2), 167–182.

Chi, Y., Jeng, W., Acker, A., & Bowler, L. (2018). Affective, behavioral, and cognitive aspects of teen perspectives on personal data in social media. In G. Chowdhury, J. McLeod, V. Gillet, & P. Willett (Eds), *Transforming digital worlds* (pp. 442–452). Cham, Switzerland: Springer.

Children's Commissioner for England. (2018). *Who knows what about me? A Children’s Commissioner report into the collection and sharing of children’s data*. London: Children's Commissioner for England.

Council of Europe. (2018). *Recommendation CM/Rec(2018)7 on the child in the digital environment*. Retrieved from [https://search.coe.int/cm/Pages/result\_details.aspx?ObjectId=09000016808b79f7](https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016808b79f7:)

Davis, K., & James, C. (2013). Tweens' conceptions of privacy online: Implications for educators. *Learning, Media and Technology, 38*(1), 4–25.

DefendDigitalMe. (2018). *The state of data. Lessons for policymakers*. Retrieved from [http://defenddigitalme.com/wp-content/uploads/2018/05/StateOfDataReport\_policymakers\_ddm.pdf](http://defenddigitalme.com/wp-content/uploads/2018/05/StateOfDataReport_policymakers_ddm.pdf:)

Department for Education (DfE). (2013). *The national curriculum in England*. Retrieved from <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/425601/PRIMARY_national_curriculum.pdf>

Department for Education (DfE). (2016a). *Media studies: GCE AS and A level subject content*. Retrieved from <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/497237/Media_studies_AS_A_level_subject_content.pdf>

Department for Education (DfE). (2016b). *Media studies: GCSE subject content*. Retrieved from <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/496451/Media_studies_GCSE_subject_content.pdf>

Department for Education (DfE). (2018). New relationships and health education in schools. Draft guidance. Press release. Retrieved from: www.gov.uk/government/news/new-relationships-and-health-education-in-schools

EPPI-Centre. (2018). *Definitions.* Retrieved from http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=334

Fontaine, C., & Dave, K. (2018). *Spectrum of trust in data: New York city parents navigating school choice*. New York: Data & Society.

Goffman, E. (1971). *The presentation of self in everyday life*. Harmondsworth: Penguin.

Gough, D. (2007). Weight of evidence: A framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education, 22*(2), 213–228.

Gough, D., Thomas, J., & Oliver, S. (2012). Clarifying differences between review designs and methods. *Systematic Reviews, 1*, 1–9.

Grant, M., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal, 26*, 91–108.

Haddon, L., & Livingstone, S. (2017). Risks, opportunities and risky opportunities: How children make sense of the online environment. In P. Brooks & F. Blumberg (Eds), *Cognitive development in digital contexts* (pp. 275–302). San Diego, CA: Elsevier.

Hasebrink, U., Livingstone, S., & Haddon, L. (2008). *Comparing children's online opportunities and risks across Europe.* London: EU Kids Online.

Information Commissioner's Office. (2018a). *Children and the GDPR*. Wilmslow: Information Commissioner's Office.

Information Commissioner's Office. (2018b). *Guide to the General Data Protection Regulation.* Wilmslow: Information Commissioner's Office.

James, A., Jenks, C., & Prout, A. (1998). *Theorizing childhood*. Cambridge: Polity Press.

Kidron, B., & Rudkin, A. (2017). *Digital childhood: Addressing childhood development milestones in the digital environment.* Retrieved from <https://5rightsframework.com/static/Digital_Childhood_report_-_EMBARGOED.pdf>

Kidron, B., Evans, A., & Afia, J. (2018). *Disrupted chidhood. The cost of persuasive design*. Retrieved from http://eprints.mdx.ac.uk/25600/

Leetaru, K. (2015, December 12). When toys are hackable and our children become data breach victims. *Forbes*.

Lievens, E., Livingstone, S., McLaughlin, S., O’Neill, B., & Verdoodt, V. (2018). Children’s rights and digital technologies. In U. Kilkelly & T. Liefaard (Eds), *International human rights of children* (pp. 1–27). Singapore: Springer Singapore.

Livingstone, S. (2009). Debating children’s susceptibility to persuasion – Where does fairness come in? *International Journal of Advertising, 28*(1), 2–6.

Livingstone, S. (2014). Developing social media literacy: How children learn to interpret risky opportunities on social network sites. *Communications. The European Journal of Communication Research, 39*(3), 283–303.

Livingstone, S. (2018). Children: A special case for privacy? *Intermedia, 46*(2), 18–23.

Livingstone, S., Stoilova, M. and Nandagiri, R. (2019) *Children’s data and privacy online: Reviewing the existing evidence*. London: London School of Economics and Political Science.

Livingstone, S., Stoilova, M. and Nandagiri, R. (forthcoming). Data and privacy literacy: The role of the school in educating children in a datafied society. In D. Frau-Meigs (Ed.), *Handbook on media education research*. London: Routledge.

Livingstone, S., Mascheroni, G., Ólafsson, K., & Haddon, L. (2014). *Children’s online risks and opportunities: Comparative findings from EU Kids Online and Net Children Go Mobile.* London: EU Kids Online.

Lupton, D., & Williamson, B. (2017). The datafied child. *New Media & Society, 19*(5), 780–794.

Macenaite, M. (2017). From universal towards child-specific protection of the right to privacy online: Dilemmas in the EU General Data Protection Regulation. *New Media & Society, 19*(5), 765–779.

Madden, M., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., & Smith, A. (2013). *Teens, social media, and privacy*. Washington, DC: Pew Research Center.

Marwick, A. E., & boyd, d. (2014). Networked privacy: How teenagers negotiate context in social media. *New Media & Society, 16*(7), 1051–1067.

Mascheroni, G., & Holloway, D. (2019). *The internet of toys: Practices, affordances and the political economy of children’s smart play*. Cham, Switzerland: Palgrave.

Montgomery, K. C., Chester, J., & Milosevic, T. (2017). Children's privacy in the big data era: Research opportunities. *Pediatrics, 140*(s2), 117–121.

Moscardelli, D. M., & Divine, R. (2007). Adolescents' concern for privacy when using the internet: An empirical analysis of predictors and relationships with privacy-protecting behaviors. *Family & Consumer Sciences Research Journal, 35*(3), 232–252.

Nissenbaum, H. (2010). *Privacy in context. Technology, policy, and the integrity of social life*. Stanford, CA: Stanford University Press.

NSPCC. (2018). Gillick competency and Fraser guidelines. Retrieved from https://learning.nspcc.org.uk/media/1541/gillick-competency-factsheet.pdf

O'Hara, K. (2016). The seven veils of privacy. *IEEE Internet Computing, 20*, 86–91.

Selwyn, N., & Pangrazio, L. (2018). Doing data differently? Developing personal data tactics and strategies amongst young mobile media users. *Big Data and Society, 5*(1), 1–12.

Shade, L. R., & Singh, R. (2016). 'Honestly, we’re not spying on kids': School surveillance of young people’s social media. *Social Media and Society, 2*(4), 1–12.

Taylor, E., Dove, E., Laurie, G., & Townend, D. (2018). When can the child speak for herself? The limits of parental consent in data protection law. *Medical Law Review, 26*(3), 369–391.

UNICEF. (2018). *Children's online privacy and freedom of expression: Industry toolkit*. New York: UNICEF.

van der Hof, S. (2016). I agree, or do I? A rights-based analysis of the law on children's consent in the digital world. *Wisconsin International Law Journal, 34*(2), 409–445.

van Gool, E., van Ouytsel, J., Ponnet, K., & Walrave, M. (2015). To share or not to share? Adolescents' self-disclosure about peer relationships on Facebook. *Computers in Human Behavior, 44*, 230–239.

Westin, A. F. (1967). *Privacy and freedom*. New York: Atheneum.

Youn, S. (2008). Parental influence and teens' attitude toward online privacy protection. *The Journal of Consumer Affairs, 42*(3), 362–388.

1. The national curriculum sets subject standards for primary and secondary schools. All except academies and private or independent schools must follow it. [↑](#footnote-ref-1)
2. The national curriculum is organised into *key stages* (KS): early years – nursery and reception (age 3–5); KS1: years 1 and 2 (age 5–7); KS2: years 3 to 6 (age 7–11); KS3: years 7 to 9 (age 11–14); KS4: years 10 and 11 (age 14–16). [↑](#footnote-ref-2)