Navigation

Document Purpose

This document is to be used with the online CSEA Data Mapping Visualization, the online CSEA Data Good Practice Bank, and the online CSEA Data Ecosystem Roadmap to improve the awareness of challenges that exist for stakeholders working within online CSEA; pointing to other resources where potential solutions might be found.

Together, the four documents act as a foundation to simplify the complexity of the system and enable forward progress towards an ideal state which is depicted in the visualization.

Introduction

This document is a summary of the key challenges and barriers that influence how data is collected, used, shared, and translated into action within the online CSEA ecosystem. It builds upon the previous work conducted by Safe Online in their 'Data for Change' initiative.

It should be read alongside three associated documents:

1. Online CSEA Data Mapping Visualization
2. Online CSEA Data Good Practice Bank
3. Online CSEA Data Ecosystem Roadmap

Together, these four documents should provide a simple and broad overview of the state of the data ecosystem and a forward view on how to improve our use of data for better outcomes for children, survivors, and society.

The key challenges and barriers are not intended to be exhaustive, but rather a summary of the insights provided by the limited group stakeholders engaged so far. It is intended that the document will be regularly updated, capturing the challenges and barriers as the system inevitably evolves.

This document is intended to be used by all those connected to online CSEA through data, as it will require a collective effort to improve the standard of data practice within the online CSEA data ecosystem.

Methodology

The current challenges and barriers affecting how data flows throughout the online CSEA response system has been analyzed by conducting a PESTLE analysis, exploring the Political, Economic, Sociological, Technological, Legal and Environmental factors. This was conducted through desk-based research, a review of previous Safe Online 'Data for Change' materials, and complemented by eight stakeholder interviews. It is intended to provide a broad yet consolidated view of factors affecting the online CSEA landscape.

This document represents the challenges and barriers that were uncovered during this process, but we recognize that this list is likely not exhaustive. There will be other significant existing challenges and barriers that require effort and attention and therefore this document should be periodically revisited to include new additions.

Impacts on victims, survivors, and children

It is crucial to incorporate the needs of victims, survivors, and children into discussion and solutions to these challenges, in the first instance this involves understanding what impacts these challenges have. This requires further exploration through discussions with this community.

Key barriers and limitations associated with online CSEA data across the ecosystem

A summary of the PESTLE analysis drawn from previous work, stakeholder interviews and wider research.

The below list represents the captured challenges and barriers within each PESTLE category. Subsequent chapters will provide more information on each, including a brief description, what were the insights from stakeholders, and if applicable, where information on best practice can be found.

Key barriers and limitations		
Political	1.01	Lack of assimilation of accepted universal terminology into national legislation and policies.
	1.02	High variance in government regulations on internet safety.
	1.03	High variance in national initiatives for digital education.
	1.04	Low political attention and exceptional policy making.
Economic	2.01	Reputational risk from participating in OCSEA-response initiatives.
	2.02	High resource cost associated with processes for online CSEA data processing.
	2.03	Funding priorities are short-term, siloed, and reactive.
Societal	3.01	Data gaps within the system from children about their experiences.
	3.02	High variance in the methodologies applied to measure OCSEA.
	3.03	Lack of recognition of data bias.
	3.04	Cultural taboos with sexuality leading to underreporting.
	3.05	Inequity in data landscape across geographies, groups, and characteristics.
	3.06	Legal complexity and ambiguity create a tension between data privacy and sharing.
	3.07	Inconsistent, damaging, and victim-blaming terminology used in reporting products or the media.
	3.08	Lack of positive outcome reporting.
	3.09	Lack of feedback of outcomes to data source owners.

Techno-logy	4.01	Increasing data volumes resulting in unsustainable demand on resources.
	4.02	Inconsistent data storage standards across stakeholders and countries.
	4.03	Inefficiency (inc. duplication) in technical activity across datasets and stakeholders.
	4.04	Inconsistent classification and data attribution between datasets.
	4.05	Limited interoperability between online platform datasets and data infrastructure.
	4.06	Access restrictions for more granular, disaggregated data
	4.07	Rapid evolution of technology reduces resilience to new offending techniques.
Legal	5.01	High costs associated with legal work.
	5.02	High variance in legal standards across jurisdictions.
	5.03	Ambiguity in data privacy and protection laws rely on organizational interpretation.
Environ-mental	-	Although no environmental considerations were identified during the review of existing material, several considerations have been included in this section as prompts for further discussion.

Political

Data within online CSEA ecosystem faces three types of political challenges:

1. How policies are made and subsequently communicated;
2. How governments decide to implement internet safety regulations; and,
3. How governments decide to implement national-scale digital education.

Government actions and policies play an important role in the use and sharing of data but high variance in national approaches has led to challenges and barriers that hinder international collaboration and weaken stakeholder data sharing.

Developing policies that are based on evidence, rather than in reactionary or superficial, whilst ensuring the inclusion of **acceptable universal terminology into national legislation and policies** is critical for an effective, credible, system response.

The regulatory response is currently fragmented internationally, with **high variance in national regulations on internet safety**. This undermines efforts to mitigate or respond to an international and highly dynamic threat by creating unnecessary complexity.

There are also significant differences in the levels of awareness, education or tools available to different demographics between nations. A **high variance in national initiatives for digital education** contributes to this, with opportunities unequally distributed putting some children at greater risk than others.

Detailed list of barriers and limitations

We found repeated themes within the stakeholder interviews, in most cases supported by further research, of challenges and barriers related to political influences.

1.01

Lack of assimilation of accepted universal terminology into national legislation and policies.

There are significant differences in how governments define and talk about online CSEA within their national legislation and policies.

Why is this an issue?
A lack of standardized and widespread adoption of agreed terminology into national legislation or policies can create legal or jurisdictional barriers by complicating frameworks. It can impact policy development, as policymakers rely on clear, universal terminology to create effective change. It also makes it difficult to promote consistent messaging and education about online CSEA, particularly across borders in response to a borderless crime.

What are stakeholders reporting?

- Governments use different terminology to discuss the same issues, even for basic concepts. In some cases, terms for types of offending do not even exist in other languages.

How could this be mitigated through better data practices?
Establishing a stronger **data governance** model for the online CSEA ecosystem would help establish norms or guidance around the terminology presented to, and used by, policy makers internationally. It would harmonize advocacy narrative based on data and simplify the access to this knowledge and evidence base for policy makers. This would be supported by improved **data analytics and storytelling** capability across organizations.

1.02

High variance in government regulations on internet safety.

There is an inconsistent landscape of internet regulation across the globe, with some countries having developed regulations and others having none.
Why is this an issue?

Online CSEA is a borderless threat, and without a coherent global response to regulation it can create challenges of data protection, managing compliance, collaboration, and consistent data practices. Having different legal requirements placed on industries associated with online CSEA creates legal complexities when data crosses jurisdictional boundaries, hindering collaboration between countries or stakeholders. Complying with multiple regulations is resource intensive and complex for industry, potentially detracting time and value from other safety-related activities. Different regulations can also stipulate different types of data collection, creating an unequal data landscape to analyze and draw conclusions from.

What are stakeholders reporting?

- In areas where regulations are now being implemented, the requirements imposed on industry sometimes fall below the standard of transparency that they were operating at in a self-regulated environment.

How could this be mitigated through better data practices?

By creating a unified voice to influence government regulations as they are developed, systemwide **data governance** models can have more power in pursuing change versus multiple independent messages. Establishing accurate methods for **performance monitoring and reporting** allows the ecosystem to identify regulatory interventions that have impact. This is supported by strong **data analytics and storytelling**. Having the capability to inform regulatory bodies on topics such as **data ethics**, **data security**, or **data privacy** will improve constructive dialogue when balancing regulation between parties.

1.03

High variance in national initiatives for digital education.

There are large variations in the prevalence and quality of initiatives aimed at raising awareness or educating populations on online CSEA.

Why is this an issue?

A lack of awareness or education about online CSEA (or digital literacy in general) will result in lower reporting levels. This will affect global data collection standards, providing skewed results with some regions under-represented. It will also impact levels of public support and advocacy for combatting the threat, in turn impacting government support for policy change or international cooperation.

What are stakeholders reporting?

- This issue is exacerbated by a generational divide between caregivers and children[1], this is felt particularly acutely in LMICs.

How could this be mitigated through better data practices?

Having stronger **data governance** will create more cohesion in messaging around key digital education guidance and will create spaces for educational organizations to make data-driven decisions in their initiatives. An improved understanding of **data training and awareness**, **data ethics**, and **data protection** will support organizations in maximizing public participation.

1.04

Low political attention and exceptional policy making.

The threat of online CSEA to society is under-represented in governments, resulting in reactive policy-making that neglects to consider evidence in its rationale.

Why is this an issue?

Developing policies that aren't entirely evidence-based represent missed opportunities for effective intervention and innovation. The interventions may not align with the needs and experiences of victims or survivors, limiting their effectiveness. This could undermine trust between governments, organizations involved in combating online CSEA, and the public. Without adopting evidence-based policymaking could also stifle innovation, including the adoption of new technology or methodologies for data collection, analysis and intervention.

What are stakeholders reporting?

- Online CSEA is a culturally sensitive topic, which can result in denial or ignorance of the scale or proximity of the problem within some governments. This hinders funding or data collection efforts.
- Social services are generally underfunded, stripping vital infrastructure for supporting children and victims. In some LMICs, there is a large reliance on NGOs to provide what would be considered elements of social services in more developed countries.

How could this be mitigated through better data practices?

[1] DH-data-insights-9-151223.pdf (safeonline.global)

Establishing systemwide **data governance** will encourage practices that increase accountability of all stakeholders, including governments. It will help provide oversight on how online CSEA is addressed in the media or in policymaking. Improved **data analysis and storytelling** will help create compelling messages to increase public support and political attention. These messages will be supported by improved methods for **performance monitoring and reporting.** Finally, **data sharing** will help create comprehensive, multi-source views of the problem that accurately reflect the scale and pervasiveness of the threat.

Economic

Economic challenges and barriers include those that prevent increased data sharing, the cost of data analysis, and the methodology in which funding is distributed across the system. Overall, there is a persistent theme (represented here but also throughout the other sections) that **there is a tension between safety and profit for parts of the ecosystem. This results in a system designed around a compromise at its heart.**

The **reputational damage from hosting or facilitating online CSEA is limiting the collaboration of some countries and industry** with other stakeholders, this is perpetuated by historical poor accountability reporting practices that have misrepresented industry data.

When collection or sharing does occur, the intensive **manual processes for reviewing, tagging, and qualifying online CSEA data are costly** – resulting in demand outweighing capacity and a reduced victim safeguarding effectiveness.

At a system level, **funding priorities are often short-term, siloed, and reactive** resulting in less emphasis on long-term strategic projects for those that deal with short-term solutions or insights. This lack of coordination can lead to duplication of effort across stakeholders, wasting resources from the system.

Detailed list of barriers and limitations

We found repeated themes within the stakeholder interviews, in most cases supported by further research, of challenges and barriers related to economic influences.

2.01

Reputational risk from participating in CSEA-response initiatives.

The significant reputational damage from online CSEA to an organization affects their transparency and cooperation with other stakeholders.

Why is this an issue?

The risk of reputational and financial damage can hinder organizations from participating in counter-CSEA initiatives, including research, cooperation with law enforcement, or industry best-practice schemes. It can result in costly, resource-intensive legal proceedings involving regulators to obtain information and creates a culture of distrust within the system. This barrier is exaggerated with poor reporting practices from research or media, where platform data is skewed or misrepresenting the prevalence of the problem – particularly for platforms who have provided data versus those who have not.

What are stakeholders reporting?

- Further discussion required.

How could this be mitigated through better data practices?
Improved organization data literacy engenders trust between stakeholders, as they perceive less risk of inappropriate data use or mishandling. Increased capability in **data architecture**, **data security**, **data protection**, **data resilience**, and **data training and awareness** are particularly key. Stronger systemwide **data governance** can provide a critical role in facilitating forums for **data sharing.** Improved **data ethics** and **data analytics and storytelling** reduces the risk of inappropriate use of data in reporting results or conclusions.

2.02

High resource cost associated with processes for reviewing, tagging, and qualifying online CSEA data.

There are significant cost implications for the time and effort spent on analyzing the huge volumes of online CSEA data.

Why is this an issue?
Relying on manual processes for a repetitive and distressing activity will result in increased mental health challenges for analysts. It also increases the response time for safeguarding victims, including removing known content, leading to increased chance of re-victimization. Manual processes are costly, introduce human error, stifle innovation, and have limited scalability. This is exacerbated when considering the increasing volume of online CSEA data being referred to analysts.

What are stakeholders reporting?

- The time and resource intensive process of collecting data in such a way that meets all stakeholder needs is resulting in a reluctance from industry to comply. In some instances, this is related to the way in which their data collection methodology aggregates data too far and too early.

How could this be mitigated through better data practices?
The need for robust infrastructure, methodology, and advanced skills is largely proportionate to incoming data volumes. Having scalable **data architecture** is now essential to provide future resilience. Improving the data literacy and capability through **data training and awareness** of teams will allow organizations to harness modern techniques such as AI in processing. Having efficient **data integration and interoperability** will allow partnerships to endure

increased data volumes without becoming burdensome. Finally, systemwide **data governance** will promote equitable access to improved processing power or AI-assisted methodologies.

2.03

Funding priorities are short-term, siloed, and reactive.

Funding across the system is complex, often based on relatively changeable immediate priorities, sometimes resulting in duplication or misplacement of effort.

Why is this an issue?
Short-term funding creates uncertainty and instability, making it difficult to plan and implement long-term data collection and analysis. It affects the quality of generated evidence, increasing the likelihood of compromised methodology, not reaching marginalized or hard to reach groups, ultimately contributing to further inequality. With funding generally not focused on capacity building, there is a relative lack of the necessary capacity for effective data management within the system. The siloed nature of funding streams results in duplication of activity between stakeholders, wasting resources devoted to online CSEA and leading to a fragmented data landscape.

What are stakeholders reporting?

- The research within online CSEA is often made up of numerous short studies, these are considered to be less impactful than longer-term, repeated studies. This creates a lack of understanding of the long-term impact of online CSEA, or the long-term benefits of online CSEA interventions.

How could this be mitigated through better data practices?
Improved system wide **data governance** will improve cohesion and collaboration, creating opportunities to create more comprehensive assessments of the impact of interventions. It will also streamline funding, reduce deduplication, and encourage partnerships with other sectors to access further funding sources. This will be supported by stronger skills in **data analytics and storytelling** to effectively demonstrate impacts from funding.

Societal

Data within online CSEA ecosystem faces three types of societal challenges:

1. Broad societal challenges that are applicable to data in many types of ecosystems;
2. data specific challenges that are applicable to any ecosystem reliant on data, particularly those that involve children; and
3. Online CSEA-specific challenges that are likely unique to this system.

Broad societal challenges that impact the accessibility and quality of data, include **cultural taboos surrounding sexuality**, the ongoing debate between **data privacy versus security**, and the **under-representation of the Global South and non-English speaking countries**.

In addition to those macro-effects, there are further issues related specifically to data, these include **recognizing data bias** during collection and reporting, a lack of **data from children about their experiences**, a lack of **feedback of outcomes to sources**, and the ethical nuances arising from **varying definitions of abuse**.

The final barriers and limitations are those associated with processes or stakeholders within the online CSEA ecosystem, these include a **lack of positive outcome reporting** within end-products or the media, and an inconsistent, **damaging use of terminology** across the ecosystem.

Detailed list of barriers and limitations

We found repeated themes within the stakeholder interviews, in most cases supported by further research, of challenges and barriers related to societal influences.

3.01

Relative lack of data within the system from children about their experiences.

There is an imbalance between the amount of data within the ecosystem that comes from children versus data that is about children.

Why is this an issue?
A lack of firsthand data from children hinders a full understanding of the prevalence and dynamics of online CSEA and without their input, interventions may not be effective in addressing their needs. As children are the primary data source for understanding scale and severity, an imbalance may be a symptom of systemic underreporting, leading to an underestimation:

- Of scale
- Nature of threat

- Risks or protective factors
- The short and long-term impact to children
- The effectiveness of interventions
- New trends and developments within online CSEA

Children's voices are essential for their empowerment of an issue where children are the victims, advocacy for this issue suffers without robust data from children. However, this is complex and nuanced, as a child's understanding or recognition of abuse varies with age or developmental milestones.

What are stakeholders reporting?

- There are gaps in the understanding of the survivor or victim online experience, particularly in the Global North. Within victim and survivor categories, the data should be broken down into sub-categories to allow sufficient insight – recognizing the unique needs of victim and survivor categories.
- It is labor intensive for helpline and hotline networks to share case study data from children, reducing the likelihood of it happening.
- Relying on hotline data to estimate the scale of the problem denies children the opportunity to share their experiences in other ways, underestimating the problem. It also reduces the data quality as often results in less data points about a child.
- Involving youth councils on how to frame questions, give appropriate definitions, to raise awareness and maximize young people's engagement who might not otherwise be able to identify abuse.

How could this be mitigated through better data practices?

Establishing systemwide **data governance** would mean there's a way of promoting best practice methodologies, including guidelines for the inclusion of children's voices in outcomes. Improved use of **performance monitoring and reporting**, with metrics focused on child inclusion would also be effective. Expanded knowledge on **data ethics**, **data protection**, and **data sharing** would also improve confidence, trust, or capability in handling or sharing data from children – rather than just data about children.

3.02

High variance in the methodology applied to measure abuse.

Abuse is defined and measured in different ways across the globe, these aren't always comparable and results in variance between datasets.

Why is this an issue?
Ethical and methodological challenges in measuring abuse leads to a lack of standardized definitions or methodologies, this means identifying global trends or patterns becomes problematic, hindering efforts to understand or address issues effectively. It also leads to gaps in understanding the true scale and severity of the problem, limiting the development of best practices and evidence-driven solutions. Overall, it weakens global advocacy efforts, without a unified approach it dilutes the impact and credibility of these efforts.

What are stakeholders reporting?

- Varying definitions between countries increases the workload of online platforms when considering what to report to hotlines, as they report everything that could be relevant in all countries they operate within (rather than tailored for specific countries or to an international standard). This increased volume is then passed onto hotlines who will do further distribution or processing.
- A lack of universal definitions ultimately hinders the ability to build a global level dataset on prevalence of online CSEA.
- Even countries with established and mature definitions of a child or abuse within legal frameworks have difficulty in standardizing cases involving 16 to 18 year olds.
- Assimilating, comparing, or building on academic studies is difficult due to the varying definitions of violence, ages, recall periods.
- Modern forms of offending are difficult for Helplines to classify, as often they could fit into multiple categories or tags, such as 'sexting' which could fit within online exploitation or within peer-to-peer relations.
- There isn't a formal global coordination mechanism between hotline and helplines to agree boundaries for reporting categories.

How could this be mitigated through better data practices?
Having common guidelines or sharing of best practice through system wide **data governance** would increase harmonization of methodologies and definitions. It could also improve the effectiveness of outputs, as these are likely to have increased alignment and therefore less risk of confusing, contradictory conclusions based on the same data.

3.03

Lack of recognition of data bias.

There is a general lack of data bias recognition within the reporting of outcomes or results.

Why is this an issue?
Clearly and explicitly acknowledging data bias within datasets and subsequent reporting is crucial for contextualizing results and addressing the limitations of datasets. Unidentified data bias can lead to inaccurate portrayal of the prevalence and characteristics of online CSEA. Furthermore, it can marginalize under-represented groups and erode trust in research, policy, and advocacy efforts leading to skepticism and resistance.

What are stakeholders reporting?

- Increasing the use of AI for extraction and analysis will perpetuate any existing bias within the AI software, on top of the bias caused by methodology.

How could this be mitigated through better data practices?
Improving the understanding of data ethics within organizations will help promote the recognition and communication of bias in data being used. Combined with systemwide and organizational data governance it will improve processes to identify ethical risk associated with data, including bias, such risk assessments, improved contextual capture during data collection, or transparency in outcomes about biases. Furthermore, improving organizational AI Governance is essential to protect against perpetuating data bias through AI-models and outputs.

3.04

Cultural taboos with sexuality leading to underreporting.

Societal attitudes toward sexuality differ between cultures impacting the ability to accurately reflect a global landscape.

Why is this an issue?
There is an enduring sensitivity and stigma associated with sexuality and sexual abuse that can prevent victims from reporting or sharing their abuse. This is nuanced, within low- and middle-income countries (LMICs) or among older children, there is an increased risk of parents or caregivers disagreeing with a child's wish to report due to the perceived shame it attaches to a family.

This can lead to underreporting, limited research participation, or reduced data quality. It can also result in different personal and legal definitions of what is considered abuse. Furthermore, it can create challenges for policy development,

where policymakers require public support in developing culturally sensitive, effective policies and interventions.

What are stakeholders reporting?

- It's estimated that there are relatively low levels of disclosure to child helplines for online CSEA compared to actual global prevalence levels. One significant factor is a lack of awareness from children that they have even been a victim of a crime.

How could this be mitigated through better data practices?
Communicating the limitations and data bias in data will improve the transparency and impact of outcomes. This capability will be strengthened through a stronger understanding in **data ethics** and harmonized across the system with **data governance**.

Under-representation of geographies, groups, and characteristics.

3.05

There is a weighted focus of effort towards the global north and English-speaking countries within the ecosystem.

Why is this an issue?
This leads to an incomplete understanding of the global prevalence and severity of online CSEA, potentially missing important regional nuances and differences. It can result in skewed statistics, over or under-representing regions with more mature data systems, or groups of children for whom more data is available. When interventions are designed, they are usually based on the evidenced needs of children in the Global North or English-speaking countries. When applying these to other regions, adaptation or contextualization is often not done in accordance (or done in the absence of) evidence or data. Overall, this perpetuates global inequalities and hinders a collaborative international response.

What are stakeholders reporting?

- There are strong reporting mechanisms for the EU, allowing the build-up of a highly contextualized EU online CSEA landscape, which in turn feeds into a relatively strong EU safety technology industry. There is a comparative lack of activity, mechanisms, or infrastructure in Asia, Africa, or South America.

- Advanced technological solutions for extracting or analyzing data based on free-text will most likely be based on the English language, limiting their use for the rest of the world.
- Helpline networks collect data in the language of their respective countries, however when this is shared it is aggregated and translated into English.
- Countries in the Global South have a higher proportion of data from children than global North; however, global policy is often set by those within the Global North.
- It is possible to conduct targeted, novel surveys aimed at under-represented study groups such as online CSEA within LGBT+ children and youth, where groups may have unique vulnerabilities or needs.
- The high resource and capacity cost to conduct household surveys and academic studies result in uneven coverage of statistics across countries.

How could this be mitigated through better data practices?
Raising the data capability in LMICs will improve their representation within the data ecosystem. This includes developing strong **data architecture**, establishing effective organizational **data governance**, **data security**, **data protection**, **data integration and interoperability**, and **data training and awareness**. This would encourage collaboration and **data sharing** between LMICs and the rest of the world. Improving this data capability could be achieved through system wide **data governance**, with a focus on improving data literacy or capability sharing.

Legal complexity and ambiguity create a tension between data privacy and sharing.

3.06

The commonly adopted view of a compromise between data security and privacy can stifle collaboration and innovation, result in a complicated legal landscape, and put victims at greater risk.

Why is this an issue?
A careful balance between protection and access is required for the highly sensitive or potentially illegal data consumed within the online CSEA data landscape. Excessive focus on privacy can prevent sharing and stifle collaboration or innovation by creating data silos, conversely sharing without appropriate security controls can increase risk of data loss or misuse. Both scenarios will increase risk to victims and survivors. Varying data protection regulations

between nations creates trust and compliance issues and challenges in integrating datasets from diverse sources.

What are stakeholders reporting?

- Proactive sharing of data on individuals (for example, prior to warrant being issued) is difficult for industry, particularly in the US, as it may contravene the 4th amendment protecting US citizens from unreasonable searches by the government. This can also impact hotline or helplines, as there is a technical and operational requirement to have separate systems for data to protect themselves against data privacy laws.

How could this be mitigated through better data practices?

Creating a repeatable legal basis for data sharing, possibly through standardized legal frameworks, will simplify data sharing amongst stakeholders. This requires strong trust between stakeholders, often facilitated by an independent organization. There may also be technical solutions for independently stewarding data. Appreciating data ethics and data protection assists making the most appropriate decision for positive child-outcomes.

Inconsistent, damaging, and victim-blaming terminology used in reporting products or the media.

3.07

There is an inconsistent standard for the use of victim-blaming, environment specific, or gender-based terminology in reporting products or media.

Why is this an issue?

A sub-standard and inconsistent reporting standard for the outcomes of research or reports to the public can be misleading, stigmatizing, or underreporting. Using inappropriate terminology can misrepresent the severity of the impact on survivors or mitigate blame on offenders, shaping public perception of the issue and perpetuating harmful stereotypes, ultimately reducing public support and reducing the likelihood of children and survivors to want to participate in online CSEA efforts. It can also create friction for data sharing by contributing to a culture of mistrust.

What are stakeholders reporting?

- Even at the highest level of international cooperation, damaging terms associated with CSAM are still being perpetuated and debated during the creation of protective treaties.
- Within law enforcement, the inconsistency of terminology has affected their operational effectiveness during cooperative operations or other collaborations.

How could this be mitigated through better data practices?
Establishing system wide data governance would help stakeholders align around a consistent set of terminology to use in media reporting or more widely, helping shape public perception of online CSEA, offending, or survivors.

3.08

Lack of positive outcome reporting.

Most of the reporting surrounding online CSEA describes a worsening situation, with positive outcomes overlooked.

Why is this an issue?
A consistent negative theme (e.g. increase in prevalence, new offending vectors) can be demoralizing to stakeholders within the system, disengaging for the public, and a barrier for further political support. It leads to perceived ineffectiveness of the system, potentially reducing funding and support for these programs. It doesn't promote a culture of sharing best practice between stakeholders, overlooking successful approaches and areas of progress.

What are stakeholders reporting?
- Further discussion required.

How could this be mitigated through better data practices?
Having increased capabilities data analytics and storytelling would improve organization's agility in their communications, including promoting positive stories from data. Supported by stronger systemwide data governance, which would advocate for cultural change through guidelines or best practice, and the improved ability in performance monitoring and reporting.

3.09

Lack of feedback of outcomes to data source owners.

Data usage is currently linear, rather than cyclical, with feedback on how data was used or what outcome it resulted in.

Why is this an issue?
Without being able to report on the outcome an organization's data resulted in, it is difficult to demonstrate the value of that data. Feedback is also crucial for learning and development, missing opportunities to adjust their data collection strategies to maximize outcome. Additionally, personal motivation will be affected if individuals are not informed of how their work has contributed to a collective mission against online CSEA.

What are stakeholders reporting?

- Analysts reviewing CSAM are often unable to know whether a child has been rescued, detrimentally impacting their mental health.
- A lack of feedback or interoperability between analysts and law enforcement means it is not always possible to know if queued suspected-CSAM content is part of an active case, and therefore should be flagged for urgent review.
- Industry don't have a clear sight on how the data they provided was used by law enforcement or other organizations, reducing their ability to report on outcomes of their effort and demonstrate value.

How could this be mitigated through better data practices?
Having the systemwide structures that facilitate feedback on data and outcomes are part of **data governance**, with one its primary aims to continuously improve the way the system is using data. Understanding more equitable methods for **data sharing**, or building feedback into **data integration and interoperability** structures, improves the system's ability to communicate about data and impact.

Technological

Data within online CSEA ecosystem faces two types of **technological challenges**:

1. Broad technological challenges that are applicable to data in many types of ecosystems; and
2. challenges that are exacerbated or unique to the online CSEA ecosystem

Broad societal challenges that impact the accessibility and quality of data, include **increasing data volumes** creating an unsustainable demand, **inconsistent data storage standards** across stakeholders and countries, and the **rapid evolution of technology** reducing resilience to new offending techniques.

Challenges that are particularly apparent within the online CSEA ecosystem are predominantly due to a lack of cohesion or compatibility between datasets or stakeholders. This includes the **inefficient duplication of technical activities or development** between datasets, **inconsistent classification and data attribution**, **limited interoperability between online platforms**, or the **lack of microdata** within shared datasets.

Detailed list of barriers and limitations

We found repeated themes within the stakeholder interviews, in most cases supported by further research, of challenges and barriers related to technological influences.

4.01

Increasing data volumes result in unsustainable demand on resources.

The amount of data associated with online CSEA is rapidly increasing, requiring more processing effort from the response system.

Why is this an issue?

It is difficult to develop the capacity to effectively process the increasing amount of data being ingested by the system, overwhelming systems and personnel involved in the response. Whilst there are higher volumes of data, the quality of data is not necessarily increasing, resulting in it becoming more difficult to find useful data amongst the volumes. There are demanding technical and skill requirements for analyzing big datasets that are often limited to private industry, making it hard for the other stakeholders to use this data effectively. Large data volumes have led to bottlenecks in the system, resulting in delays to safeguarding.

What are stakeholders reporting?

- Demand is outstripping capacity within law enforcement organizations, with some technical systems not being utilized to their maximum capacity.

How could this be mitigated through better data practices?

The need for robust infrastructure, methodology, and advanced skills is largely proportionate to incoming data volumes. Having scalable **data architecture** is now essential to provide future resilience. Improving the data literacy and capability through **data training and awareness** of teams will allow organizations to harness modern techniques such as AI in processing. Having efficient **data integration and interoperability** will allow partnerships to endure increased data volumes without becoming burdensome. Finally, systemwide **data governance** will promote equitable access to improved processing power or AI-assisted methodologies.

4.02

Inconsistent data storage standards across stakeholders and countries.

There are no universal standards for storing online CSEA data amongst stakeholders.

Why is this an issue?

A lack of universally agreed standards reduces compatibility between datasets, making it difficult to integrate and share data effectively. It reduces overall accessibility of data sets and coordination and collaboration in the system response – creating operational inefficiency with individual, rather than collective, data sharing agreements. Overall data quality and integrity is reduced across the system, as a lack of standards can introduce inaccuracies or inconsistencies. The sensitive nature of this data also means it is at an elevated risk of security vulnerability.

What are stakeholders reporting?

- Further discussion required.

How could this be mitigated through better data practices?

Improving broad quality of data literacy across the system empowers individual organizations to adopt better standards of data storage. This could be done through highlighting best practice, sharing guidance, efforts to harmonize data storage standards through **data governance**; improved **data architecture** and **data integration and interoperability** to create the technical infrastructure that meets all stakeholder requirements; or improved **data security** and **data**

protection reducing the risk of data being mishandled or inappropriately stored; or through the implementation of better collaboration through data sharing.

4.03

Inefficiency (including duplication) in technical activity across datasets and stakeholders.

Activity can be duplicated between stakeholders and the extent to which this is happening is not known.

Why is this an issue?
Duplication of activity wastes resources that could be engaged in other, unique, efforts. It also adds to the complexity of the data landscape, with reporting or analytical output being based on similar data but drawing differing conclusions. It also makes it more likely that the system is producing a fragmented view of the problem, as coordination is essential for a comprehensive view and response.

What are stakeholders reporting?

- Without a formal community of practice or method for sharing intelligence, there are organizations performing similar roles but without coordination. Varying mandates between helplines and hotlines make processes and focuses different.
- There is limited matchmaking to help stakeholders find what they need. Facilitation across the ecosystem should raise the visibility of user challenges, and connect them to relevant intermediaries.
- There is limited support around governance issues. While many intermediaries specialize in the technical aspects of data access, processing and analysis, there are fewer equivalent supports on the governance side.
- There is limited tracking of the use of tools, guidance, and other materials. There is a disconnect between the expressed needs for support among less experienced practitioners and the many tools, guidance materials, and principles that exist. It would be valuable to better understand if these tools are reaching their intended audiences and meeting their goals.

How could this be mitigated through better data practices?
Data governance will improve coordination across the system, increasing efficiency by creating space for knowledge sharing, improved visibility of the

impacts of interventions, awareness of ongoing work, and information for donors on under-developed areas within the ecosystem.

4.04

Inconsistent classification and data attribution between datasets.

There is no universal schema, or other meta-data methodology, consistently employed at scale across the system.

Why is this an issue?
Differences in classification and data attribution methodology between organizations or countries makes it difficult to amalgamate databases of CSAM. This is often occurring due to differing laws within each country on the legal definitions of abuse. Without consistency it can increase resourcing cost due to duplication of effort required in classifying the same image multiple times. It also reduces operational efficiency and effectiveness, with organizations unable to check unknown images against a single, global, repository of hashes. It also results in varying levels of trusted databases, due to tagging methodologies providing higher levels of reassurance in some countries compared to others.

What are stakeholders reporting?

- Interoperability is significantly reduced by different hash sets, classification schema, and meta-data tagging methodologies.

How could this be mitigated through better data practices?
Establishing a universal terminology and classification schema relies on a harmonized system. Systemwide **data governance** would advocate for this, providing guidance or principles that would increase momentum towards a more harmonized system. Increased **data sharing** would organically begin to align terminology closer, as sharing would rely on certain levels of **data integration and interoperability**.

4.05

Limited interoperability between stakeholder datasets and data infrastructure.

The structure, content, and format employed by stakeholders when developing datasets is unstandardized, limiting interoperability.

Why is this an issue?
Lack of standardization between stakeholders creates a fragmented view of the system. This is particularly important given the cross-platform offending nature relevant to online CSEA; stakeholders within the system lose time manipulating data into coherent datasets from multiple platforms.

What are stakeholders reporting?

- The proliferation of different systems with different formats for data is a universal problem whether this is within or across countries The need to standardize terminology, data structures and API formats is universal.

How could this be mitigated through better data practices?
Increased efforts in **data sharing** would facilitate the need for certain levels of **data integration and interoperability**. Becoming more innovative in the ways data is shared, adopting models used in other ecosystems would make the case for increased focus on improving integration and interoperability. **Data governance** would also be an advocating voice for this, particularly with international bodies to generate political will for change. Any movement towards improved data integration and interoperability would need to be supported by good **data architecture** between sharing organizations.

4.06

Access restrictions for more granular, disaggregated data.

Data is often shared as an aggregated set of results, with the microdata being unavailable for further analysis (or re-analysis) to validate results or understand bias.

Why is this an issue?
Aggregating data and not retaining microdata for sharing reduces the granularity of results, potentially obscuring details and patterns for others to find. This might perpetuate or hide bias or miss an opportunity to highlight effects on under-represented groups. It limits analysis in general, as many statistical techniques aimed at identifying trends, risk factors or correlations often require microdata to be available. Overall, it reduces flexibility for exploring hypotheses

and reduces transparency in the shared dataset, making it difficult to compare trends or patterns over time, particularly for sub-groups.

What are stakeholders reporting?

- Anonymized or aggregated datasets make it hard for law enforcement to find individuals.
- Shared material from industry lacks richness in its meta-data, for example, it often does not include IP address, material traded, or additional context from the individual's account.
- Reported material from industry to law enforcement is often a report of a single instance, even when the same account is associated with multiple reports on that platform's system.
- Often share aggregated data rather than raw data, would like to implement ML on this data to allow it to be shared (IJM)
- Raw data is closely guarded, with concerns about individual safety or inappropriate re-use of data being cited as the most common reasons.

How could this be mitigated through better data practices?

Best-practice methodologies and approaches to data collection and sharing shared by a systemwide **data governance** model would help unify the system in understanding what is possible to share between stakeholders, and what the benefits of doing so might be. Subsequently, having appropriate **data security** and **data protection** is essential to protect that data. This is coupled with knowledge of how to appropriately handle that data, increasing capabilities in **data ethics** and **data training and awareness**.

4.07

Rapid evolution of technology reduces resilience to new offending techniques.

Advances in technology are creating new offending vectors and behaviors, which the system tries to keep pace in its understanding and response.

Why is this an issue?

The technical landscape in which online CSEA is occurring is evolving, becoming more complex and increasing in scale. This can render preventative or responsive techniques obsolete. The associated data is being captured from an increasingly diverse set of sources, with increasingly higher levels of privacy. This can cause a lag for the system to respond.

What are stakeholders reporting?

- The process for getting new data types agreed within helpline and hotline networks is slow, often being outpaced by technological change.
- Data on emerging technology trends is lacking from a global perspective, with live streaming being cited as an example of this.

How could this be mitigated through better data practices?

Strong systemwide **data governance** would provide guidance and deduplication across innovation and research activities, improving the cost effectiveness of funding for responding to future threats. It would also promote global coverage of issues, supporting LMICs in their efforts to access technological advancements. Furthermore, improved **data training and awareness** assists those maximize the value from data or use it in innovative ways.

Legal

Detailed list of barriers and limitations

We found repeated themes within the stakeholder interviews, in most cases supported by further research, of challenges and barriers related to legal influences.

5.01

High costs associated with legal work.

Legal costs are disproportionately high in the implementation of data sharing, reducing impact and discouraging collaboration.

Why is this an issue?
Sharing data related to child sexual abuse across organizations often involves navigating a complex web of legal requirements, including data protection laws (e.g. GDPR), confidentiality agreements, and child protection laws. Ensuring compliance with these legal frameworks requires substantial legal expertise, which can be expensive. Organizations, particularly smaller ones or nonprofits may struggle to afford the necessary legal counsel to ensure they are sharing data legally and safely. Now, most data sharing agreements are bespoke and non-standardized, requiring concentrated legal effort and ongoing support. This removes funds from other critical areas within an organization, potentially reducing overall impact on online CSEA, or discouraging organizations from engaging in data sharing at all.

What are stakeholders reporting?

- Further discussion required.

How could this be mitigated through better data practices?
Introducing and promoting standardized, repeatable basis for legal **data sharing** through better system wide **data governance** would increase the access and opportunities for sharing data between stakeholders.

5.02

High variance in legal standards across jurisdictions.

The high variance in legal standards across jurisdictions significantly complicates the global response to and prevention of child sexual abuse.

Why is this an issue?
Different definitions of abuse, varying legal protections, and inconsistent enforcement practices can create gaps in safeguarding children, making it difficult to establish a universal standard for what constitutes abuse and how it should be addressed. This inconsistency hampers international cooperation, particularly in cross-border cases where conflicting legal frameworks can delay or obstruct justice. Data sharing is also hindered by varying privacy laws, making it challenging for organizations and authorities in different countries to collaborate effectively. Moreover, global initiatives aimed at establishing uniform protections often falter when individual countries struggle to align with international standards due to their unique legal landscapes and resource limitations. For victims, these disparities can lead to uneven access to justice and support, further complicating their ability to seek redress and protection, and potentially leading to re-victimization as they navigate disparate legal systems.

What are stakeholders reporting?

- Further discussion required.

How could this be mitigated through better data practices?
Through improved data governance it is possible to establish the infrastructure to facilitate independent promotion of data sharing between stakeholders, enabling the setting of guidelines (or minimum criteria) to promote trust between partners despite high variability. There are also innovative data sharing practices that mitigate high variability in legal differences between partners.

5.03

Ambiguity in data privacy and protection laws rely on organizational interpretation.

Ambiguity in data privacy and protection laws, coupled with the increased reliance on organizations to interpret these laws, creates significant challenges in how data is handled and shared.

Why is this an issue?
When laws are unclear, organizations often adopt varying interpretations, leading to inconsistent practices that can result in critical information being withheld or not shared promptly, undermining efforts to protect children. The fear of legal repercussions may cause organizations to be overly cautious, leading to under-sharing of essential data, which can hinder coordinated efforts between agencies. The need for organizations to seek legal counsel or internal approvals due to legal uncertainties can also cause delays, which are particularly harmful in

cases where timely data sharing is crucial for preventing abuse or intervening in ongoing cases. This combination of factors results in missed opportunities for early intervention, weakening the collective response to child sexual abuse and potentially allowing harm to continue.

What are stakeholders reporting?

- Further discussion required.

How could this be mitigated through better data practices?

Through improved **data governance** it is possible to establish the infrastructure to facilitate independent promotion of data sharing between stakeholders, enabling the setting of guidelines (or minimum criteria) to promote acceptable agreements between partners. There are also innovative **data sharing** practices that leave room for interpretation and degrees of flexibility in areas such as data control or privacy.

Environmental

Although no environmental considerations were identified during the review of existing material, several considerations have been included below as prompts for further discussion.

Environmental considerations impact stakeholders working to improve data collection, storage, sharing, and use in response to online child sexual exploitation and abuse.

- Data processing power is a critical factor, with unequal access to energy creating disparities in the ability to manage and analyze large datasets.
- The high cost of data storage further compounds these challenges, particularly in resource-constrained areas. For example, access to cooling water is currently necessary in data storage.
- Climate-related humanitarian crises often expose weak technical infrastructure, reducing capacity for child safeguarding efforts and making it harder to protect vulnerable populations.
- Climate-displaced populations, such as refugees, face increased risks of exploitation due to their migratory status and lack of basic services.
- The absence of comprehensive data about children affected by the climate crisis, including climate refugees, limits the ability to develop targeted interventions to protect them from exploitation.
- Natural disasters exacerbated by climate change can damage data centers, or other technical infrastructure, and disrupt connectivity, reducing the availability of crucial child safeguarding systems during emergencies.
- A lack of funding for sustainable data management practices - such as optimizing storage and using green energy – is leading to reduced resilience to environmental shocks.

Conclusion

The preceding sections have discussed the challenges and barriers associated with the collection, sharing, and use of data within the online CSEA ecosystem. Many of those can be attributed to several *major root cause challenges* facing the ecosystem, these are presented below.

The absence of strong governance in the system leads to significant inconsistency and variability, undermining the effectiveness of protective efforts. This lack of coordination affects various critical aspects, including the balance between data ethics - where the rights to privacy must be weighed against the benefits of sharing information - and the risk of duplicating efforts due to poor coordination. It also creates conditions whereby inconsistencies in terminology used to describe abuse and report it to society can flourish, while missing opportunities to identify shared challenges or solutions among stakeholders. Additionally, a lack of governance fails to promote equal access to tools, guidance, and materials that could strengthen the system and increase overall capacity. Finally, stakeholders are more isolated and siloed in the current state, leaving the ecosystem exposed to symptoms such as lack of feedback between stakeholders on outcomes from sharing data, missing opportunities to improve processes or improve relationships.

High variance and inconsistencies across the global system results in inefficient effort, stakeholder friction, and a weakened political will for change. These disparities are evident in areas such as data sharing and online safety regulations, political commitment to addressing online CSEA, data privacy and protection laws, stakeholder data literacy, and the definitions and methodologies used to measure abuse. Consequently, the increased effort required to develop solutions, combined with the friction between stakeholders, undermines the effectiveness of collective actions and weakens the overall response to online CSEA, highlighting the need for greater global alignment and coordination.

This, in part, **leads to low levels of trust between stakeholders, reducing collaboration and the quality of data outputs.** The high variance in methodologies, laws, and organizational motivations further exacerbates this trust imbalance. Stakeholders are often uncertain whether shared data will be managed and used appropriately due to varying levels of technical capacity and differing data management standards. Additionally, the datasets that are shared are often overly aggregated, lacking in microdata, which limits the ability to verify or validate results and recognize biases. The absence of contextual data, such as metadata, further hinders effective analysis by failing to provide crucial information on collection methodologies, biases, and limitations.

Where laws on data protection, privacy, and sharing exist, they often require significant interpretation by organizations, leading to high legal costs and reducing

the motivation to collaborate between stakeholders. This interpretation is affected by fear-based public narratives that drive further sensitivity or bias in weighing rights and action. This results in increased caution, resulting in the under-sharing of crucial data, causes delays and unnecessarily long processing times, and prevents technical collaboration and innovation to overcome challenges.

The technical landscape is fragmented, with a significant lack of interoperability between stakeholder systems. This fragmentation is exacerbated by high variance in data storage standards, technical capacity, and classification and attribution methodologies. Even within sectors, there are considerable differences in how data is collected and stored. The absence of technical solutions employed within the system, such as APIs, further complicates data sharing. This situation is particularly problematic given the increasingly cross-border and cross-platform nature of online CSEA threats, making seamless and effective data exchange more critical than ever.

All these challenges require coordinated, concentrated effort from each stakeholder in the system. However, **the system is characterized by an uneven distribution of funds, resources, and capacity – particularly technical and legal – which are becoming overwhelmed by increasing data volumes.** This is exacerbated by the capacity and capability of the system being unevenly distributed across the globe, resulting in an incomplete understanding and inefficient response to the threat of online CSEA.