WeProtect GLOBAL ALLIANCE

GOUVERNEMENT
Liberté
Égalité
Fraternité

Safe Online

Safe Digital Futures for Children

2023 Data for Change Technical Workshop

BRIEFING NOTE

Over 50 experts convened on 8-9 November 2023 in Paris on the margins of the Paris Peace Forum to take stock of progress on the online child sexual exploitation and abuse (CSEA) data ecosystem, and jointly plan what's next. This is part of the **Data for Change Initiative** led by Safe Online in partnership with key actors including WeProtect Global Alliance, the EU Parliament in 2022, and the French Government in 2023.

> *The lack of systematic and harmonised data collection and infrastructure for robust analysis of online risks and harms for children in the context of other forms of violence and across levels remains a critical barrier to securing a safe internet for children -Safe Online*

This briefing outlines key actions to take the community forward and discusses emerging findings to **strengthen the data ecosystem** and support effective advocacy, practice and solutions at global, regional and national levels. The workshop emphasised creative data sharing and interdisciplinary collaboration. Acknowledging the challenge of rapid technological developments, the community prioritised filling data gaps, refining definitions, and aligning funding priorities. The emphasis was on harmonising data taxonomies, capturing children's experiences more robustly, and building interoperable data systems.

Where did we start?

One year ago, the initial gathering identified five focus areas: Language, Evidence, Systems, Advocacy, and Networks. Efforts were coordinated around these areas with short-term priorities to move the community forward – for more information see the 2022 Workshop briefing note.

Quality data is essential to inform and ensure the quality and impact of activities across levels. We need better and reliable data and collection methods to build a more comprehensive understanding of the threats, and we also need collaboration, capacity and sustained political support to ensure data is used effectively. There is an immediate need to improve capacity, transparency, responsible data use and investments in data efforts - 2022 Data for Change Workshop

What did we come away with? And what's next?

What we are building as a data community is at an inception stage, and it will grow slowly before we see large-scale impact in our work. We can start with some of the following key takeaways:

Holistic View of the Data Ecosystem

Learning: The call for a more holistic and collaborative approach to the data ecosystem underscores the importance of **creative ways of data sharing and interdisciplinary collaboration** rather than of creating new structures and new processes as a default. Sharing expertise, legal and ethical frameworks, technical infrastructure, relationships, etc. within the community can bolster data flows to combat online CSEA more effectively. **We need to shift the focus from potential risks of data sharing to the consequences of not sharing data.**

Actions:

- Conduct a more granular assessment of **current barriers to data sharing**.

- **Map the legal landscape** to identify barriers and facilitate sharing of established good practice of legal and ethical frameworks or agreements.

Address Data Gaps

Learning: We need to think about targeted ways to **fill data gaps and improve data quality.** While there is an overwhelming volume of data from a law enforcement perspective for example, there are also certain parts of the ecosystem that lack data – about certain phenomena, or marginalised groups of children. Efforts should intentionally **address under-represented geographies and groups** within the online CSEA data ecosystem.

Actions:

- Establish a **framework for coherent data governance**, prioritising areas lacking adequate data, such as marginalised groups of children or specific phenomena.
- Engage and learn from **related fields** such as GBV, terrorism, cybersecurity, etc.

Improve Terminology & Taxonomies

Learning: When creating or refining terminology, finding the **right balance** between being strong enough to capture the hideous nature of this crime and making language accessible is crucial. There is also an important reconciliation between rapidly changing trends with slower moving frameworks such as in criminal justice settings. Increased harmonisation in data taxonomies is greatly needed across countries, sectors, and organisations. **Promoting better data triangulation and interoperability** is essential.

Actions:

- Promote **child-centric approaches** to adapting or defining terms.
- Engage the **tech sector and criminal justice actors more in terminology** work.

Technology Development

Learning: The **fast pace of technology development** and its impact on accelerating evolution of the threat landscape is a challenge, but it doesn't need to hinder our efforts for data collection and generation. Technology can also be used or developed as tools and solutions to tackle challenges in the field as well.

Actions:

- **Evaluation of the technology tools** being used by actors globally for better understanding of these tools and increased interoperability with other systems.

Enhance Funding Alignment

Learning: Funding and resources set priorities in the field. **Donor alignment** is deemed critical to open the space for intentional and aligned work in combating online CSEA.

Actions:

- **Convene donors** to enhance understanding of the critical needs and potential mechanisms for adaptive approaches within the online CSEA data ecosystem.

Promote Responsible Data Use

Learning: Data is never neutral and is as good as the stories being told and who tells these stories comes with their own biases. This highlights the importance of **always keeping use cases in mind** whether that is providing services, advocacy whether through policy or research, or even responding to the public in explaining, justifying or contextualising data.

Actions:

- Collaboratively develop and apply **core data principles** for responsible data generation and use within the community.
- We must ensure **children's and survivors' voices** in approaches, including the need to prioritise **lived experience experts' inclusion** across the data lifecycle.

What have we learned about the current data ecosystem?

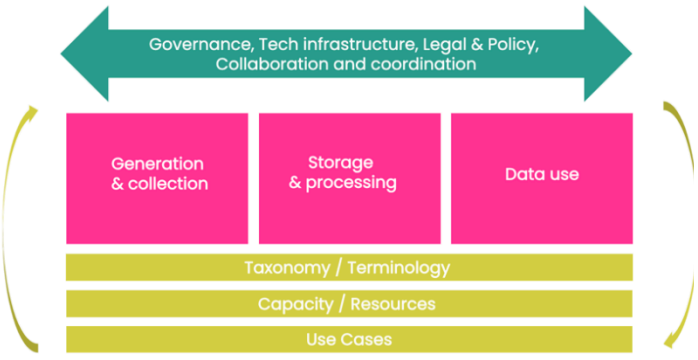
The emerging findings presented at the workshop to ground the discussion were not meant to be comprehensive, but rather speak to dynamics across the ecosystem and begin to spark conversations about **how to do things differently and intentionally together**. The findings shared emerged from conversations with multiple actors and presented challenges or gaps that can help us dig into **the way we think about and action data**.

Two asks were presented for everyone to have in mind in moving through the discussions:

- *As an ecosystem, we have been thinking about data as datasets and not about the full data ecosystem surrounding data assets. How are we thinking about data governance, data sharing, data agreements, legal frameworks or ethics around data collection for example?*
- *How do we move from seeing data as an organisational asset to acting as a holistic data ecosystem? Is there a way to frame data more as a public good and how can we account for the gaps in certain parts of the ecosystem with capacities from other parts of the ecosystem?*

Preliminary findings of the data landscape analysis

The landscape analysis[1] began to frame this across the data lifecycle, highlighting a few core building blocks of taxonomy/ terminology, capacity/ resources, and use cases as well as cross-cutting themes of governance, tech infrastructure, legal and policy, collaboration and coordination.



Demand for Unified Language: There is an urgent need for standardised data taxonomies and terminology worldwide, spanning sectors and organisations.

[1] *The data landscape analysis framework was developed and KIIs implemented with the support of Development Initiatives*

Crucial Gaps in Data: Children's experiences are not consistently or robustly captured; barriers to non-English language as well as gender data are significant.

Enhanced Triangulation: Better data triangulation is needed when datasets move between actors and attributes are changed based on purpose or role of the data holder.

Technology Strengthening: There is a clear need for more interoperable data systems, increased processing power (currently a rare commodity in the ecosystem), and robust data-sharing infrastructure.

Ecosystem Intricacies: Challenges underscore the interconnected data ecosystem. We as a Community must more proactively tackle technology's impact on trends-be it harm, good practices, investments, professional well-being, performance, or public reception of data-driven narratives. There is no room for these considerations as mere afterthoughts; they are integral to our work.

Some findings about gaps and needs were also often common to other data ecosystems. This points to the possibility to more easily learn from other fields. For example:

- Data Generation and Collection
    - Need for standard classifications
    - More data in terms of scale and frequency needed for robust statistics
- Data Storage and Processing
    - Multiple administrative systems and data formats do not speak to each other.
    - Limits in quality of data – e.g. metadata and levels of disaggregation
- Data Use
    - Lack of data inventories and data dictionaries
    - Ethical and responsible use questions around common practice of data reuse

A sample of emerging findings on gaps and needs across the data lifecycle that were more unique to the online CSEA ecosystem include:

- Data Generation and Collection
    - The ecosystem becomes global at the beginning of the data lifecycle
    - Difficult to capture online – in-person linkages in existing systems
    - Fragmentation of resourcing of data efforts globally
    - Lack of official statistics from national systems on the issue
- Data Storage and Processing
    - Lack of cohesion in classifications
    - Duplication of efforts during sharing and reuse of data
    - Needs for technological and human processing capacity
- Data Use
    - Rapid evolution of technology affects trends in manifestations of harm
    - Fragmented MEL efforts due to lack of standards
    - Feedback loops are critical not only for technical efficacy but also for human wellbeing
    - Huge legal and communications resources needed to navigate fuzzy legal environment

Towards an improved data ecosystem

The evolving landscape of online child safety advocacy demands agile and collaborative approaches. This synthesis captures key insights from discussions, highlighting the need for creative data sharing, interdisciplinary collaboration, acknowledging biases in data storytelling, adaptation to the rapid pace of technological development, and ensuring the inclusion of children's and survivors' voices.

Effective Data Harmonisation

Methodologies for data collection, storage and processing need to be **clearly demonstrated and evaluated**. If there is understanding on how the data was collected and how it is processed and stored, comparability becomes easier. **Consistency in definition of data attributes** was raised as a challenge; these tend to be defined by practitioners either processing content or supporting victims. A specific example of this is age grading of victims in child sexual abuse material - some use numerical ages and some use maturity levels. This varies in national and international law as well as in the different systems being used to collect data. The languages used in these descriptions are **not usually fit for advocacy or legislative purposes**. There is also a significant problem with continuing use of victim blaming terminology.

Way Forward: Immediate steps in data efforts should focus on how to facilitate harmonisation, how to translate technical language into advocacy for wider audiences, and how to correct victim-blaming language.

Priorities Drive Data Ecosystem Development

This field seems to be **highly responsive to emerging threats**, especially those related to new technologies like for example recently AI. **Institutional mandates tend to drive design of data collection systems** along with data definitions and classifications within organisations. And **funding priorities** tend to impact objectives or purpose driving data collection or use. This reactivity does require a refocus of internal capacity within organisations on new issues and adaptation of team expertise and design of systems required to best tackle it.

Way Forward: There is strong opportunity for direct influence at a donor and Government level via prioritisation and allocation of funds to promote increased alignment.

Fragmentation in the Data Landscape

There is a fragmented data landscape as well as **fragmented investments across geographies**. For example, Disrupting Harm provides ground-breaking data on children and survivor experiences of online CSEA across countries in the Global South while this targeted data effort is largely missing in the Global North. Evidence in **non-English languages is a huge gap**. Even if data sources exist in certain languages, standards such as hashsets or codes must also be in those languages to be able to effectively process, analyse and apply the data to certain use cases. Country specific sensitivities can also be a challenge, especially in non-English speaking contexts.

Way Forward: Organisations leading efforts can intentionally address under-represented geographies and groups within the ecosystem and expand the Community to engage and align strategically with related and intersecting fields and agendas such as cybersecurity, privacy, digitisation, etc. Efforts – internal capacity, external communications, criteria for inclusion – also can be increased in non-English languages to fill these data gaps.

Responsible Data

Clarity on data ownership and storage is essential. When using data, particularly related to child, victim, or survivor information, acknowledgement of the original collection purpose is crucial for appropriate and ethical use. In this ecosystem, data undergoes sharing, reclassification, and repurposing. For instance, industry platforms report to hotlines and attributes vary often by platform or service making the report, that data is then processed for victim identification or response, shared to law enforcement for both victim and offender identification, and may even be repurposed for advocacy or improved frontline services based on survey data. **Defining the purpose of data collection** is vital for responsible leveraging and effective communication with various stakeholders to ensure child online safety.

Way Forward: Principles around responsible data generation and use need to be developed and applied as a Community. Data collection, sharing and use need to be informed by and include as key collaborators those with lived experience of abuse.

Translation is Contextualised

There is the challenge of **bridging the gaps in language** being used by experts and practitioners as compared to daily language being used by media, parents, young people, etc. For example, *nudes* and *sexting* need more nuanced definitions. We need balance between using terms that are **strong and correct but avoid sensationalism**. We also need to make sure language speaks to the people that it concerns, emphasising **child friendly terminology** when appropriate. Some definitions ignore the **gender dimension**. For example, 'upskirting' is not only a female only problem and can be related to boys. We must reconsider reframing some terms that are not victim-centred such as 'revenge porn'. We must also avoid the use of misleading terms such as 'dark web offending'. There are no such terms for other platform / environment specific offending. An offender is an offender in any environment – whether online or offline. Terms that are used in actual practice or have direct implications for prevention policies or practices.

Way Forward: Children and lived experience experts' perspectives are crucial. They need to be involved from the beginning in setting up the agenda in what terms need to be defined and understanding some of the terms that they are using to describe their experiences online. Meaningful, ethical approaches to child participation need to be prioritised in efforts.

Implications of Technology Development

Technology infrastructure is fragmented across the ecosystem and very few actors have the level of data science expertise or processing power to most effectively handle the ever-increasing volume of some datasets available in the space – namely industry and reporting

data. There is a great need to **reduce the backlogs in processing of data** and most of these processes are managed manually. **Technology tools and solutions** can be used or developed for data processing and triaging. There are also **risks from emerging technology,** for example AI that will magnify the volume of the data and increase the level of noise and complexity when processing the materials which will add to the backlogs.

Way Forward: More granular mapping of the data ecosystem would complement efforts for a better understanding of how information flows and how tools can be shared or made more interoperable.

Incentives and Barriers to Data Sharing

Social and reputational incentives for data sharing are some of the first to be identified by actors in the ecosystem but equally, these drivers can also pose the greatest risks as key barriers to data sharing in the absence of **robust, clear legal basis, standard ethical guidance and contractual agreements.** The level of resource required for addressing legal data sharing arrangements is enormous. A fear that sharing could potentially cause harm, whether to reputation of an actor or even more significantly cause physical harm to children, victims or professionals, is such a huge barrier to data sharing that the potential benefits – and ability to measure these clearly - must be significant to overcome that in most cases. For example, data sharing can be a personal risk particularly in countries in Global South which may not have robust laws on data protection in place. Further complicating this issue, **laws for data protection and privacy** are not harmonised across countries and legal data agreements will need to be in place for sharing of data which requires legal expertise and resources.

Way Forward: Allocate resources towards generating new and sharing existing legal and regulatory frameworks or agreements to support data sharing across jurisdictions, e.g. the Lanzarote Convention. UK Online Safety Act and EU GDPR, Data Protection Impact Assessments in academia or law enforcement, Voluntary principles within the CSR space (WeProtect five country group). Also look at other areas such as terrorism, health, disability, etc. to see if there are promising practices that have worked well and can be translated to the online CSEA data ecosystem.

Further Barriers to Collaboration

Legal and procedural barriers were flagged as surprisingly significant when it comes to data collection, processing and analysis as well. **Technical and technological barriers** are common. Working with academics can bolster CSO technical capacity in data science but data protection and ethics can be challenging in several respects. For example, **sensitivity or even illegality of some data** presents huge challenges to bringing in data scientists to support analysis. And peer review and ethics processes at universities can create long timelines for research and evidence generation that impacts the timeliness of insights.

Way Forward: Striking a balance between data protection and ethics with timeliness and impact of data is crucial. Legal resources are critical in addressing these barriers. Develop standardised methods of data collection and assessment as a community.

Examples of promising models and practices

A few data initiatives exemplifying good practices and models were shared throughout the workshop for further reference.

- INHOPE Universal Classification Schema
- International Classification of Violence against Children - CVAC, UNICEF
- D.I.S.C.O.V.E.R. Project, Safernet Brasil
- Lantern, Tech Coalition
- Whole ecosystem approach, UK Home Office
- Luxemburg Terminology Guidelines Revision process

What Comes Next

Based on the emerging findings, some **foundational ideas to take efforts forward** were shared.

- **How we talk about data matters.** Data means so many things to different people so we need to be more specific when we talk about it; for example, clearly distinguishing metadata or anonymised data or aggregated data, etc.
- **All data sources in this ecosystem are needed to inform the full picture.** We can increase complementarity and learn from the strengths across data sources.
- **The data ecosystem is more than just data.** It covers data infrastructure, governance mechanisms, technical safeguards, ethical frameworks, data assets, etc.
- **There is a lack of data governance in the online CSEA ecosystem**. We need more rigour in our ecosystem in how we approach and govern data.
- **Responsible use and intentional approaches are key**. We need to be always working with the purpose of the data in mind and conscious that the building blocks will be slow but there are already some collaborations and successful practices underway.
- **There are different sensitivities for different datasets or even data attributes**. Due to the sensitivity, safeguarding issues, the critical role tech industry data plays, and the uncertain legal status of technology solutions within the sector; the data ecosystem will be impacted by legislative change more than in other sectors.
- **We need to understand better the data we collect and offer as a Community**. Share learnings on strengths and weaknesses across data sources and networks. We still struggle to deconflict some of the online CSEA data, and we are not able to answer some hard questions about the data we collect and offer.
- **Ensuring children's and survivors' voices in approaches**, including the need for more interdisciplinary coordination, and a victim-centred approach to tackle these issues.
