Standards-based approaches to help keep kids away from adult content online

Nick Doty & Aliya Bhatia
Center for Democracy & Technology
August 2025

In an effort to block kids from online content intended for adults, some have argued that age-verification or age-assurance tools offer the possibility of simple, effective guardrails. In our brief to the Supreme Court last year, CDT laid out serious concerns these tools raise regarding privacy and free expression – in addition to questions about their efficacy. But that doesn't mean technical work can't address some valid concerns about minors' access to adult content online. In particular, two categories of work related to internet standards (labeling and signaling) are worth pursuing right now, and a third may be worth evaluating in the future (anonymous attestations). We propose principles to evaluate those and any other technical mechanisms, including their efficacy, accessibility and impacts on user privacy, safety and free expression.

Submission to the IAB/W3C Workshop on Age-Based Restrictions on Content Access (agews).

I. Labeling

Parents who wish to prevent their children from accessing adult websites can already use technical means to do so, as most devices can be set to block adult websites. This functionality depends in part on sites labeling themselves as adults-only via metadata. Most pornography-focused content sites are happy to label themselves as adults-only: it's cheap and easy, and allowing minors to view their content raises legal, regulatory, ethical and commercial concerns that sites would rather avoid.

This mechanism isn't limited to a particular jurisdiction's requirements or a specific implementation. Users can't trivially work around this approach by accessing a site located in another country or using a VPN to direct internet traffic through another country.

Making these tools more robust — through well-defined standards that are widely used by websites and interpreted by web browsers and parental control tools — could make them more widely adopted and more helpful to parents. Updating past work to recognize the current practices of parental control software could also be useful. Age rating in mobile device app

stores, a form of self-labeling for age appropriateness, has also become common and includes some level of third-party review.

Many past projects have focused on this category of work, and they offer lessons for future work. In particular, simplicity of implementation is significant for broad adoption by publishers; third-party labeling is possible as an addition to self-labeling, but introduces more complexity for readers. Self-labeling currently relies on voluntary adoption and accuracy is similarly dependent on the publisher's own classification. Where third-party labeling is used, the parent or consumer has to develop a trust relationship with a labeler.

Prior work in this area includes (not an exhaustive list!):

- Platform for Internet Content Selection (PICS) (1996),
- ICRA (2000-2008) (using PICS and RDF Content Labels),
- RTA Label (2006),
- Protocol for Web Description Resources (POWDER) (2009),
- age.xml
- Machine-readable and Interoperable Age Classification Labels in Europe (MIRACLE) (2014)

II. Signaling

Just as users may request a "safe mode" when using Google search or YouTube, devices could be configured to request "safe mode" when visiting other sites on the web. Proactively alerting sites that there's a young person (or just someone avoiding not-safe-for-work content) on the other end of the connection has the advantage of working on platforms that contain content appropriate for general audiences alongside content for adults only.

Safe modes are widely implemented in site-specific ways, although mostly not triggered by standard signaling. However, Prefer: safe (RFC 8674) is one very simple option that has seen some implementation.

Declared age ranges could provide similar functionality with more nuance and detail to accommodate apps that reflect different content and behavior for multiple age ranges, rather than just generally safe or generally unrestricted. Apple has implemented a Declared Age Range API on iOS that would communicate a voluntary, declared age-range for the user, only to apps that proactively use that information. Similar mechanisms could be extended more broadly to the web with client hints, including server-side opt-in, voluntary client-side indication, and enrollment and monitoring to detect and cut off abuse of the signals.

Signaling necessarily introduces some privacy risk: clients must expose that they are looking for safe content or content appropriate to a particular age range. This is a trade-off for those users, and should be paired with rules and accountability measures for services that solicit and consume these signals. We believe many users are willing to accept this signaling tradeoff, because the data shared is voluntarily disclosed and fairly coarse.

III. Anonymous attestations

Some have proposed using cryptographically-verified attestations of age from a range of issuers, including government-issued identity documents or banking information. Verifiable credentials have the capacity for selective disclosure (revealing only the property of age, or a predicate property such as age-over-18, rather than other inherently linkable fields like a person's name or passport number) and unlinkable presentations (revealing age to multiple websites won't confirm to those websites that the same person is accessing all of those sites). Zero-knowledge proofs or large-scale batch issuance may be ways to provide verifier unlinkability. The goal and premise of these proposals is to provide anonymous proofs of age without requiring or revealing additional information about the user.

However, this technology is early in its development, and risks to privacy and free expression remain significant. Implementations continue to introduce significant threats that we expect users won't understand. In order for granular trust decisions to be made, the issuer of the attestation is often revealed to the verifier: that information may be substantially revealing on its own (revealing a user's home location, or affiliation, or what banks or services they use) and could certainly be used to fingerprint, re-identify or track users across different contexts. Phoning home and verifier-issuer collusion risks remain, especially for any sites that want to confirm that credentials haven't been revoked and any services hosting disfavored speech where a government might want to uncover the identity of those speaking or seeking information. Centralization is an architectural risk: for the sake of privacy and ease of implementation, the ecosystem could centralize control in a small number of issuers, as noted in similar architectural designs (see RFCs 9576 and 9518 for further discussion).

Many of these issues might be mitigated with enough effort, and we anticipate future work, including standards development, implementation guidance, testing and auditing, self-regulatory practices and legal and regulatory protections. It's possible that systems could be designed using blind signatures and similar techniques to make it less feasible for issuers (like government agencies) to learn from services who is accessing what content, or group signatures or intermediary signatures to prevent services from learning who issued an age attestation. But until those challenges are addressed, these proposals won't reveal only the simple information that the user is over 18.

Proposals here include:

- Google's Credential Manager API and Sparkasse partnership
- European Commission's European Age Verification Solution

IV. Principles for evaluation

In evaluating age-based restrictions to content and other age verification and assurance requirements, CDT is working with others to develop principles that can be used to evaluate

requirements and implementations. While we hope to publish more detailed documentation in the future, we sketch some initial principles here for the sake of workshop discussion.

In short, age verification and age-based restrictions should be:

- proportionate
- narrowly tailored
- user-controlled
- accurate
- accessible
- private and secure
    - unlinkable
    - data minimized
    - limited retention
    - purpose-restricted
    - securely implemented
    - not shared or distributed
- transparent
- accountable
- community-developed
- broadly deployable

Restricting unwanted access to adult content by children can be done most successfully with less-invasive mechanisms such as those highlighted above. In instances where the goal is to limit minors from accessing content that is lawful for them to access but nonetheless age-inappropriate, equipping both minors and parents with controls to shape their own information environment will be more effective, tailored and dynamic. Labeling and signaling interventions enable such user control with fewer privacy and security risks (though as noted above, risks persist).

In the limited contexts where age verification systems based on service-side data-collection may be deployed, approaches that are uniformly accurate and accessible should be prioritized. Approaches with disparate efficacy, accuracy or availability rates across demographics should not be preferred. Data collection often includes sensitive data, including biometrics, financial transactions or government-issued high assurance identity documents. As such, those approaches should be implemented in a way that minimizes the collection, retention, and sharing of data and metadata to ensure inappropriate access and abuse of sensitive data is mitigated. Examples of privacy protections include:

- Limiting the linkability between presentations of ID (meaning that the issuer of the ID should not learn where the ID is presented, and services seeking to access ID and verify age should not be able to connect users between services);
- Not revealing who the issuer of the ID is, nor collecting or storing data not directly related to the purpose of verifying age, including data related to location, birthplace, etc.;

- Only retaining and sharing the data required to verify a user is above age, including limiting the retention and sharing of birthdate data;
- Storing no data on the services and types of services where the user presented proof of age;
- Minimizing or not collecting at all any sensitive biometric data; and
- Deleting proof of age data after proving age and meeting the minimum conditions of legal requirements.

Any age verification approach should offer users clear disclosure on the method used, what data was collected and stored and with whom it was shared if at all. Furthermore, users should have the ability to know who is offering the age verification system and how to meaningfully request deletion of data and remedy inaccurate age classification when it occurs.

About CDT

The Center for Democracy & Technology (CDT) is the leading nonpartisan, nonprofit organization fighting to advance civil rights and civil liberties in the digital age. We shape technology policy, governance, and design with a focus on equity and democratic values. Established in 1994, CDT has been a trusted advocate for digital rights since the early days of the internet.

Protecting privacy and free expression while designing practical mechanisms to promote child safety online has been an area of work for CDT since its founding, and continues to be a topic of research, advocacy and design to this day. In the 1990s, CDT advocated for less-invasive user-controlled filtering mechanisms, including as one of the first civil society participants in the World Wide Web Consortium, working on the Platform for Internet Content Selection, and as a supporter of legal challenges to the Communications Decency Act in *Reno v. ACLU*. More recently, CDT has published research reports on building reporting processes for non-consensual intimate imagery and summarizing effective online child safety features and advocated in the US and Europe for laws and regulations to protect children's privacy and ability to use the internet.

Additional references

CDT Files Amicus Brief in Free Speech Coalition v. Paxton, Challenging TX Age Verification Law
https://cdt.org/insights/cdt-files-amicus-brief-in-free-speech-coalition-v-paxton-challenging-tx-age-verification-law/

Using Internet Standards to Keep Kids Away from Adult Content Online
https://cdt.org/insights/using-internet-standards-to-keep-kids-away-from-adult-content-online/
(blog post with overview of these alternatives)

The Kids are Online: Research-Driven Insights on Child Safety Policy
https://cdt.org/insights/the-kids-are-online-research-driven-insights-on-child-safety-policy/

Rapid Response: Building Victim-Centered Reporting Processes for Non-Consensual Intimate Imagery
https://cdt.org/insights/rapid-response-building-victim-centered-reporting-processes-for-non-consensual-intimate-imagery/

RFC 9576: The Privacy Pass Architecture
https://datatracker.ietf.org/doc/rfc9576/

RFC 9518: Centralization, Decentralization, and Internet Standards
https://datatracker.ietf.org/doc/rfc9518/

NIST IR 8525:
Face Analysis Technology Evaluation: Age Estimation and Verification
https://pages.nist.gov/frvt/reports/aev/fate_aev_report_IR8525.pdf