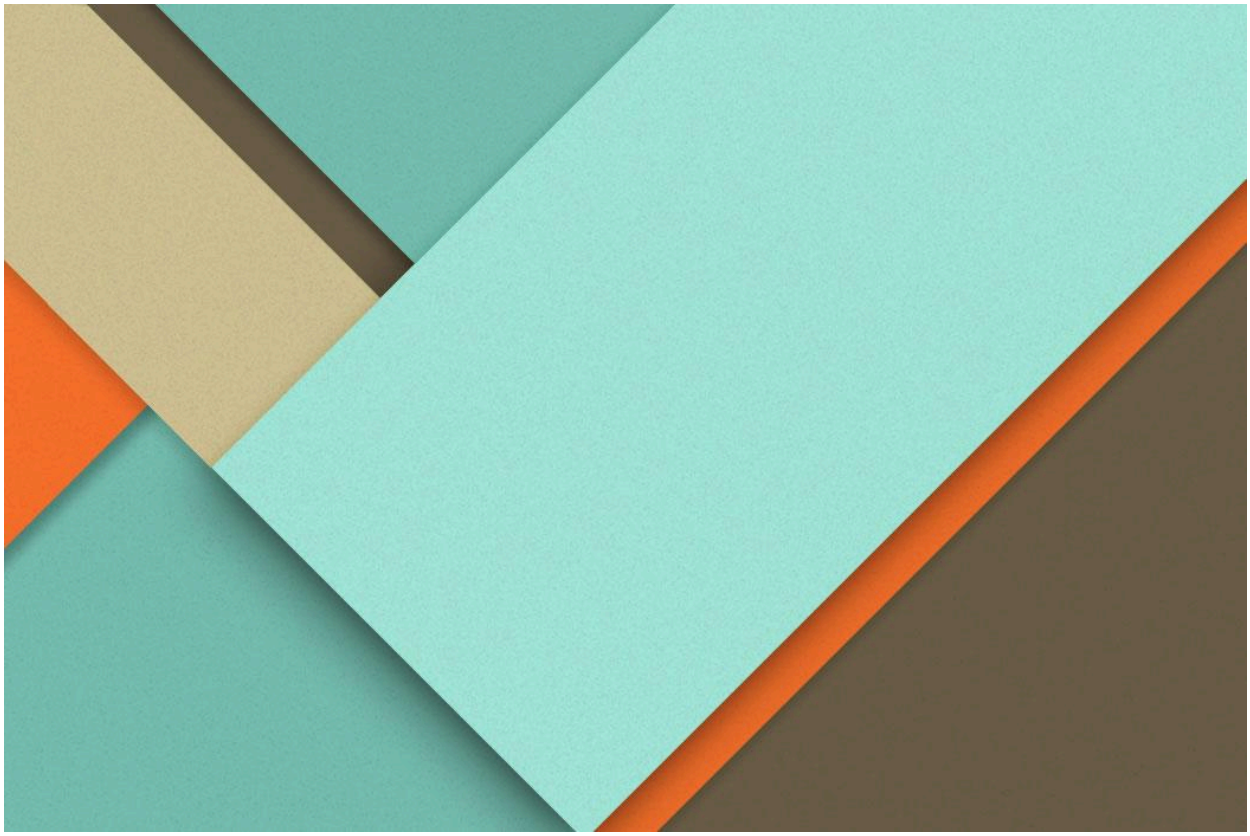

# MyWeb - Personalisation

05.07.2024


Web Civics
Australia


## Summary


The MyWeb initiative initially seeks that websites provide basic information about the suitability of their website for children and teens; as to make the web a safer place for kids.

## Overview

The Web Ratings initiative seeks to initially provide an easily deployed solution for providing content rating information about websites in a manner that then allows clients to block websites if the user is too young. This is thought to be categorized into two to three groups, distinguished by the age of a user who is not classed as an adult by law.

Whilst the method is considered simple, more complex related works could be developed overtime to progress the useful purpose of these initial steps. Historical solutions for content moderation have sought to identify specific keywords and similar, which has led to unintended consequences; such as, blocking sex and biology education content via methods that attempt to pornography. This method expects websites to provide information about their own sites and this is in-turn provided as part of the connection to that site, whereby clients can then be set-up to deny access to those sites based on conditions; such as, that the user is a child.

This method does not require identity information or payment relationships to be established with online sites considered to be 'adult sites' or similar. The method seeks to maintain privacy of users whilst addressing the real-world problem of adult content being consumed by children.

In future, the general method could be applied also to content that is stored within websites to dynamically alter what content is supplied to consumers based upon these sorts of settings. This is considered more complicated & something to be considered down the track if the initial works prove successful.

As with all things, there are still a variety of attack vectors, such as websites that claim to be suitable for children but are not. It is suggested therefore that a mechanism to report sites be defined, and that somehow reported sites then need to be processed appropriately.

There has been a long-track-record of alternative agendas being prosecuted with the 'flag' that its about protecting children, when often the consequence either does not or materially worsens circumstances for children in some way, set aside by those who did it.

There are many material risks and threats from unaccountable entities, groups and adults generally otherwise; that needs to be part of the consideration when materially seeking to protect children. Sometimes those considered most trusted in society, are the worst offenders. These are not baseless conspiracy fictions - but rather, sad matters of fact.

## Goals

I. Define ontology that can be easily deployed by websites.

II. Create a tool that helps websites implement the ontology by loading a file onto their server that can be easily identified and processed by clients.

III. Create example client-side solutions that process these files and act accordingly.

IV. Promote the opportunity for browsers, operating systems and apps to block adult content based upon settings that can be deployed and protected via a password.

V. Ensure the solution is decentralized and managed socially, not otherwise..

VI. If possible, try to ensure the model benefits the website for discovery by appropriate users.

## Specifications

The preferred option is to use RDF as the encoding method, with support for serialization in json-ld, RDFa, turtle and backwards compatibility to json. The solution has two parts.

### Server file or content notations

The server (ie: website.tld) has a file located at the root or ./well-known/file.file-extension. That provides information that is then able to be used by agents.

### Client side

The client has some code that allows it to look for and process the file, and then act in accordance with the settings defined in the client.

This can be achieved via a web-extension but is ideally integrated into browser software, once a specification has been suitably defined.

Depending on the method used; content could be parsed and then specific elements could be redacted by the client if the server is unable to redact elements of the content itself.

Examples:
A. A Website is blocked due to the browser being set to 'primary school aged child' mode.

B. Specific articles in a feed of content is blocked, whilst the rest of the content is available.

NOTE: There is a question about how 'profiles' are defined and then managed.

### Language Support

The code should support a multitude of languages, not simply American english. For this reason, it is considered better if the underlying function is provided via a code rather than words, and that the words are then able to be associated and/or defined in relation to the users language or language group.

# UseCases

## VII. Block Porn Sites

This use-case is about blocking porn-sites from being able to be used by people under the legal age which is generally 18.

## VIII. Safety for Younger Children

There are a variety of online sites, services, platforms and portals that are not suitable for children under a certain age.. Therein, the consideration about Age-Groups and related developmental levels.
Conversely also, some sites are specifically designed for children of a certain age-group, which could be made more easily discoverable.

## IX. Altering Feeds & Site Behaviour

Some social network environments are designed in such a way that means particular types of posts can be identified as having content that is sought not to be made available to children and these posts can be redacted from the content that is made available to consumers based upon configuration files being available for the client. Alternatively, platforms are also able to modify what content is made available if there is a content-negotiation process relating to the connection to the site, that is then respected by the site. This function could also be defined for AI platforms.

## X. Encouraging Support

It is believed that the vast majority of online sites that have content problematic for children, will make attempts to resolve the problem if given the opportunity to do so. These considerations also extend to both browser & operating system vendors.

IT is believed to be important to form a mutually supportive engagement to encourage this behavior by website providers. The initial expectation is that a simple file be uploaded to the root of their website that provides guidance and that this is then able to be processed and acted upon by clients via client-software, initially a web-extension but in future it is hoped to be integrated into devices.

Additional tools could be provided to assist websites, particularly self-hosted sites such as those provided via content management systems such as wordpress; which is thought able to be done easily via a plugin - that could be easily deployed by people who do not have a high-level of technical skills and/or expertise.

A side benefit of these sorts of tools may also be that specific sites that are about children or defined for children, may become more easily discoverable.

More consideration about how the ontology could be defined to support these benefits should be considered.

## XI. Further opportunities

There are several further opportunities in the general field that are thought usefully considered.

a. The ability to self-moderate content that is not wanted by users / Users being able to elect not to see content of a certain definable type.

   For example:
   - Older people not wanting to see young peoples (legal) 'sexy media'.
   - Problem Gamblers wanting to deny access to online gambling sites.

b. Promoting content that is OF Interest: for example, prioritizing posts on social networks that relates to areas of interest that consumers want more information about and/or would prefer to see prioritized.

c. Content Categorisation: the ability to define content in particular categories.

These opportunities are sought to be considered both in design, but moreover not the focus of implementation works in the first instance. Broadly otherwise, consideration that may be usefully considered in the near-term includes;

d. Whether children should have the ability to make payments, and how personal identifiers might be protected in the interests of children.

e. How notification methods might be usefully applied either directly or via api, to protect from other forms of online harms.

### f. Means to respond to bad actors

It is sadly the case that some sites may accidentally or intentionally, seek to define their sites as being suitable for children, or similar; when it is plainly not the case.

A reporting function would be useful alongside a means to consider how to address actors who are intentionally providing false information as to engender harm.

## Threat Models

There are a few threats considered already; namely,

1. A desire to maintain privacy in particular due to the implication that these works are intended to protect children; and that,

2. There is also a threat that people may misuse the method as a means to specifically target children.

3. Content negotiation & personalisation methods could be used for censorship purposes in ways that are not good for societies, human rights, democracies.

4. Alternative solutions seek to;

    a. Deploy 'digital identity' methods; which is considered a threat due to the design of these systems failing to consider human rights generally.
    b. Revoke access to encryption; which is overall more broadly, impossible and/or a very serious threat to societies, people, safety and human rights.

    Actors who are focused on these sorts of solutions generally have underlying reasons that are not declared and difficult for honorable people to discover.

5. Arbiter of truth: There is a threat of using these sorts of initiatives for censorship and manipulation. Where applied, meaningful compensation may be unavailable.

6. That various tactical engagements work to overwhelm and prevent delivery of a usefully reasonable solution; whether that be for the purpose of engendering demand for an alternative, that has unwanted consequences or otherwise.

    Better solutions that are different, are of course welcomed.

It is expected that the threat modeling will develop and this list will increase substantively.

### Nomenclature related considerations

Whilst the initiative is based upon a focus of seeking to create a simple solution to make the web a safer place for children, means to define this ecosystem solution as a tool to protect children specifically to the exclusion of others, may not be the best way of looking at it. There are various examples where the means to improve online experiences for people by ensuring they're able to have more influence over their own choices, lives, in the global agora that is the internet - is something that is also beneficial for adults, of various ages. Different threats target young adults, elderly persons; and everyone in between.

Consideration should be made about how to ensure that the approach be defined inclusively; this may in-turn provide additional benefits, in-terms of reducing the identifiability of children online in any unwanted way more broadly.

### Agent Discovery Protocol

Initially, this initiative sought to establish a broader implementation that considers various other use-cases extending well being the scope of defining a method to more specifically focus on addressing the child-safety related problems and means to address that problem.

This method has been called 'agent discovery protocol' or ADP. Whilst consistency is sought, whether and how this initiative seeks to advance these ADP works is undefined.

## Milestones

### I. Define Basic Illustration

The first step is to create a simple, initial implementation, demonstration and set of examples that can then be made comprehensible to persons & groups to advance it.

### II. Group Development

Develop a group of people thereby capable of implementing it.

### III. Nomenclature

There are various questions about how best to define the language around it, and thereby also the implementation method. Whilst initial considerations have thought about using language that focuses on age, ie: children - more sophisticated considerations may actually consider how these 'preferences' are able to be defined in a way that improves agency for persons of all ages, including those to whom guardianship relations is important - therein - children.

### IV. Development of implementation

The implementation is not considered difficult, moreover, the strategy of how it is defined is of most importance. Simple examples can be generated by an LLM.

### V. Promotion and Marketing

There both needs to be support for promoting the availability of these tools; and in-turn, the process of encouraging websites to implement it.

Workbook;

https://docs.google.com/spreadsheets/d/14wAuZ1N1nPQddYWM7uWRIoTxlLDHv407hZxObriUQnQ/edit?usp=sharing

## Background Information

There are a few initiatives being undertaken in the name of improving child-safety online that are in-turn attached to various forms of concerns.

Below is some information about those initiatives.

Age Gating https://en.wikipedia.org/wiki/Age_verification_system

ChatControl

**Related Implications**

'Digital Identity' (authentication)

mandatory age verification

## Technical Information

This is the place to put technical information.

**Implementation Considerations**

Desktop / Laptop Environments

Mobile & Tablet Devices

Modes

Simple mode & Advanced Mode should be supported. Simple mode, requires only one reference; advanced mode, requires more comprehensive customisation

Inline method

Linked Notes:
GitHub Issue (now closed)

https://github.com/schemaorg/schemaorg/issues/3560

Related Posts:
https://lists.w3.org/Archives/Public/public-humancentricai/2024Aug/0000.html

https://lists.w3.org/Archives/Public/public-humancentricai/2024Sep/0000.html
https://www.linkedin.com/feed/update/urn:li:activity:7236338418157297664/

https://groups.google.com/g/coreinternetvalues/c/aASEr-6dXl8