Privacy and the Web

To boldly go where no one has gone before...

Amelia Andersdotter & Anders Jensen-Urstad

IFIP Summer School, Karlstad, August 2016

dataskydd.net

Raison d'être

- Sweden and history...

Raison d'être

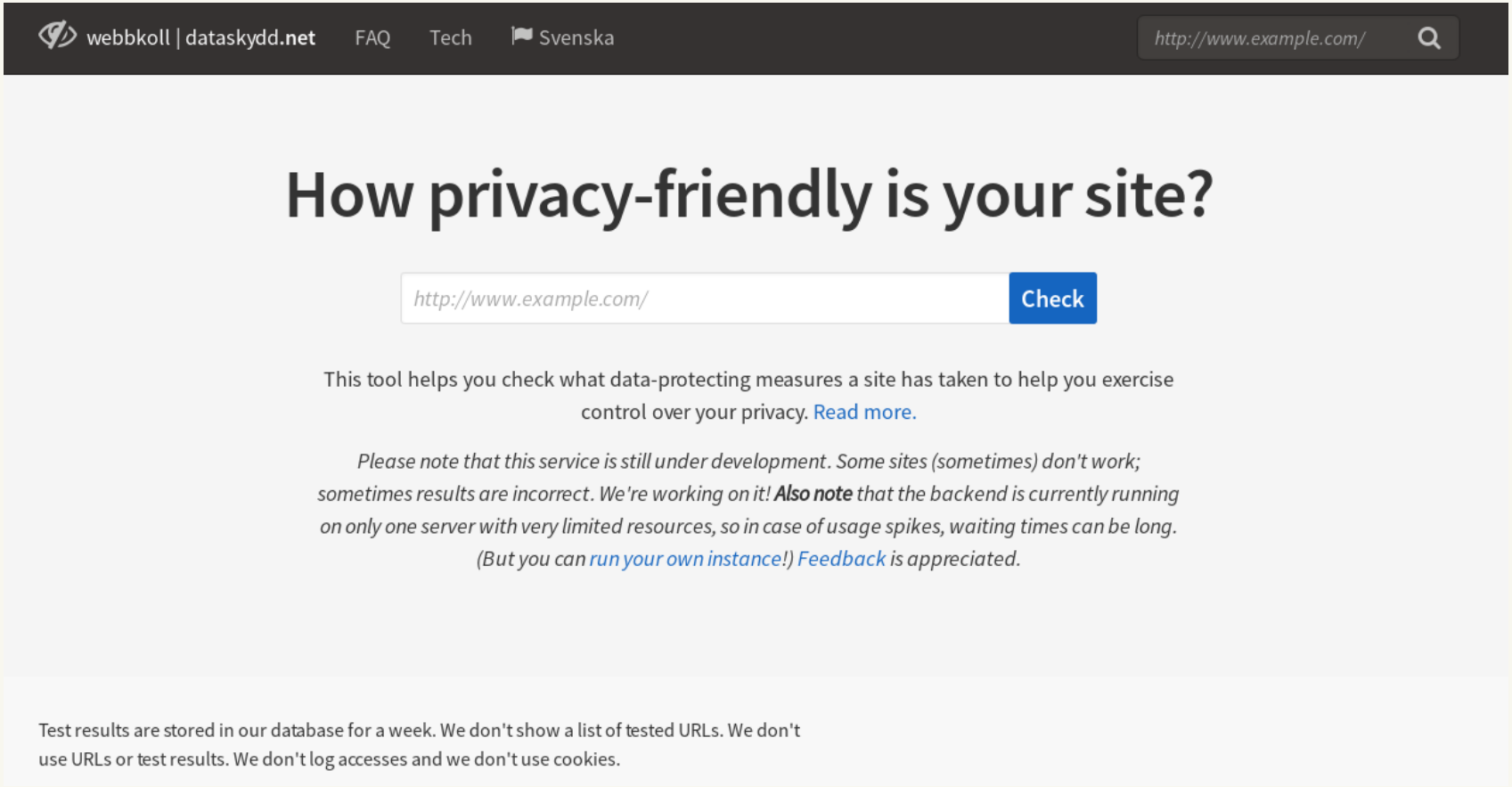
- Sweden and history...
- Data protection in 2016 is teh shock!

Raison d'être

- Sweden and history...
- Data protection in 2016 is teh shock!
- Creating https://webbkoll.dataskydd.net

A web-based tool to check websites against some privacy metrics.

- Data leakage to ISPs, schools, work, etc.
- Data leakage to adjacent websites.
- Data leakage to advertisers, CDNs, font providers, etc.



Internetfonden

IIS

Funding: Internetfonden / The Internet Foundation IIS

Results for **nwt.se**

Check again

2016-08-22 07:42:16

Input URL: http://nwt.se/

Final URL: http://nwt.se/

Insecure	Referrers leaked	44 Cookies	194 Third-party requests	50 Third-parties contacted

Insecure connection

`nwt.se` does **not** use HTTPS by default.

HTTPS encrypts nearly all information sent between a client and a web service. Properly configured, it guarantees three things:

- **Confidentiality**. The visitor's connection is encrypted, obscuring URLs, cookies, and other sensitive metadata.
- **Authenticity**. The visitor is talking to the "real" website, and not to an impersonator or through a "man-in-the-middle".
- **Integrity**. The data sent between the visitor and the website has not been tampered with or modified.

A plain HTTP connection can be easily monitored, modified, and impersonated. Every unencrypted HTTP request reveals information about a user's behavior, and the interception and tracking of unencrypted browsing has become commonplace.

The goal of the Internet community is to establish encryption as the norm,

To enable HTTPS on a website, a **certificate** for the domain needs to be installed on the web server. To get a certificate that browsers will trust, you need one issued by a trusted certificate authority (otherwise a visitor's browser will show a warning).

Let's Encrypt is a non-profit certificate authority (sponsored by Mozilla, EFF, Cisco, Facebook and others) providing free *domain-validated* (

To get a DV certificate, you only need to prove that you control the domain. To get an *Extended Validation* (EV) certificate, you must pass a more thorough identity verification process.

There is no difference in encryption between DV and EV certificates, but they are typically displayed differently in browsers. EV certificates generally result in the domain owner's name appearing in the browser URL bar that visitors see.

DV certificates are the most common. Let's Encrypt only issues DV

Referrers leaked

When you click a link, your browser will typically send the HTTP *referer* [sic] header to the webserver where the destination webpage is at. The header contains the full URL of the page you came from. This lets sites see where traffic comes from. The header is also sent when external resources (such as images, fonts, JS and CSS) are loaded.

The referrer header is privacy nightmare as it allows websites and services to track you across the web and learn about your browsing habits (and thus possibly private, sensitive information), particularly when combined with cookies.

Let's say you're logged in on Facebook. You visit a page with the URL `http://www.some-hospital.com/some-medical-condition`. On that page, you click a link to their Facebook page. Your browser then sends `Referer: http://www.some-hospital.com/some-medical-condition` to `facebook.com`, along with your Facebook cookies, allowing Facebook to associate your identity with that particular page.

The problem is made worse by the fact that many websites load resources like images and scripts from dozens of third-parties, sending referrer information to all of them, with the typical visitor having no idea that this is happening.

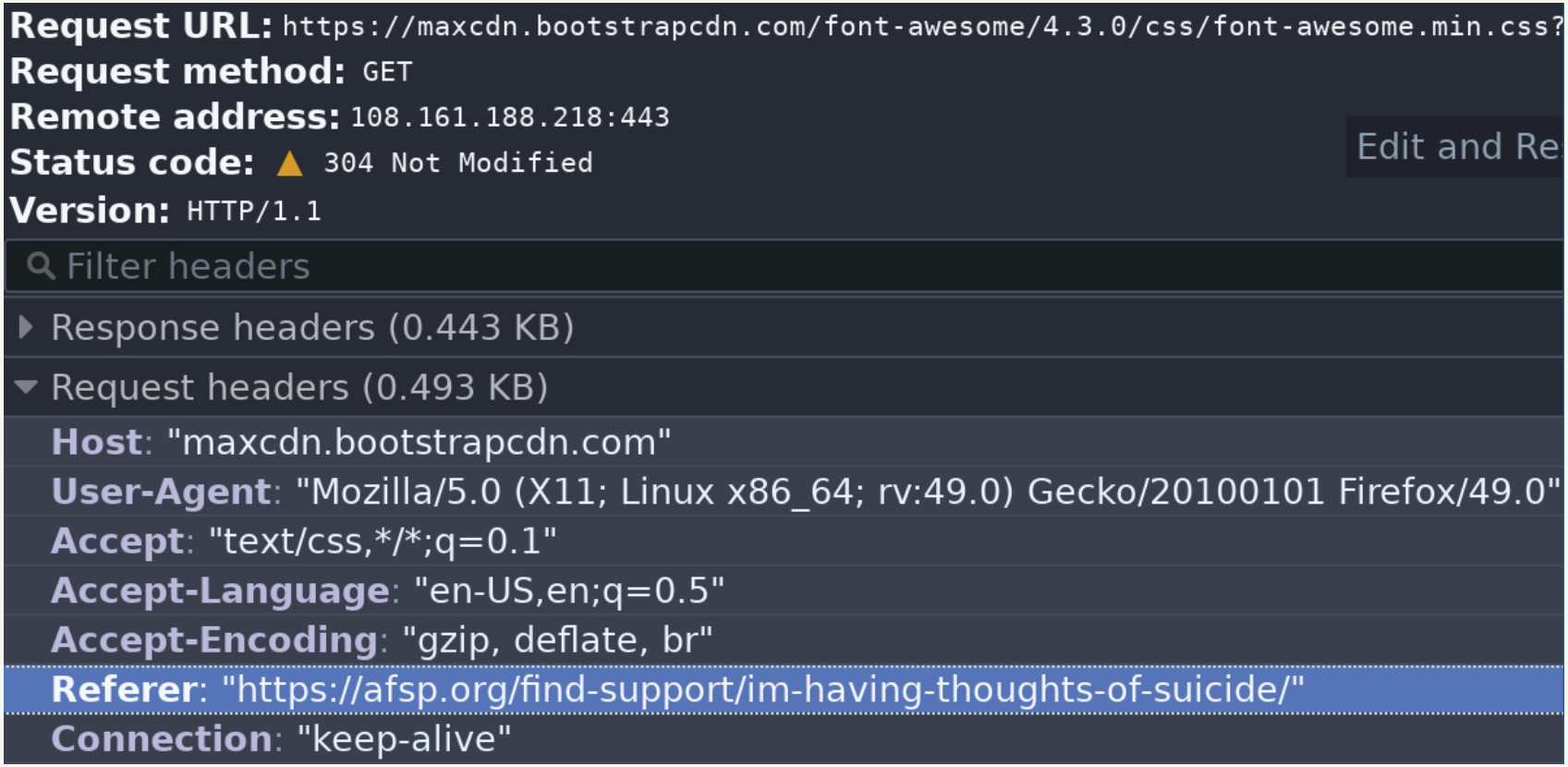
Thanks to a fairly recent development, *Referrer Policy*, it's finally possible for websites to tell browsers to not leak referrers. It lets you specify a policy that's applied to all links clicked, as well as all other requests generated by the page (images, JS, etc.).

A few different policies are offered, such as `origin` (strips everything except the origin) and `origin-when-cross-origin` (sends full URL with same-origin requests, otherwise stripped). The only one we recommend is `no-referrer`, which kills the referrer header entirely for all requests, no matter the destination.

A referrer policy can easily be set with a `<meta>` element in your HTML. Simply include this inside the `<head>` section:

```
<meta name="referrer" content="no-referrer">
```

While still a work in progress, Referrer Policy is now supported by all major browsers (except Internet Explorer, although it *is* supported by Edge, the new browser in Windows 10).


Request URL: https://maxcdn.bootstrapcdn.com/font-awesome/4.3.0/css/font-awesome.min.css?
Request method: GET
Remote address: 108.161.188.218:443
Status code: 304 Not Modified
Edit and Re
Version: HTTP/1.1
Filter headers
Response headers (0.443 KB)
Request headers (0.493 KB)
Host: "maxcdn.bootstrapcdn.com"
User-Agent: "Mozilla/5.0 (X11; Linux x86_64; rv:49.0) Gecko/20100101 Firefox/49.0"
Accept: "text/css,*/*;q=0.1"
Accept-Language: "en-US,en;q=0.5"
Accept-Encoding: "gzip, deflate, br"
Referer: "https://afsp.org/find-support/im-having-thoughts-of-suicide/"
Connection: "keep-alive"

"Note: Because the source of a link may be private information or may reveal an otherwise private information source, it is strongly recommended that the user be able to select whether or not the Referer field is sent."

— RFC 1945, HYPERTEXT TRANSFER PROTOCOL–HTTP/1.0, 10.13, MAY 1996

Referrer Policy, W3C draft

```
<meta name="referrer" content="no-referrer">
```

Just do it!

Third-party services

The site is loading libraries from one or more CDN:s.	Self-host the files.
The site is using the Disqus comment system.	Consider a self-hosted platform.
The site is using Google Analytics. While this is a powerful tool, we think you should respect your users' privacy and not tell Google about them — at least not without your users' consent.	Piwik is an excellent alternative. It's free software (PHP & MySQL) and you run it on your own server, meaning *you* are in control of the data. It offers various privacy settings and, unlike Google Analytics, it can be used without cookies. *(While analytics might be considered essential by some websites, another alternative is* don't track people just because you can. *Visitors do not, in fact, have an implicit obligation to help you optimize things.)*

First-party cookies

8 first-party cookies.

Domain	Name	Value	Expires on
.nwt.se	_ga	GA1.2.1137558315.147...	2018-08-22 07:42:02
.nwt.se	cX_P	is5qig9be1tiz3cs	2017-08-22 07:42:01
.nwt.se	_cX_segmentInfo	e4995e35ac2954a02732...	2017-08-22 07:42:01
.nwt.se	__gads	ID=f1d88caa98cf7592:...	2018-08-22 07:42:01
.nwt.se	_gat	1	2016-08-22 07:52:01
.nwt.se	cX_S	is5qigbiqdzxm122	session
nwt.se	OAS_SC1	1471851723480	session
nwt.se	adblockTracked	true	session

Third-party cookies

36 third-party cookies.

Domain	Name	Value	Expires on
.addthis.com	um	2JP.'57baaccd6d41f30...	2018-08-22 07:42:05
.addthis.com	uid	57baaccc9cd7af0e	2018-08-22 07:42:05
.addthis.com	loc	MDAwMDBFVUdCRU4yMzE1...	2018-08-22 07:42:05
.addthis.com	uvc	1%7C34	2018-08-22 07:42:05
.addthis.com	vc	2	2018-08-22 07:42:04
.addthis.com	di2	OCAWW0.6NI~1TOH~UYM	2018-08-22 07:42:04

Third-party requests

194 requests (39 secure, 155 insecure) to **50** unique hosts.

A third-party request is a request to a domain that's not `nwt.se` or one of its subdomains.

Host	Classification
acdn.adnxs.com	Advertising (AppNexus)
admin.youplay.se	
adserver.cxad.cxense.com	Advertising (cXense)
ajax.googleapis.com	Content (Google)
api.cxense.com	Advertising (cXense)
cdn-content-production.cxpublic.com	
cdn.cxense.com	Advertising (cXense)
cdn.cxpublic.com	
cdnjs.cloudflare.com	
code.jquery.com	
comcluster.cxense.com	Advertising (cXense)
content.youplay.se	
cs.go.affec.tv	
csi.gstatic.com	Content (Google)
d2s91iffsebk9p.cloudfront.net	Content (Amazon.com)
eas.mediekompaniet.com	

HTTP headers

Header	Set?
Content-Security-Policy	✖ **NO** Content Security Policy is an effective measure to protect your site from XSS attacks. By whitelisting sources of approved content, you can prevent the browser from loading malicious assets.
Public-Key-Pins	✖ **NO** HTTP Public Key Pinning protects your site from MiTM attacks using rogue X.509 certificates. By whitelisting only the identities that the browser should trust, your users are protected in the event a certificate authority is compromised.
Strict-Transport-Security	✖ **NO** HTTP Strict Transport Security is an excellent feature to support on your site and strengthens your implementation of TLS by getting the User Agent to enforce the use of HTTPS.
X-Content-Type-Options	✖ **NO** X-Content-Type-Options stops a browser from trying to MIME-sniff the content type and forces it to stick with the declared content-type. This helps to reduce the danger of drive-by downloads. The only valid value for this header is "X-Content-Type-Options: nosniff".
X-Frame-Options	✖ **NO** X-Frame-Options tells the browser whether you want to allow your site to be framed or not. By preventing a browser from framing your site you can defend against attacks like clickjacking.
X-Xss-Protection	✖ **NO** X-XSS-Protection sets the configuration for the cross-site scripting filters built into most browsers. The best configuration is "X-XSS-Protection: 1; mode=block".

```elixir
defp get_cookies(cookies, registerable_domain) do
  {first, third} =
    Enum.partition(cookies, fn(x) ->
      (x["domain"] |> String.trim(".") |> get_registerable_domain) == registerable_domain
    end)
  %{"first_party" => first, "third_party" => third}
end

defp get_cookie_count(cookies) do
  %{"first_party" => Enum.count(cookies["first_party"]),
    "third_party" => Enum.count(cookies["third_party"])}
end

defp get_meta_referrer(content) do
  content
  |> Floki.find("meta[name='referrer']")
  |> Floki.attribute("content")
  |> List.to_string
end
```

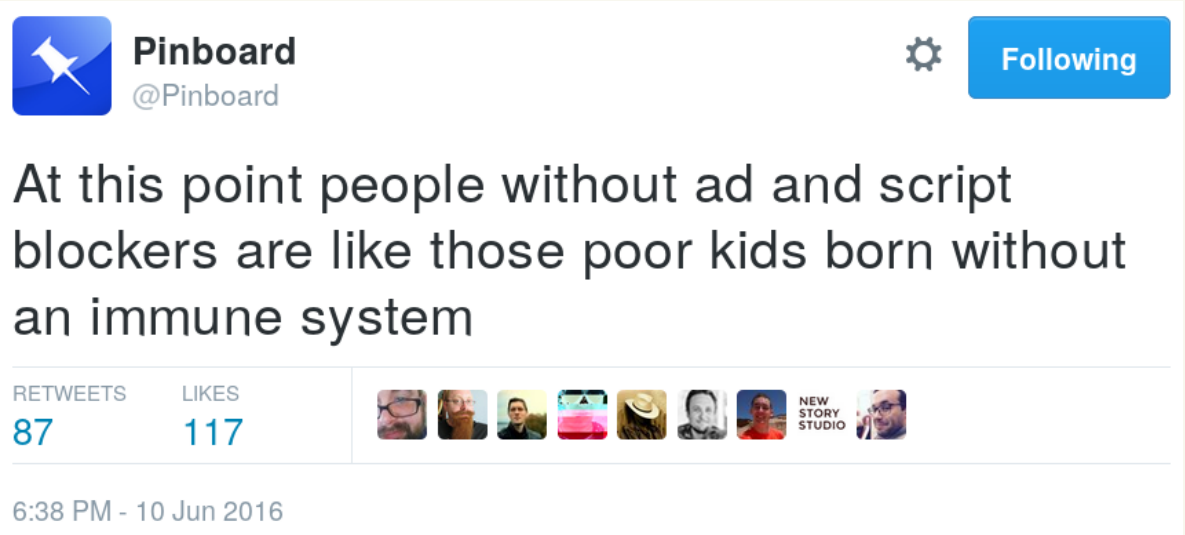
Try it: `https://webbkoll.dataskydd.net/` (English/Swedish) (*no cookies!*)

Code: `https://github.com/andersju/webbkoll` (MIT license)

Checking municipal websites

- Why do we need to use PETs when visiting public institutions?

- What are municipalities doing to decrease our need for PETs?



Scoring

A: HTTPS, no third-party requests*, no insecure requests, no referrers leaked, HSTS
B: HTTPS, no third-party requests*, no insecure requests
C: HTTPS, third-party cookies and third-party requests (but no insecure), *or* HTTP and no third-party requests
D: HTTPS, third-party cookies and third-party requests (but makes insecure requests), *or* HTTP+no third-party cookies
E: HTTP, third-party cookies, third-party requests

* *Implies no third-party cookies.*

https://dataskydd.net/kommuner/

First check on May 30th 2016:

Webbkoll: Kommunundersökning | dataskydd.net Frågor & svar Begrepp & tips Metodologi

Hur **privatlivsvänlig** är din kommun?

Vi har undersökt webbplatserna för Sveriges 290 kommuner och tagit reda på vilka dataskyddande funktioner de använder — eller *inte* använder — för att hjälpa dig utöva makt över ditt privatliv.

Webbplatserna betygsattes enligt en skala A-E. Klicka på ett kommunnamn för detaljerad information.

I korthet:
0 A
0 B
16 C
56 D
217 E
Antal med HTTPS: 14

Tips: använd Dataskydd.net:s Webbkoll för att testa din egen sajt (eller någon annans)!

Visa 25 kommuner Sök : karl

Kommun	Betyg	HTTP/HTTPS	Läcker referrers	Kakor totalt	Kakor 1:a	Kakor 3:e	Tredjeparter
Karlsborg	D	HTTP	Ja	6	6	0	2
Karlshamn	E	HTTP	Ja	7	6	1	8
Karlskoga	E	HTTP	Ja	6	5	1	10
Karlskrona	E	HTTP	Ja	9	6	3	6
Karlstad	E	HTTP	Ja	7	6	1	10
Älvkarleby	D	HTTP	Ja	4	4	0	4

Visar 1 till 6 av 6 kommuner (filtrerat från totalt 289 kommuner) Föregående 1 Nästa

https://dataskydd.net/kommuner/

Second check on August 20th 2016:

Webbkoll: Kommunundersökning | dataskydd.net Frågor & svar Begrepp & tips Metodologi

Hur **privatlivsvänlig** är din kommun?

Vi har undersökt webbplatserna för Sveriges 290 kommuner och tagit reda på vilka dataskyddande funktioner de använder — eller *inte* använder — för att hjälpa dig utöva makt över ditt privatliv.

Webbplatserna betygsattes enligt en skala A-E. Klicka på ett kommunnamn för detaljerad information.

Tips: använd Dataskydd.net:s Webbkoll för att testa din egen sajt (eller någon annans)!

I korthet:
0 A
0 B
20 C
65 D
204 E
Antal med HTTPS: 19

Visa 25 kommuner Sök : karl

Kommun	Betyg	HTTP/HTTPS	Läcker referrers	Kakor totalt	Kakor 1:a	Kakor 3:e	Tredjeparter
Karlsborg	D	HTTP	Ja	6	6	0	3
Karlshamn	E	HTTP	Ja	18	17	1	8
Karlskoga	E	HTTP	Ja	10	5	5	10
Karlskrona	E	HTTP	Ja	9	6	3	7
Karlstad	E	HTTP	Ja	11	6	5	12
Älvkarleby	D	HTTP	Ja	4	4	0	4

Visar 1 till 6 av 6 kommuner (filtrerat från totalt 289 kommuner) Föregående 1 Nästa

Resultat för **Karlstad** E

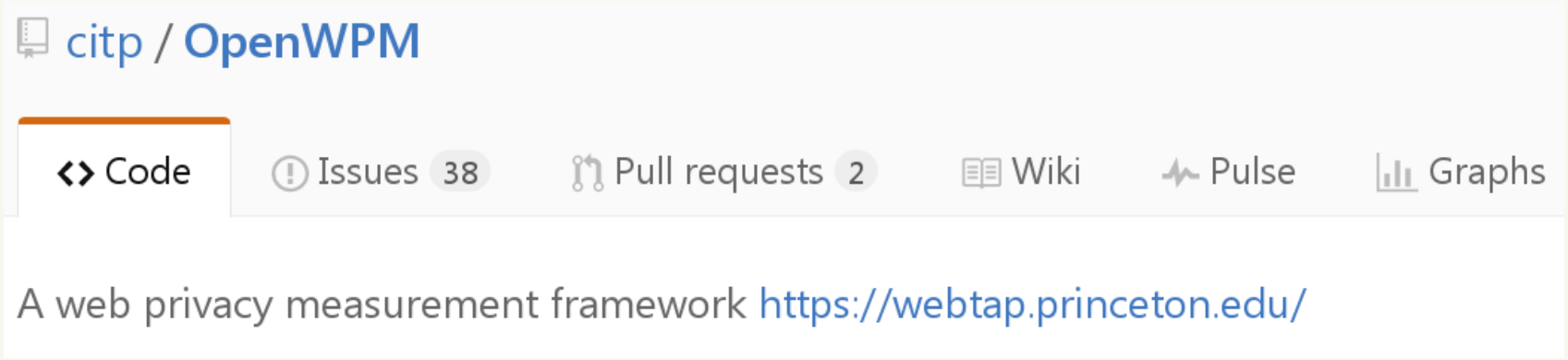
(http://www.karlstad.se/)

Osäker	Referrers läcks	11 Kakor	16 Tredjepartsförfrågningar	12 Tredjeparter kontaktade

Resultat för **Hammarö** C

(https://www.hammaro.se/)

Säker	Referrers läcks	5 Kakor	1 Tredjepartsförfrågningar	1 Tredjepart kontaktad



Englehardt, S. and Narayanan, A., Online tracking: A 1-million-site measurement and analysis. [Technical Report] Draft: May 2016.

WebTAP – Princeton Web Transparency & Accountability Project

Our code:
`https://github.com/andersju/municipality-privacy`

Results: `https://dataskydd.net/kommuner` (*in Swedish*)

Going forward...

Visualizations of website privacy analysis results?

Similar mappings of public sector in other EU countries?

Join us & implement data protection!

Thank you!

Amelia Andersdotter
@teirdes
amelia@andersdotter.cc
ameliaandersdotter.eu

Anders Jensen-Urstad
@ndrsju
anders@unix.se
anders.unix.se

dataskydd.net