

# Big Data & Tracking

# Gliederung

- I. Big Data
- II. Sammeln von Daten
  - I. Pro/Kontra
  - II. Rechtliche Lage
- III. Webtracking
  - I. Logdateien
  - II. Pixel/Beacons
  - III. Tags/URL-Tracking
  - IV. Cookies
  - V. Fingerprinting
  - VI. Federated Learning of Cohorts
- IV. Schutz
- V. Diskussion

# Big Data

Datenmengen, die

- zu groß
- zu komplex
- zu schnelllebig
- zu schwach strukturiert

sind, um sie mit herkömmlichen Methoden zu verarbeiten.

# Big Data als Buzzwort

- zunehmende Überwachung von Menschen
- zunehmende Intransparenz der Datenspeicherung
- Verletzung von Persönlichkeitsrechten
- Verarbeitung von großen Datenmengen zur Personalisierung für Wettbewerbsvorteil

# Datensammeln – Pro und Kontra

- Entscheidungen anhand von Nutzerverhalten treffen
- Telemetrie, um Softwarefehler zu beheben
- ermöglicht kostenlose Produkte
- verbessert Funktionalität
- Prognosen (Marktforschung)
- Prävention (Kriminalistik, Terrorismusbekämpfung)
- Personalisierung von Werbung, insbesondere politische Werbung
- große Menge an sensiblen und personbezogenen Daten, die einem Profil eindeutig zugeordnet werden
  - kann allgemeinem Persönlichkeitsrecht verletzen
  - schadet Vertraulichkeit von Kommunikation
- Missbrauch, Verkauf, Veröffentlichung
- Bildung/Förderung von sozialen Blasen

# Rechtliche Lage

- muss DSGVO und BDSG unterliegen
- Daten müssen anonymisiert oder pseudonymisiert werden
  - anonymisiert: keine personenbezogenen oder personenbeziehbaren Daten
  - pseudonymisiert: ursprünglich persönliche Daten werden nur einem Pseudonym zugeordnet
- protokolliertes, einsehbares und verbindlich widerrufbares Opt-In
- keine festen Löschfristen

# Webtracking

Bewegungen von Benutzern im Internet allgemein oder innerhalb einer Webseite werden erfasst und ausgewertet.

# Loganalyse

```
1 0.0.0.0 - - [17/Jun/2021:00:34:40 +0000] "GET /_profiler/phpinfo HTTP/1.1" 404 493 "-" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.15; rv:77.0) Gecko/20100101 Firefox/77.0"
2 0.0.0.0 - - [17/Jun/2021:00:34:40 +0000] "GET / HTTP/1.1" 302 473 "-" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.15; rv:77.0) Gecko/20100101 Firefox/77.0"
3 0.0.0.0 - - [17/Jun/2021:00:34:42 +0000] "GET / HTTP/1.1" 301 5278 "-" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.15; rv:77.0) Gecko/20100101 Firefox/77.0"
4 0.0.0.0 - - [17/Jun/2021:00:34:43 +0000] "GET /wiki/Main_Page HTTP/1.1" 200 20753 "-" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.15; rv:77.0) Gecko/20100101 Firefox/77.0"
5 0.0.0.0 - - [17/Jun/2021:00:37:46 +0000] "GET /wiki/Main_Page HTTP/1.1" 200 20737 "-" "Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:89.0) Gecko/20100101 Firefox/89.0"
```

IP, Datum, Pfad, Ergebnis, User Agent String



# Pixel/Beacons

- sehr kleine Bilder, die beim Laden Informationen über z.B. Aufrufszahlen liefern
- gängige Methode für Lesebestätigungen bei E-Mails

# Tags/URL-Tracking

[https://www.amazon.com/dp/exampleProduct/ref=sxin\\_0\\_pb?\\_mk\\_de\\_DE=ÅMÅŽÕÑ&keywords=tea&pd\\_rd\\_i=exampleProduct&pd\\_rd\\_r=8d39e4cd-1e4f-43db-b6e7-72e969a84aa5&pd\\_rd\\_w=1pcKM&pd\\_rd\\_wg=hYrNI&pf\\_rd\\_p=50bbfd25-5ef7-41a2-68d6-74d854b30e30&pf\\_rd\\_r=0GMWD0YYKA7XFGX55ADP&qid=1517757263&rnid=2914120011](https://www.amazon.com/dp/exampleProduct/ref=sxin_0_pb?_mk_de_DE=ÅMÅŽÕÑ&keywords=tea&pd_rd_i=exampleProduct&pd_rd_r=8d39e4cd-1e4f-43db-b6e7-72e969a84aa5&pd_rd_w=1pcKM&pd_rd_wg=hYrNI&pf_rd_p=50bbfd25-5ef7-41a2-68d6-74d854b30e30&pf_rd_r=0GMWD0YYKA7XFGX55ADP&qid=1517757263&rnid=2914120011)

<https://www.amazon.com/dp/exampleProduct>

# Cookies (LernSax)

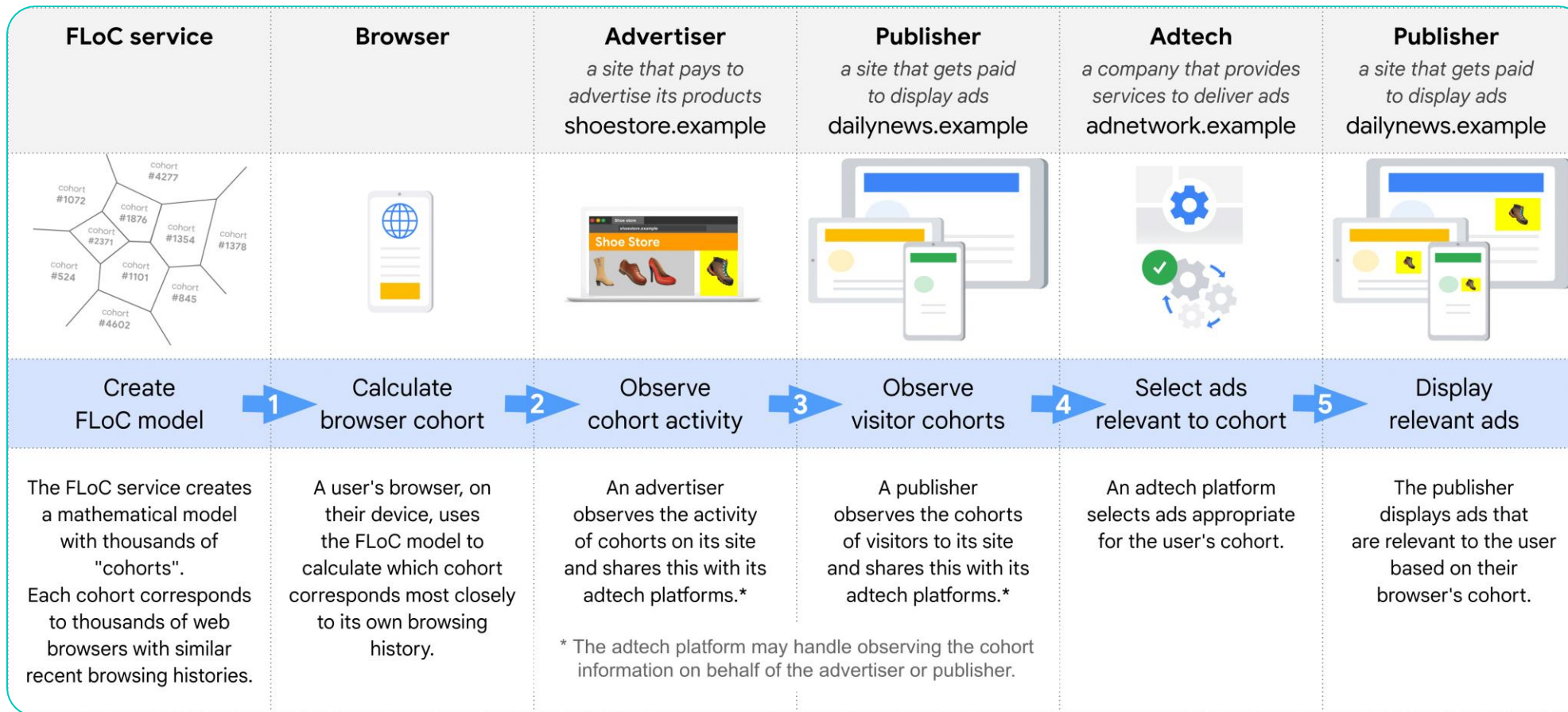
Einträge durchsuchen

Name	Wert	Domain	Path	Läuft ab / Höchstalter	Größe	HttpOnly	Secure	SameSite	Zuletzt zugegriffen
wwsc	404033447175653608 <b>1</b>	www.lernsax.de	/	Sitzungsende <b>2</b>	22	true	true	None	Wed, 16 Jun 2021 17:50:44 GMT
wwspc	1	www.lernsax.de	/	Sitzungsende	6	true	true	None <b>3</b>	Wed, 16 Jun 2021 17:50:44 GMT

1. Sitzungs-ID
2. Ablaufdatum (bis zum Ende der Sitzung)
3. Zugriffsbeschränkung auf eigene Seite (deaktiviert)

# Fingerprinting

- User Agent String: Mozilla/5.0 (X11; Linux x86\_64; rv:90.0) Gecko/20100101 Firefox/90.0
- Zeitzone: -120 Minuten; Europe/Berlin
- Bildschirm: 1920x1080, 24bit Farbtiefe
- Schriftarten
- Canvas Hash: 113e60305ab0f12d65ac9ac94b732cca
- WebGL Hash: 8b4cbedc1ffa6757996415257cfbe083
- ...



# Federated Learning of Cohorts (FLoC)

Lösungsvorschlag von Google,  
<https://web.dev/floc/>

# Schutz

- alternative Suchmaschinen (z.B. DuckDuckGo)
- privatsphäreorientierte Browser (z.B. Firefox, Brave)
- Proxies/VPNs
- Browsererweiterungen
  - Blockieren von Werbung und Trackern
  - Reinigen von URLs
  - Generalisieren/Wechseln von übermittelten Angaben

Big Data eröffnet Möglichkeiten des informationellen Machtmissbrauchs durch Manipulation, Diskriminierung und informationelle ökonomische Ausbeutung – verbunden mit der Verletzung der Grundrechte der Menschen.

Thilo Weichert, Datenschutzbeauftragter von Schleswig-Holstein, 2013

If you're not paying  
for the product, you  
are the product.

häufiges Sprichwort zum Thema (ohne Verfasser)



# Diskussion

- Würdet ihr für soziale Medien Abonnements abschließen, um mehr Privatsphäre zu erhalten?
- Schützt ihr euch aktiv gegen Tracking?
- Was haltet ihr von FLoC?
- Wie bewertet ihr die bisherige Gesetzeslage bzw. die bisher getroffenen Maßnahmen?

# Quellen

- <https://coveryourtracks.eff.org/>
- [https://de.wikipedia.org/wiki/Big\\_Data](https://de.wikipedia.org/wiki/Big_Data)
- <https://www.datenschutz.org/webtracking/>
- <https://developer.mozilla.org/en-US/docs/Web/HTTP/Cookies>
- [https://de.wikipedia.org/wiki/Federated\\_Learning\\_of\\_Cohorts](https://de.wikipedia.org/wiki/Federated_Learning_of_Cohorts)
- <https://www.eff.org/deeplinks/2021/03/googles-floc-terrible-idea>
- [https://de.wikipedia.org/wiki/Web\\_Analytics](https://de.wikipedia.org/wiki/Web_Analytics)
- <https://web.dev/floc/>