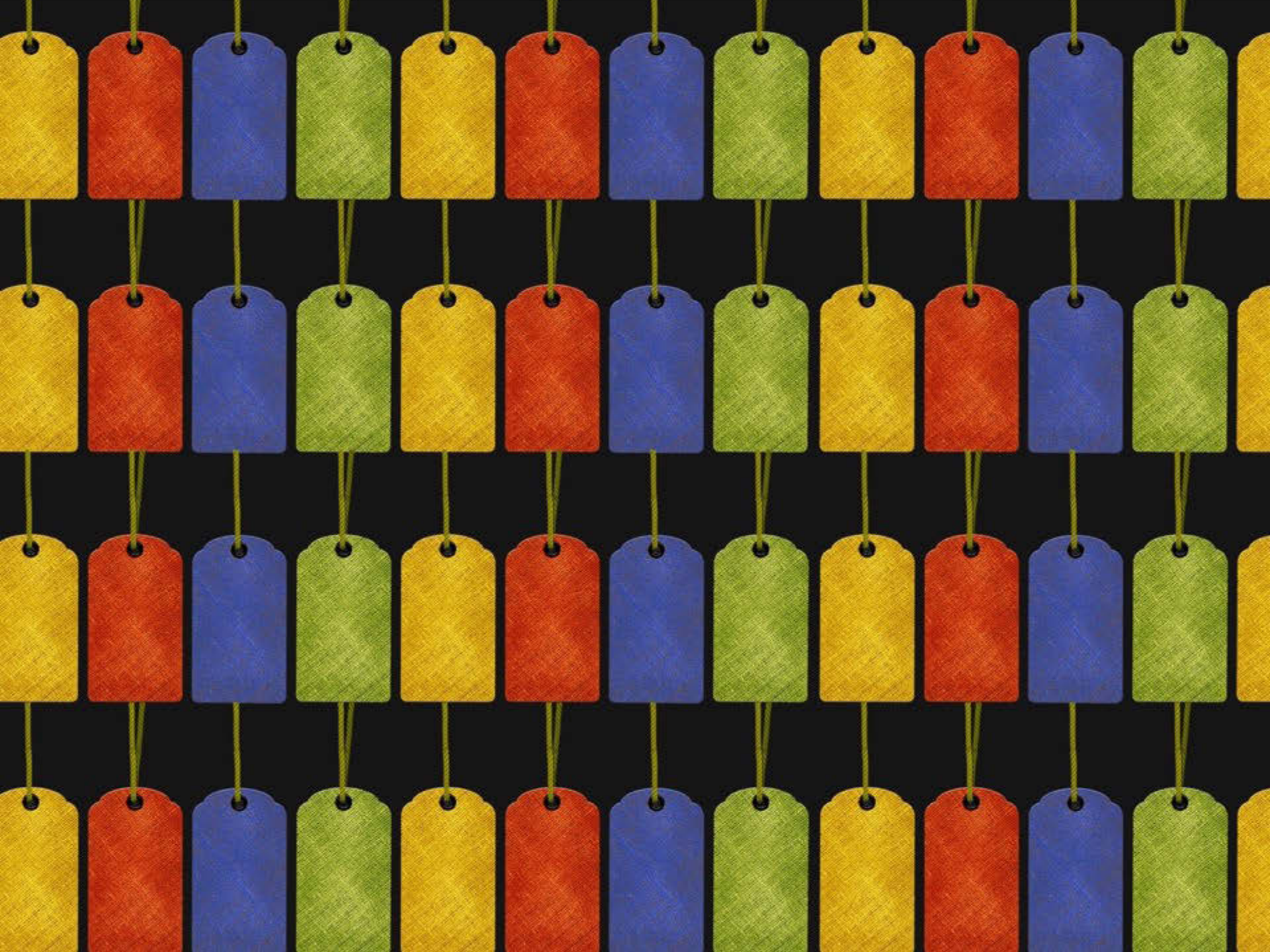
Using Apache Spark to create the Internet's first audience data firewall

Empowering digital enterprises to regain control of online data

mezzobit

Joseph Galarneau
CEO and cofounder
jg@mezzobit.com
@mezzobit



Control = $$$

Over third-party access to audiences, data, ads, and technology

$10+ billion

of estimated publisher ad revenue lost to data leakage, low viewability, and ad blocking

Source: "Selling Off Privacy at Auction" (The Internet Society, 2014)

Data control spans all verticals

Media

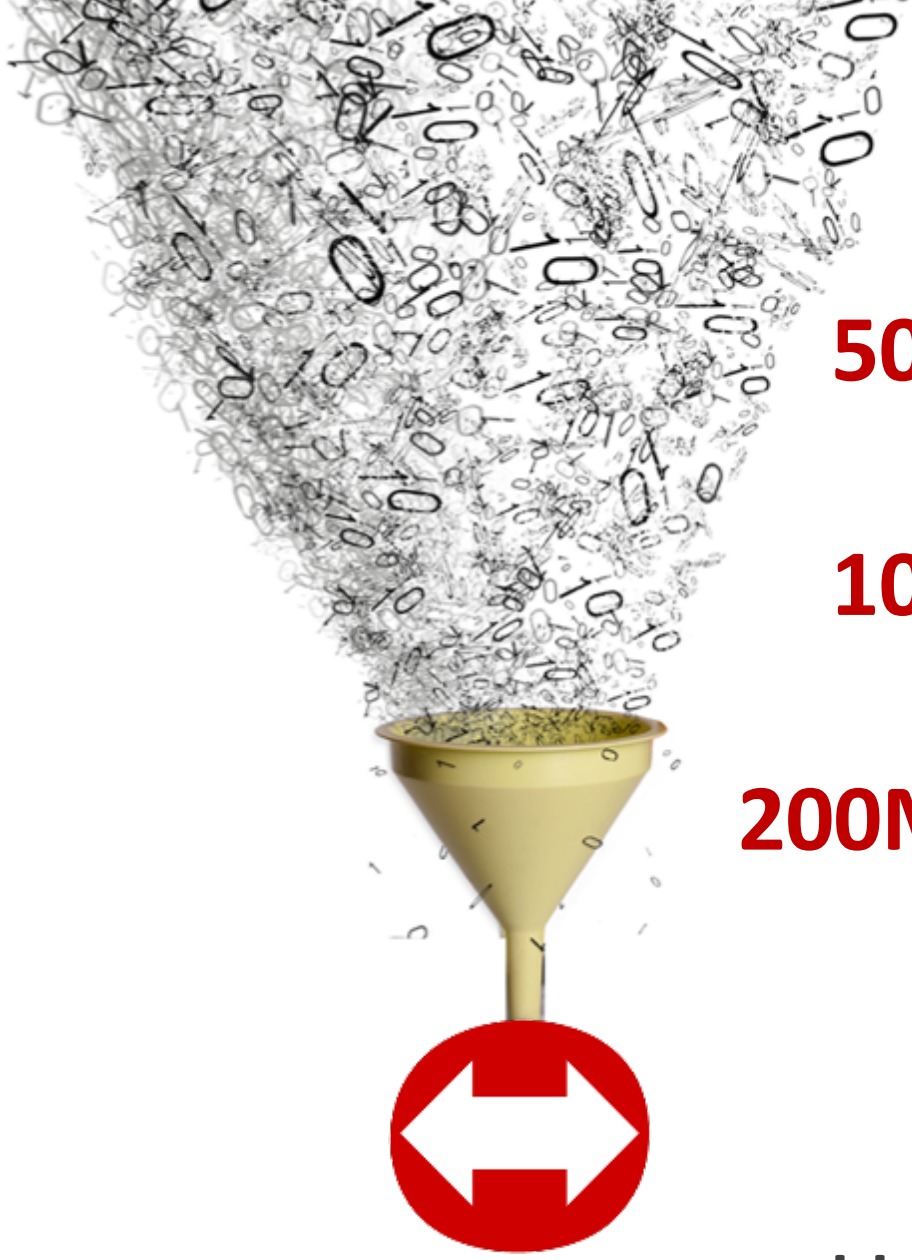
Brands

Education

E-commerce

Financial services

Healthcare



1 website

times

50 tags per pageview

times

100 elements per tag

times

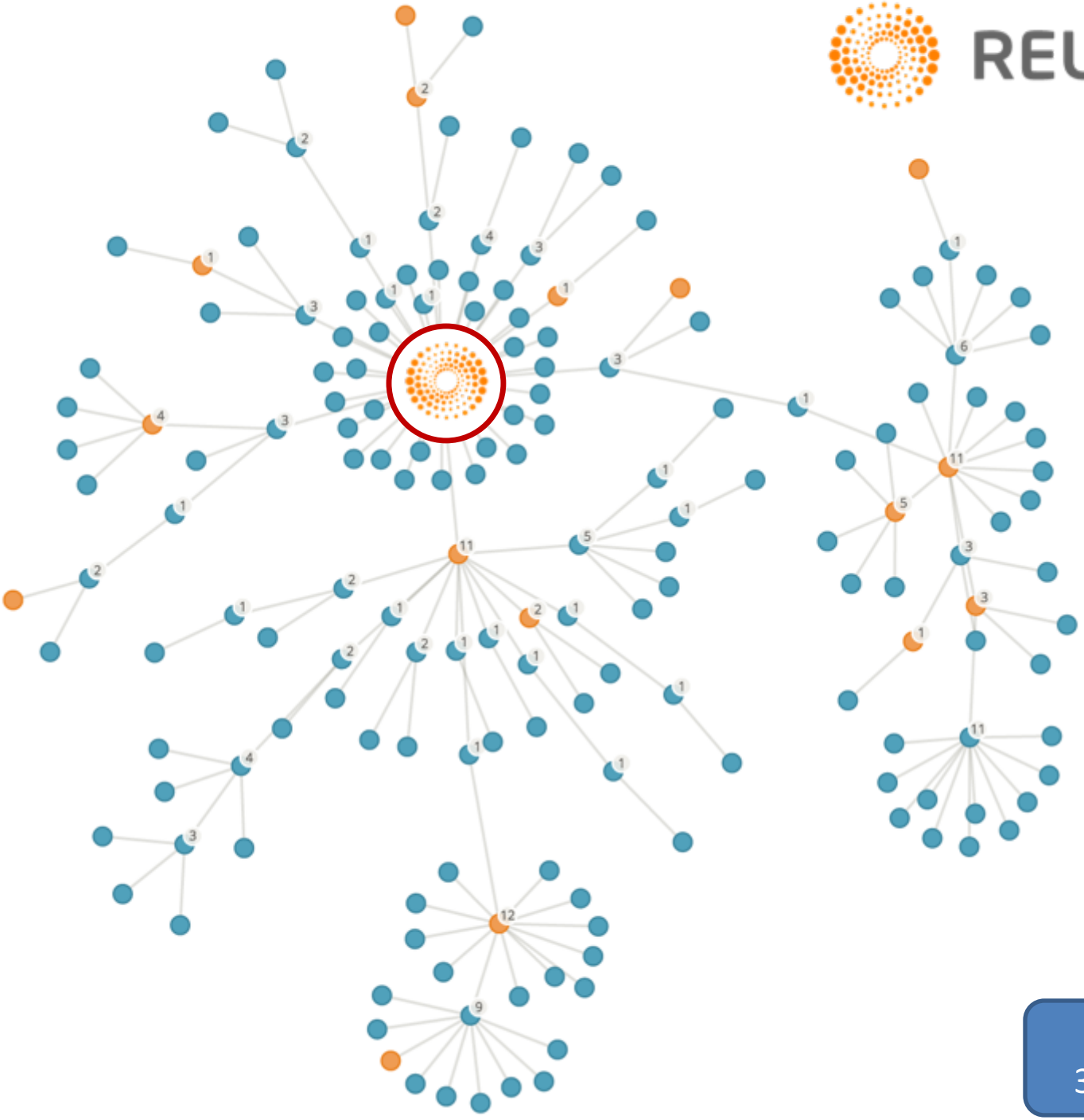
200M monthly pageviews

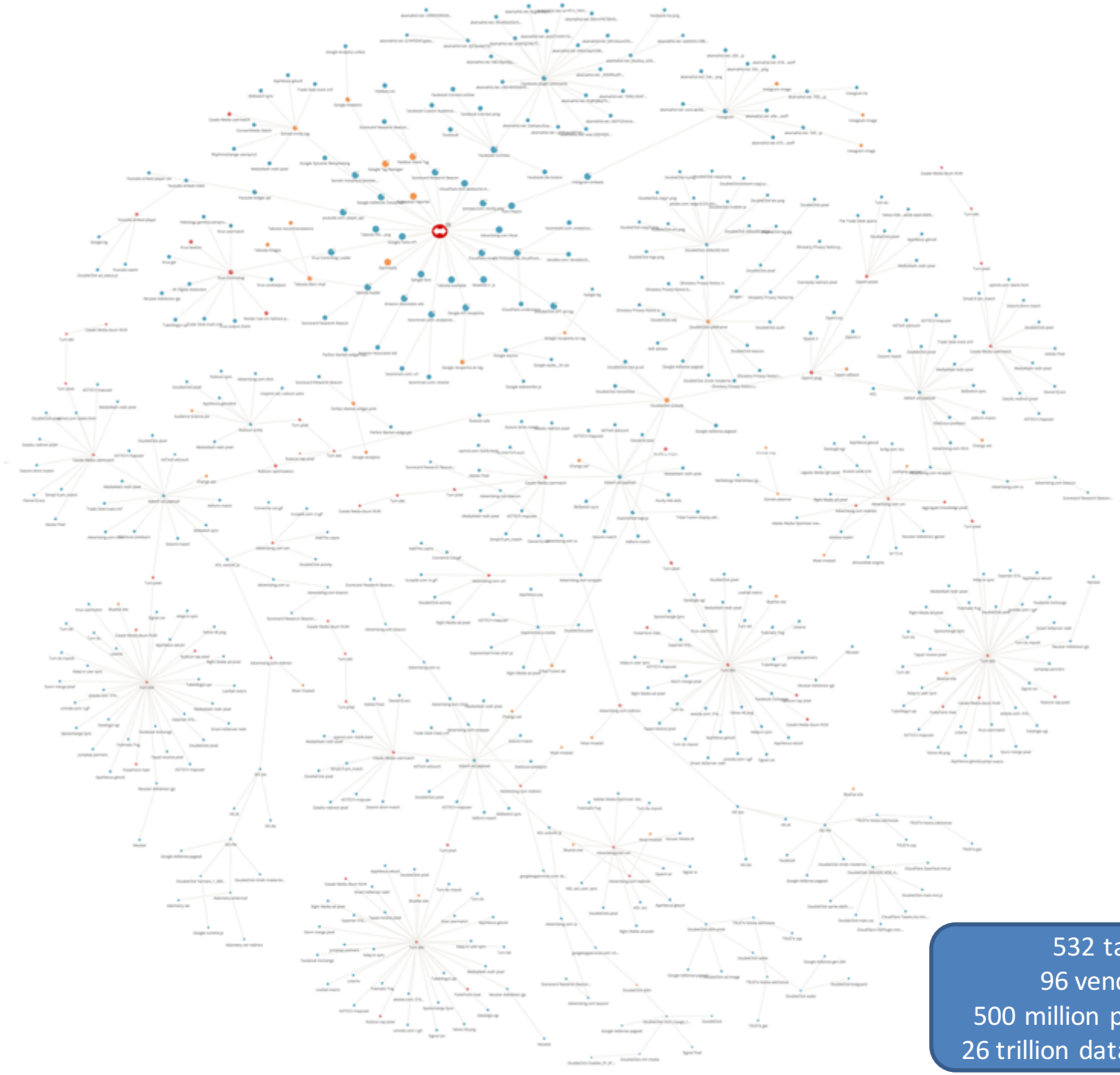
equals

1 trillion

monthly data points for **one** customer

(32 tps, when aggregated)


REUTERS
196 tags
32 vendors



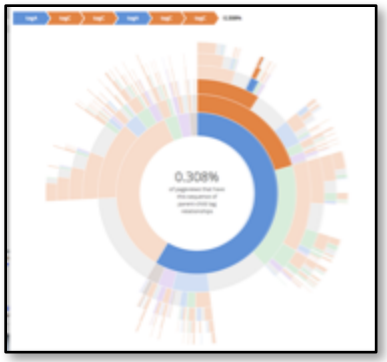
532 tags
96 vendors
500 million pageviews
26 trillion data elements

Mezzobit data needs

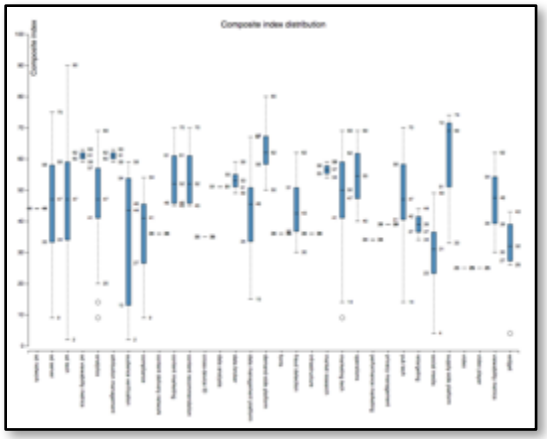
Near real-time dashboard

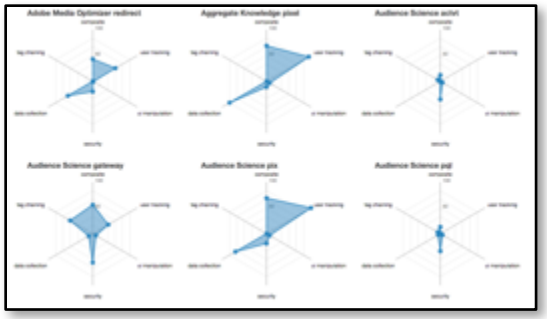






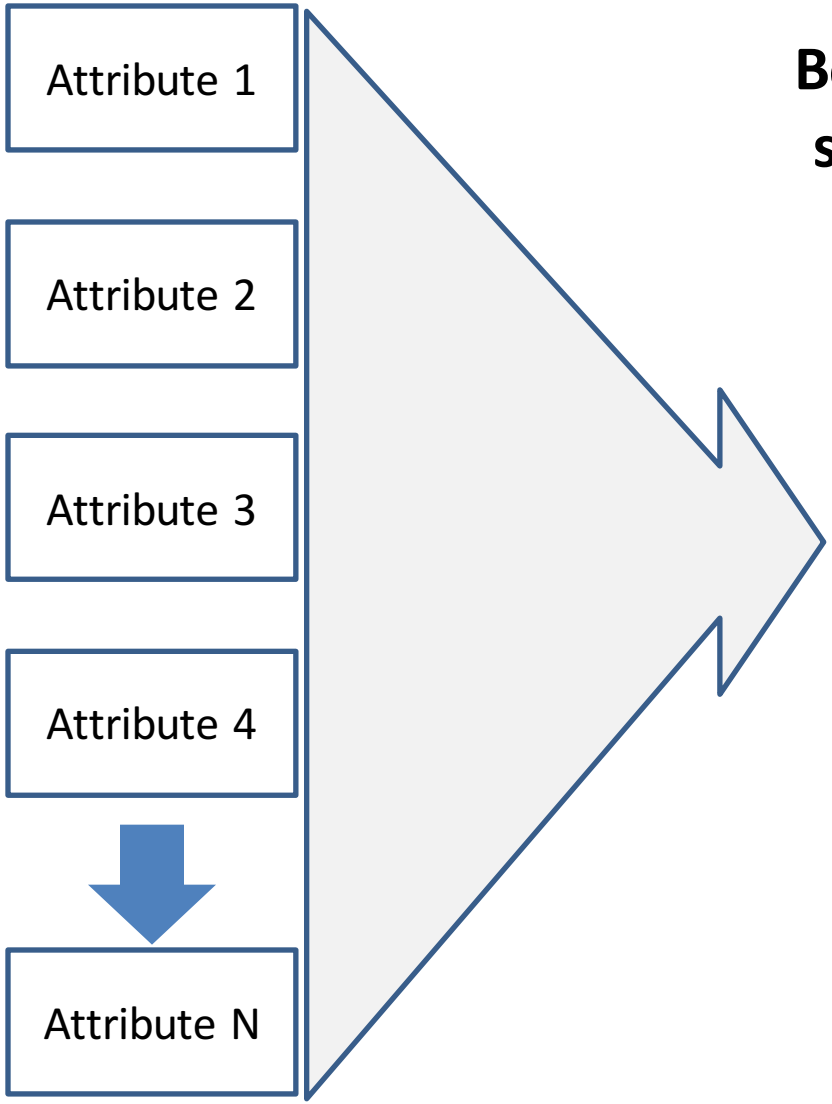
Data science



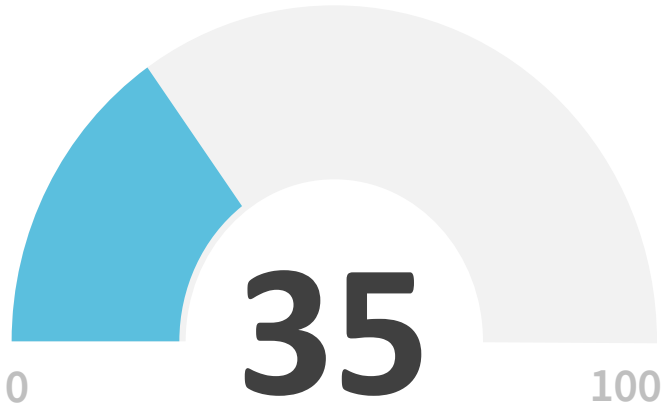


Mezzobit stack

Data science: Tag and vendor indices



Boiling complex tag behavior into a single number that can be used in dashboards, reports and rules

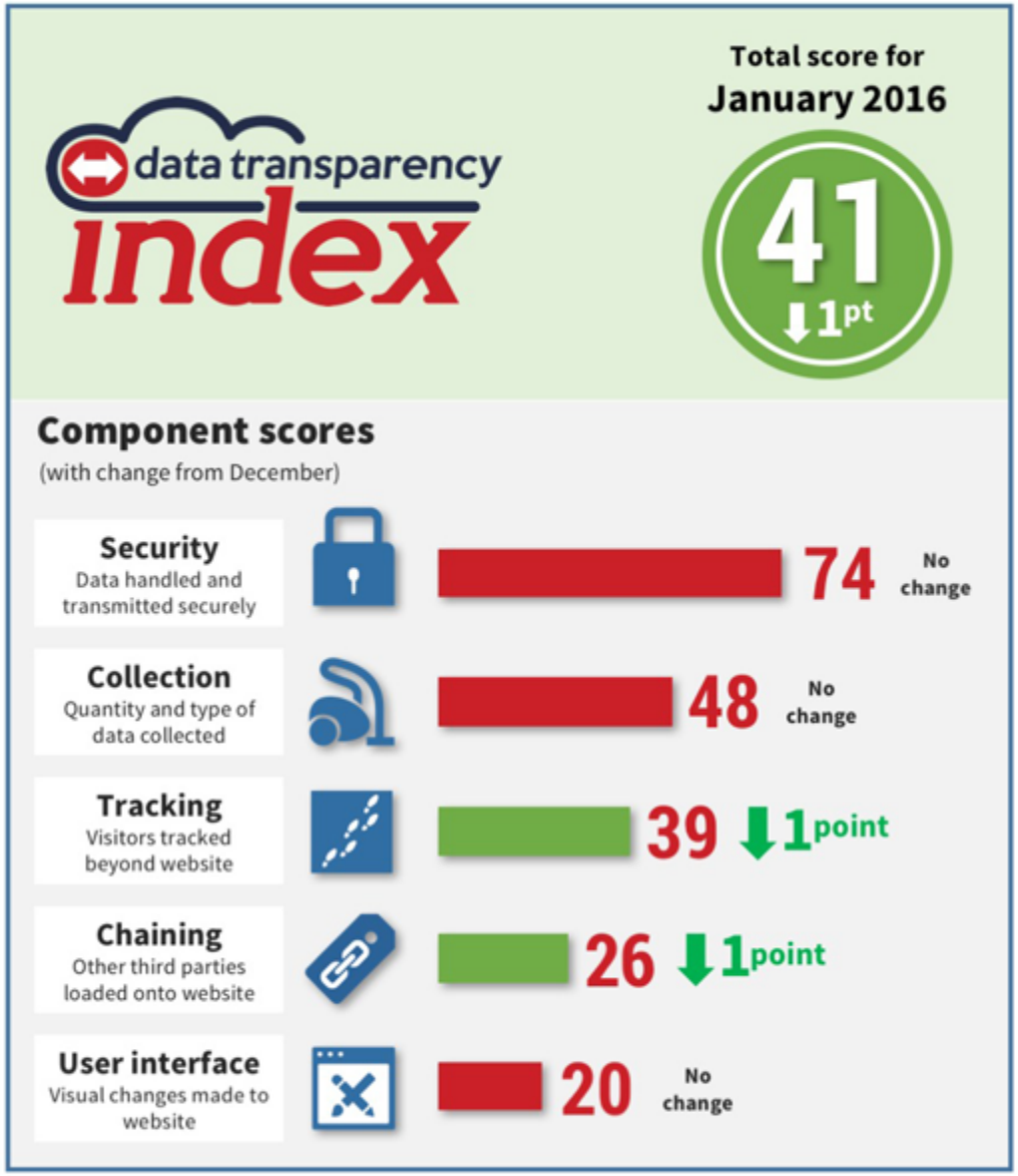


Data collection

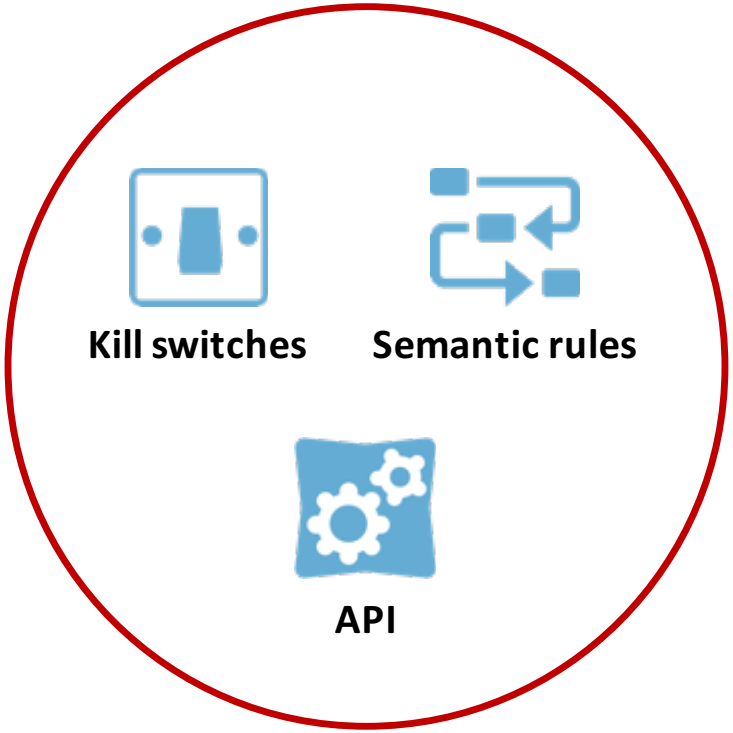
Interpretation

This tag collects and shares data more than 35% of the thousands of other tags we've seen

Data science: Tag and vendor indices

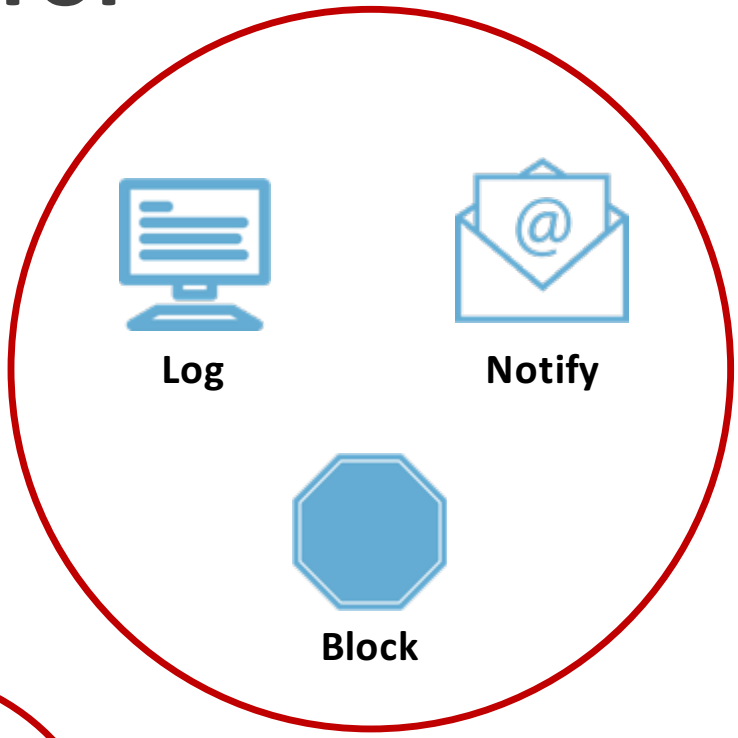


Data control

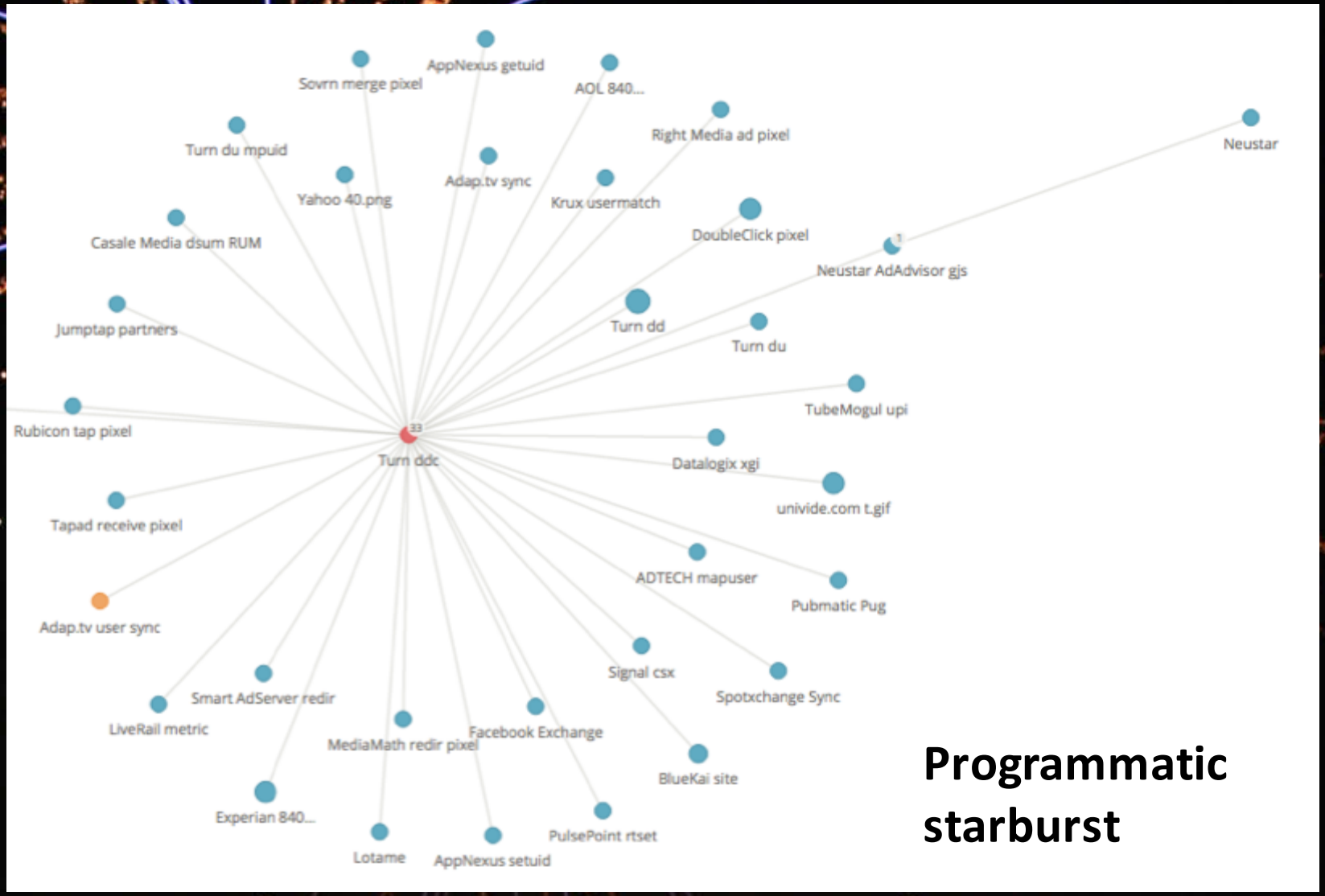


Available control tools

Scope

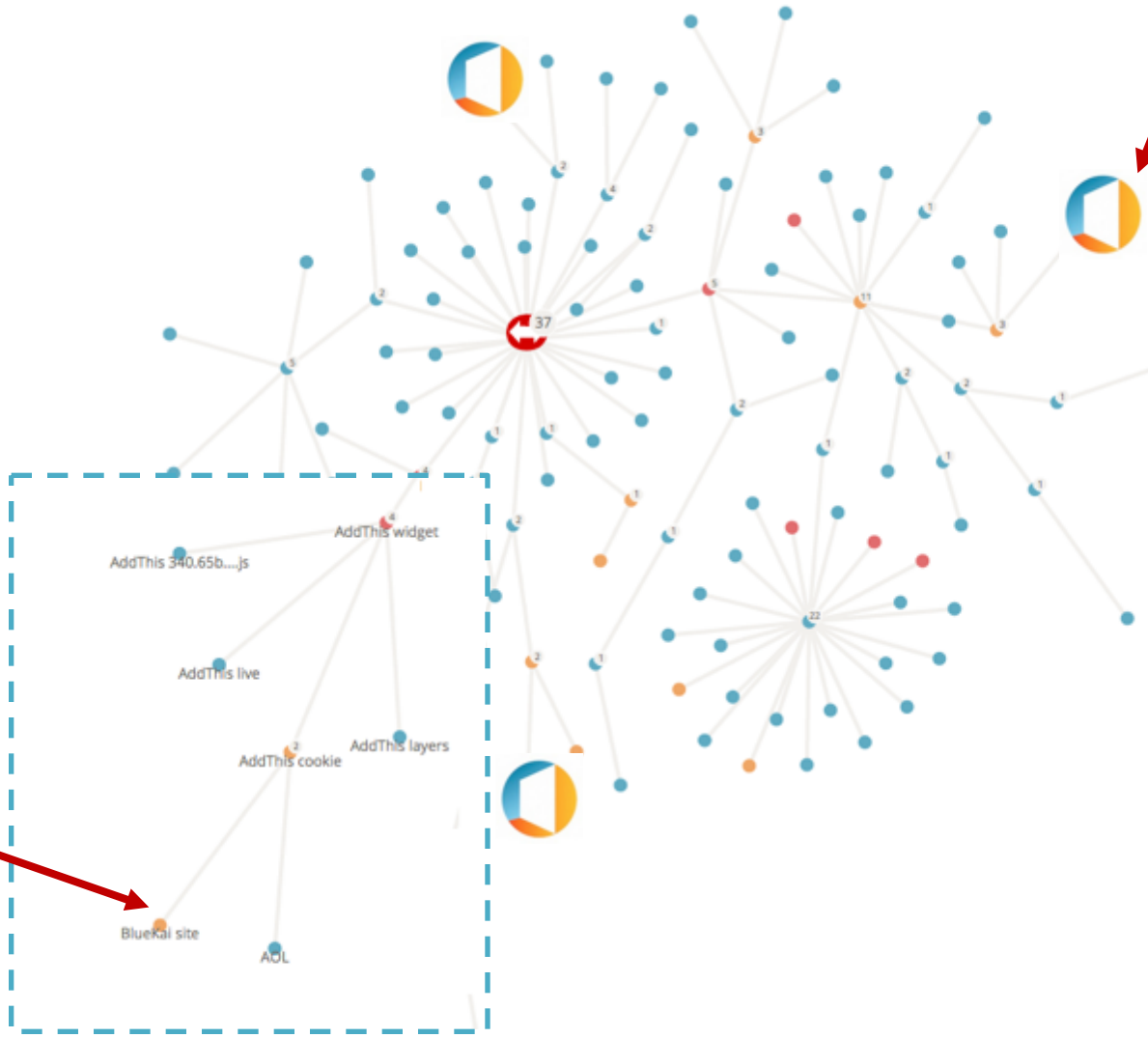


Types of control



Programmatic starburst

Tag storms: Tags can sneak in under the radar



Extra third-party analytics tags coming in through three other vendors.

Two retargeting tags riding with social sharing widget

Publishing example