

The Tracker Guide to the Cloud

Research

Diego Dacal, Kalina Dancheva, Carolin Gerlitz, Anne Helmond, Sara Minucci, Sabine Niederer, Lonneke van der Velden, and Esther Weltevrede.

Design

Alessandro Brunetti
Gabriele Colombo

Second edition

September 2015

First edition

July 2012



D E N -
S I T Y
G N +



CITIZEN
DATA LAB

Table of Contents

Introduction

How To - Find Traces of Trackers

How To - Detect Content Delivery Networks

How to - Deal with Dynamic Content

Equipment - Tools, Plugins and Research Protocols

Glossary - Overview of Trackers and CDNs

Maps - Tracking Ecologies for Sets of URLs

Maps - What Happens When You Visit a Webpage?

Findings - Distinguishing Between CDNs and Trackers



INTRODUCTION

~ Welcome to
Trackers and
Content Delivery
Networks

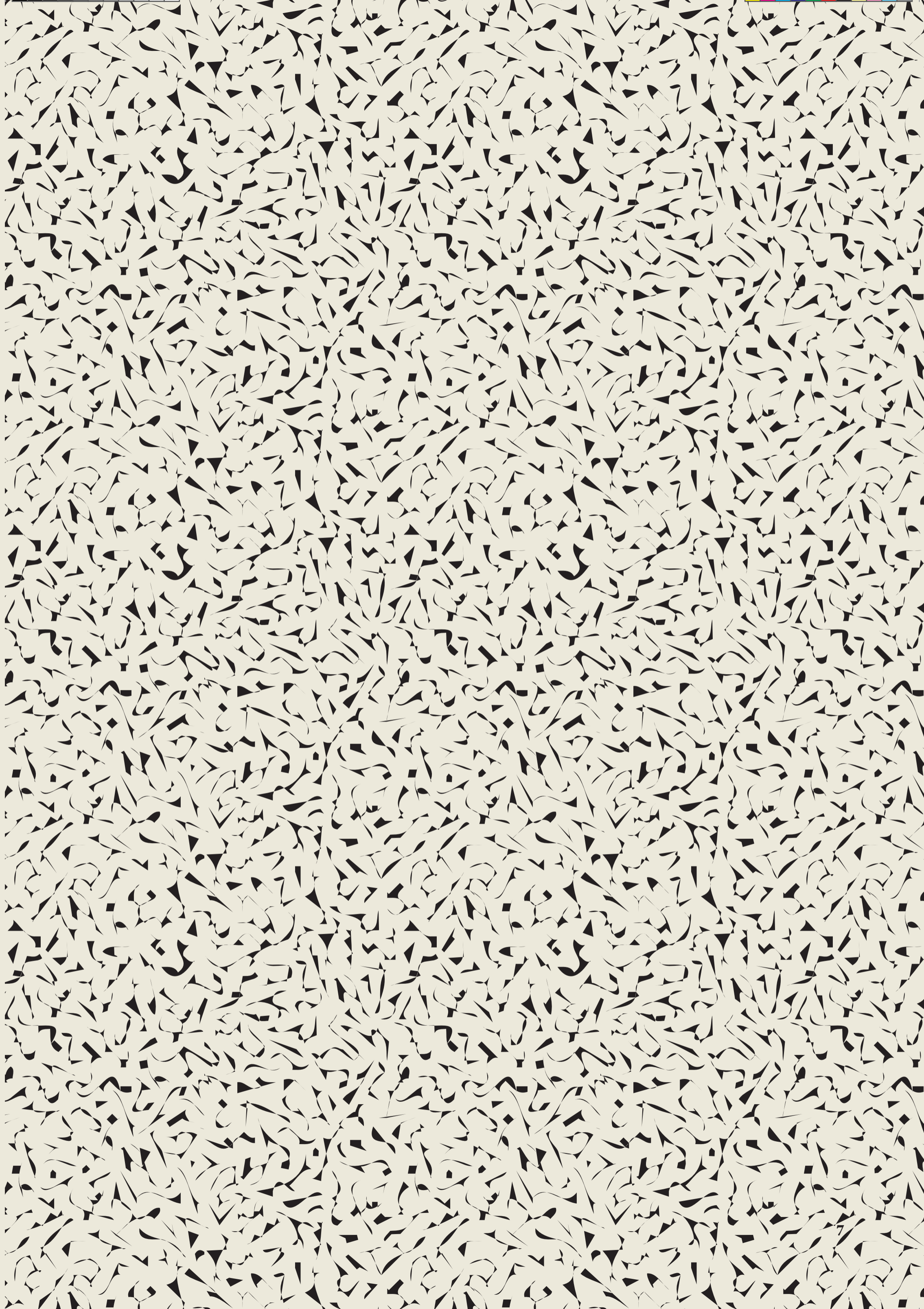
This field guide offers an introduction to recognizing the traces and fingerprints of online trackers, and content served through the Content Delivery Networks of the cloud. The term “the cloud” already suggest that it could be difficult to grasp, let alone recognize its particles. And indeed it is.

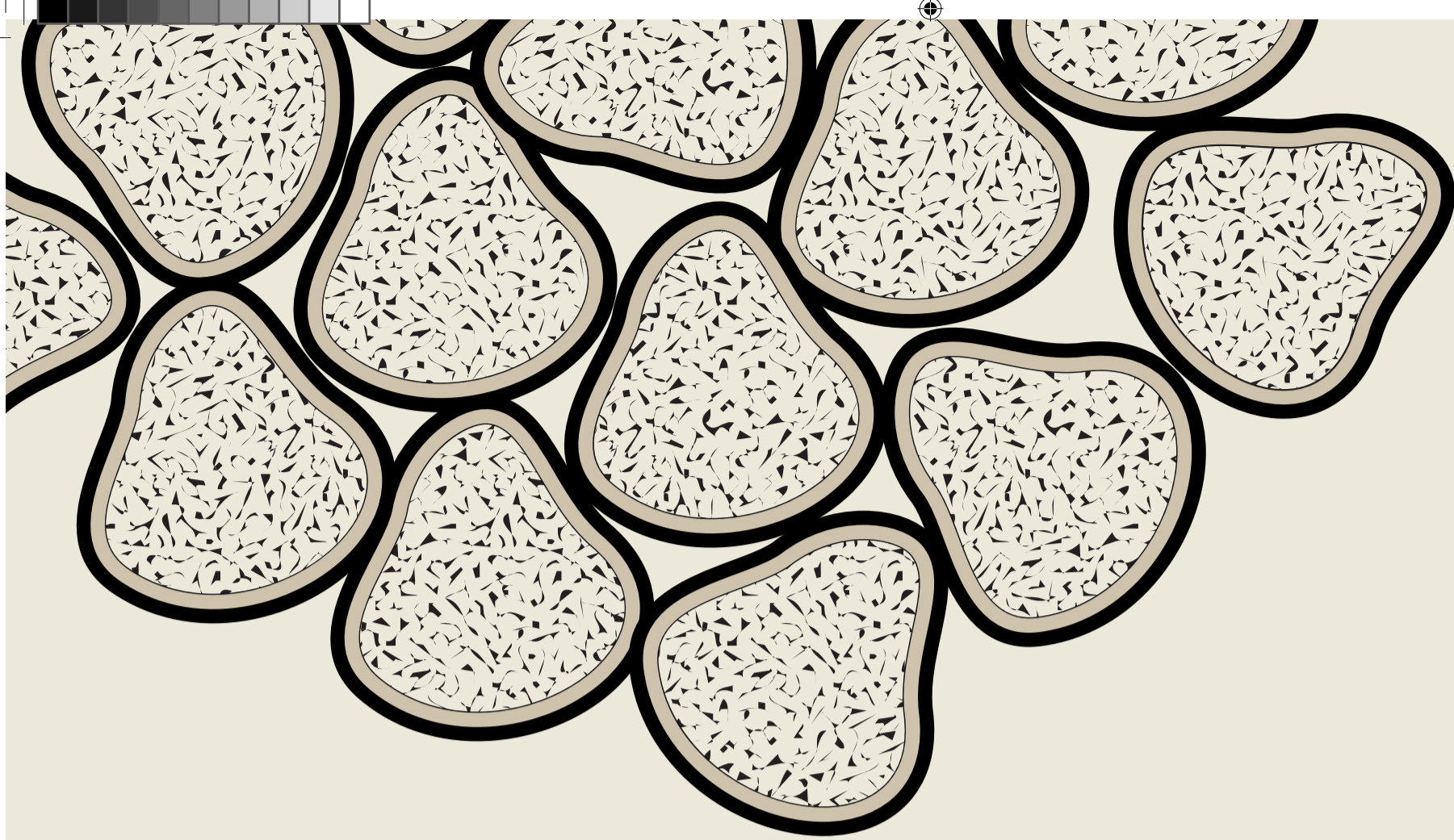
This field guide will provide you with a glossary of trackers and content delivery networks, as well as point you at the available equipment that can be used for your inquiry. This equipment contains a research browser (clean and de-personalized), specific browser settings as well as browser add-ons and even dedicated tools.

In this field guide, we use a wide defini-

tion of “tracker,” including a number of devices that allow for user-data collection, such as first-party tracking devices, bugs, widgets, third-party analytic services and further interfaces to the cloud. For the cloud consists of many types of services, of which hosting is probably the most widely known, here we focus only on Content Delivery Networks (CDNs).

The Tracker Guide to the Cloud will help you to better understand what happens when you open a website in your browser. Which content is being “pulled in” from the cloud, and which user data is being collected by the present trackers. It will also help you relate to cloud critique, and formulate your own viewpoints and research projects.





HOW TO

~ Find traces of trackers

Trackers, widgets, beacons, ads and analytical services are operating in the back end of many websites. These devices track user behaviour online and can share a range of datapoints about the tracked users to the associated services, such as IP address, demographic data, search history and cookie data.

Such trackers function as interfaces to the cloud, allowing data produced on one website to be shared with associated services. Some require the active participation of the user (through widgets), others are automated (through tags, web bugs, pixels and beacons). Most websites contain both first party and third party tracking devices, some of which are used to keep the page running, most, however, are deployed to track and profile web us-

ers for analytics, advertising and behavioural targeting. Whilst often not visible in the front end, they can be detected by tracing their fingerprints in the source.

The growing presence of such tracking devices across the web has led to the development of numerous tracker detection and blocking applications and plugins. In order to identify the presence of trackers on a set of websites, we have built a tool that repurposes the analytical capacities of the anti-tracking browser plugin Ghostery that informs users which companies are present on websites they visit. The Tracker Tracker tool allows to identify automated tracking devices for a set of URLs, based on the fingerprints provided by Ghostery and classifies them into widgets, trackers, ads and analytics.

The Tracker Tracker tool therewith allows to make visible the invisible web and has so far been used to define the connections between websites and tracking services present on them via widgets, beacons and cookies. It defines these relations based on their connection to the same tracking devices, giving insight into the fluidity of content.



HOW TO

~ Detect Content Delivery Networks

Detecting Content Delivery Networks (CDNs) from a set of websites (or a single website) involves a three-step process.

First, hostnames are fetched from the HTML of the website, then for each found hostname a Domain Name Server (DNS) lookup is done looking at the Canonical Name Record (CNAME) that specifies that the domain name is an alias of a canonical domain name.

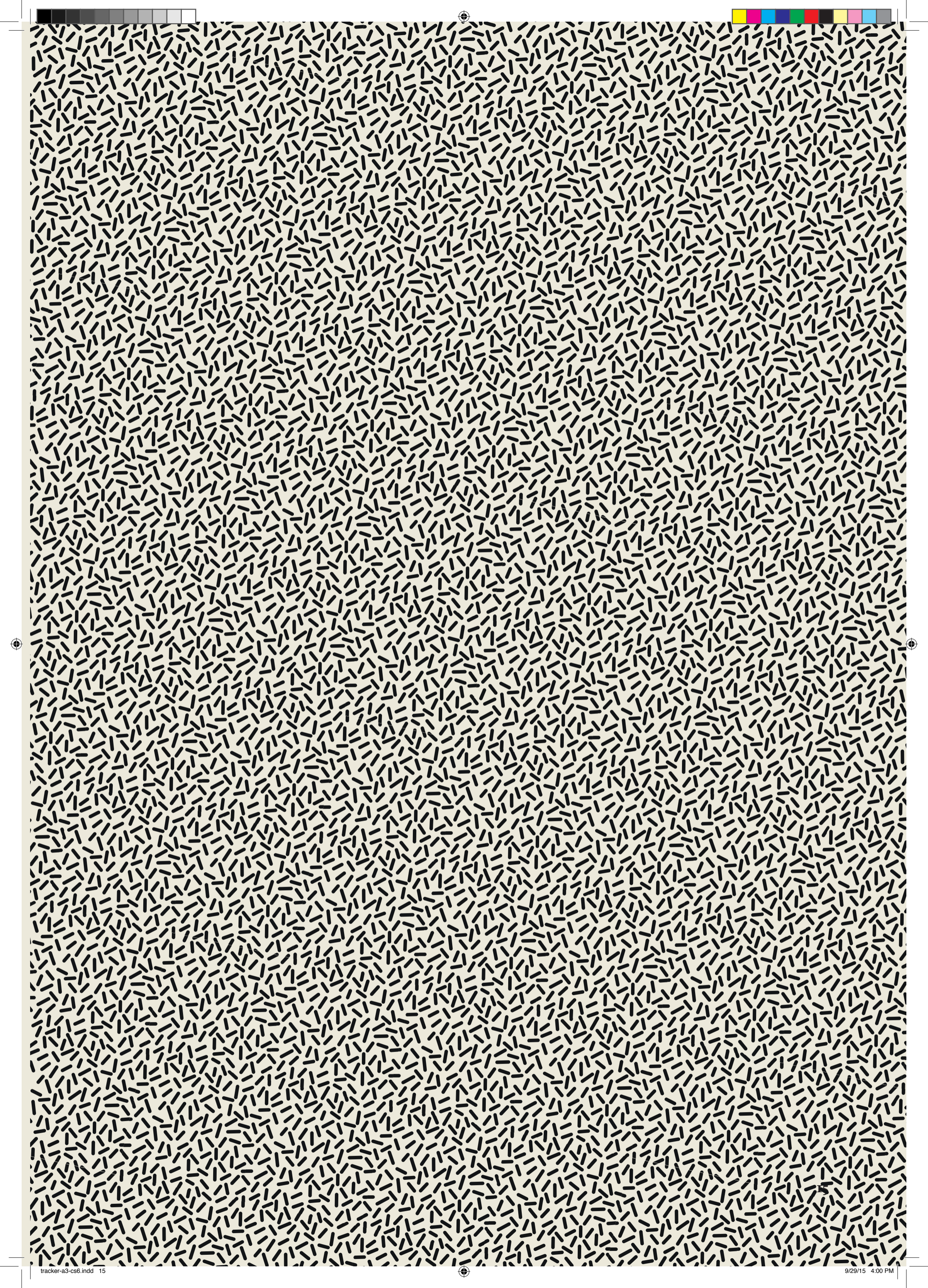
Lastly, these domain names are compared against a list of CDN fingerprints (see table 1).

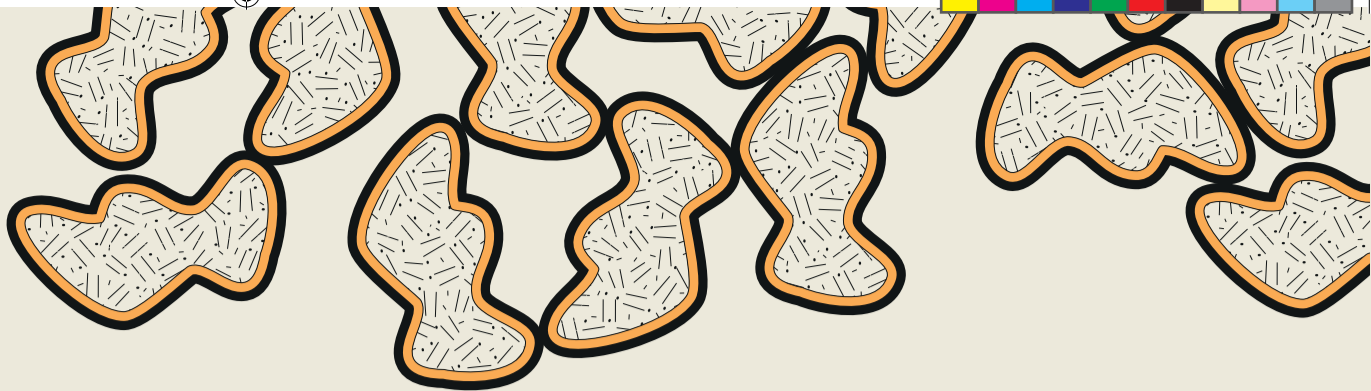
N.B. There are tools that automate this three step process.

see also → Equipment

Table I - Content Delivery Networks and their fingerprints. Source: CDN Planet.

FINGERPRINT	CDN
<i>Fingerprint</i>	<i>CDN</i>
.afxcdn.net	afxcdn.net
.akamai.net	Akamai
.akamaiedge.net	Akamai
.gslb.tbcache.com	Alimama
.cloudfront.net	Amazon Cloudfront
.att-dsa.net	AT&T
.cdn.bitgravity.com	Bitgravity
.bluehatnetwork.com	Blue Hat Network
.bo.lt	BO.LT
.cachefly.net	Cachefly
.cdn77.net	CDN77
.panthercdn.com	CDNetworks
.cdngc.net	CDNetworks
.gccdn.net	CDNetworks
.gccdn.cn	CDNetworks
.ccgslb.com	ChinaCache
.cloudflare.com	Cloudflare
.cotcdn.net	Cotendo
.systemcdn.net	EdgeCast
.edgecastcdn.net	EdgeCast
.fastly.net	Fastly
.google.	Google
googlesyndication.	Google
youtube.	Google
.googleusercontent.com	Google
.hwcdn.net	Highwinds
.internapcdn.net	Internap
.footprint.net	Level3
.llnwd.net	Limelight
.lxdns.com	lxdns.com
.netdna-cdn.com	MaxCDN
.netdna-ssl.com	MaxCDN
.netdna.com	MaxCDN
.instacontent.net	Mirror Image
.rncdnI.com	Reflected Networks
.simplecdn.net	Simple CDN
.swiftcdnI.com	SwiftCDN
.gslb.taobao.com	Taobao
.clients.turbobytes.com	Turbobytes
.voxcdn.net	Voxel
.vo.msecnd.net	Windows Azure
.ayi.b.yahoo.com	Yahoo
.yimg.	Yahoo





HOW TO

~ Deal with Dynamic Content

One of the difficulties in retrieving fingerprints, beacons and other traces of tracking and cloud activity, is the hyper-dynamic nature of the websites under study. On a site like Pinterest.com, even refreshing the page will render different content, sometimes served by different CDNs.

There is only one way of dealing with this, which is to work fast and collect your data in the shortest possible period of time. Furthermore, always note the date and time in which you found the fingerprints and other traces, and include a screenshot taken at the time you collect your data.

see also → Equipment



EQUIPMENT

~ Research
Protocol



Aim — CDN fingerprinting a selection of websites

1. Create a selection of websites to fingerprint
(For instance, the Alexa top 100 websites).

2. Enter each website into tool

www.cdnplanet.com/tools/cdnfinder.

3. Copy and paste “hostname” and “cdn” into a spreadsheet.

Aim — Analyzing which content type is hosted on which CDN per website (e.g. images, video, styling, scripts)

1. Per website, look up which content is hosted on which CDN and add this to the spreadsheet.

1.1. Open website and “view source” (in Firefox this can be found under Tools > Web developer > Page source).

1.2. Apple F (find) the “hostname” in the source code and determine what type of content it is by reading the extension of the URL (for example .jpg or .css or .js or .png).

Note: Evaluate if the site is a frontpage, or login page, common in social networking sites such as Facebook or Twitter, or whether we should go “deeper” into the site to show the platform with all its content.

If it is a site that does not necessarily require a login (such as Twitter, where you can view other profiles without a Twitter login) go to a public page, copy paste the URL and redo in the CDN finder.

Facebook also has some public pages, some groups: <https://www.facebook.com/barbarian.group>.

If the site does require a login (such as Facebook) you can follow the following steps:

2.1. “Walled gardens”, or sites behind login, have additional extra steps.

2.2. Open the website (or page with a lot of divers content) and view source.

2.3. Copy and paste source code into the Harvester tool and run it: <https://tools.digitalmethods.net/beta/harvestUrls/>.

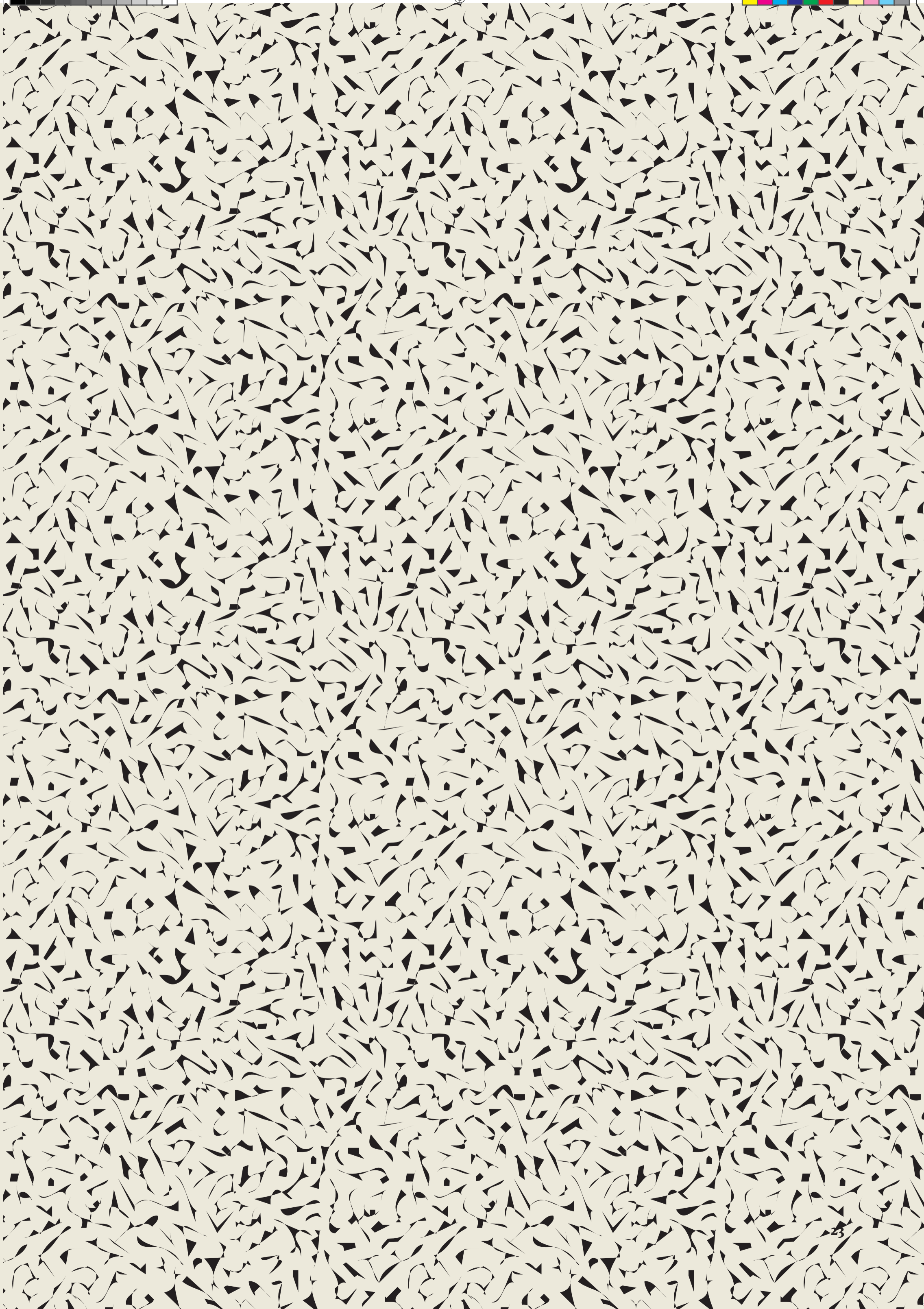
2.4. Evaluate which URLs are content links to the cloud and run these through a DNS resolver, such as <http://www.mxtoolbox.com/SuperTool.aspx>.

2.5. Update spreadsheet.

Aim — Annotating homepages with visible aspects of CDNs and trackers when a site is loaded in the browser.

For each website, if possible, locate where the CDN content is located and annotate on a screenshot:

1. Install Skitch (<http://skitch.com/>).
2. Take the lists of CDN's and trackers, and load the website in a browser that has Ghostery and Firebug plugins.
3. Annotate the screen shot.
4. Collect all information about the trackers and CDNs (and widgets) active in the page from the glossary.





EQUIPMENT

~ Tools and Plugins



Tracker Tracker

The Tracker Tracker tool detects from a predefined set of ‘fingerprints’ of web technologies, such as social buttons and trackers, and outputs a list of technologies used.

<https://tools.digitalmethods.net/beta/trackerTracker/>

CDN Finder

CDN Finder fetches the HTML and parses it with very simple logic to get to a list of hostnames. For each found hostname CDN Finder then does a DNS lookup, look at the CNAME and regex matches that hostname to our list of CDN hostnames.

<http://www.cdnplanet.com/tools/cdnfinder>

Ghostery

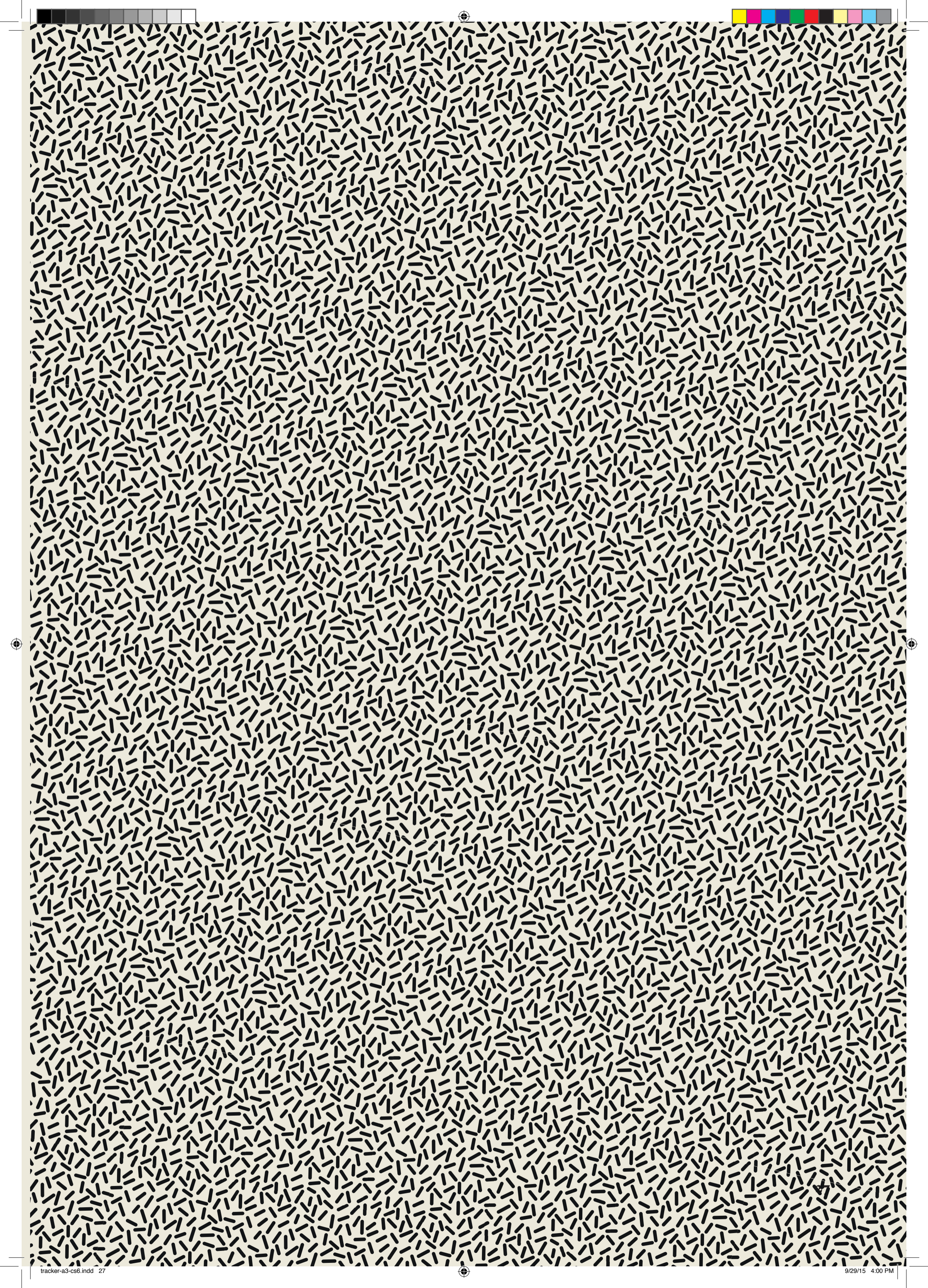
Ghostery alerts users about the web bugs, ad networks and widgets on visited web pages.

<http://www.ghostery.com/download>

Firebug

Firebug is a web development tool that, among other things, can be used to inspect HTML, dynamic elements, style and layout in real-time.

<https://getfirebug.com/downloads>



GLOSSARY

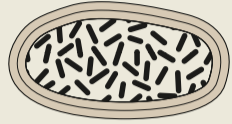
~ Introduction

Beside detecting the presence of trackers and CDNs on websites, we gave particular interest to the quantification and qualification of these interfaces to the cloud. Starting from the top 100 Alexa pages, we

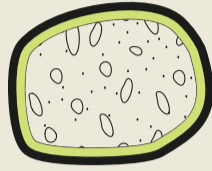
determined the key tracking devices and content delivery networks and developed a glossary which features both the total presence of CDNs within the URL set, the presence of trackers per category, but also the data in- and out-flows they enable. It functions as a field guide to detect CDNs and trackers, a description of the associated services and companies, as well as insights into what data is being collected or integrated into webpages.

The CDN glossary maps out the top 18 cloud service providers. It identifies a series of key fingerprints for each CDN based on the findings from <http://www.cdnplanet.com/tools/cdnfinder> and further contains a description of the cloud service as well as its network map and its geolocate presence and distribution.

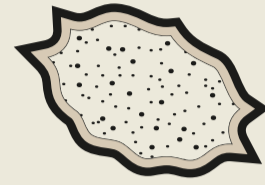
Type → Shape



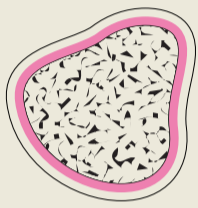
A D



A N A L Y T I C S



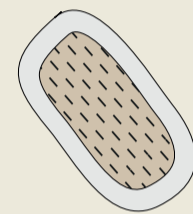
B E A C O N



T R A C K E R

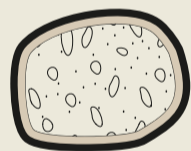


W I D G E T

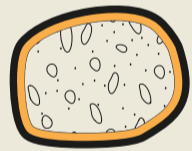


C D N

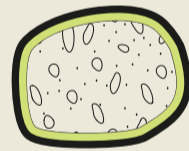
Data-in → Inner circle color



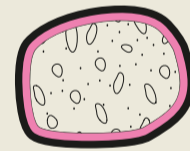
N O D A T A



S O C I A L

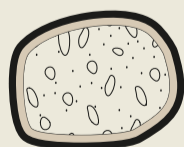


A D

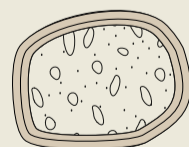


S U R V E Y

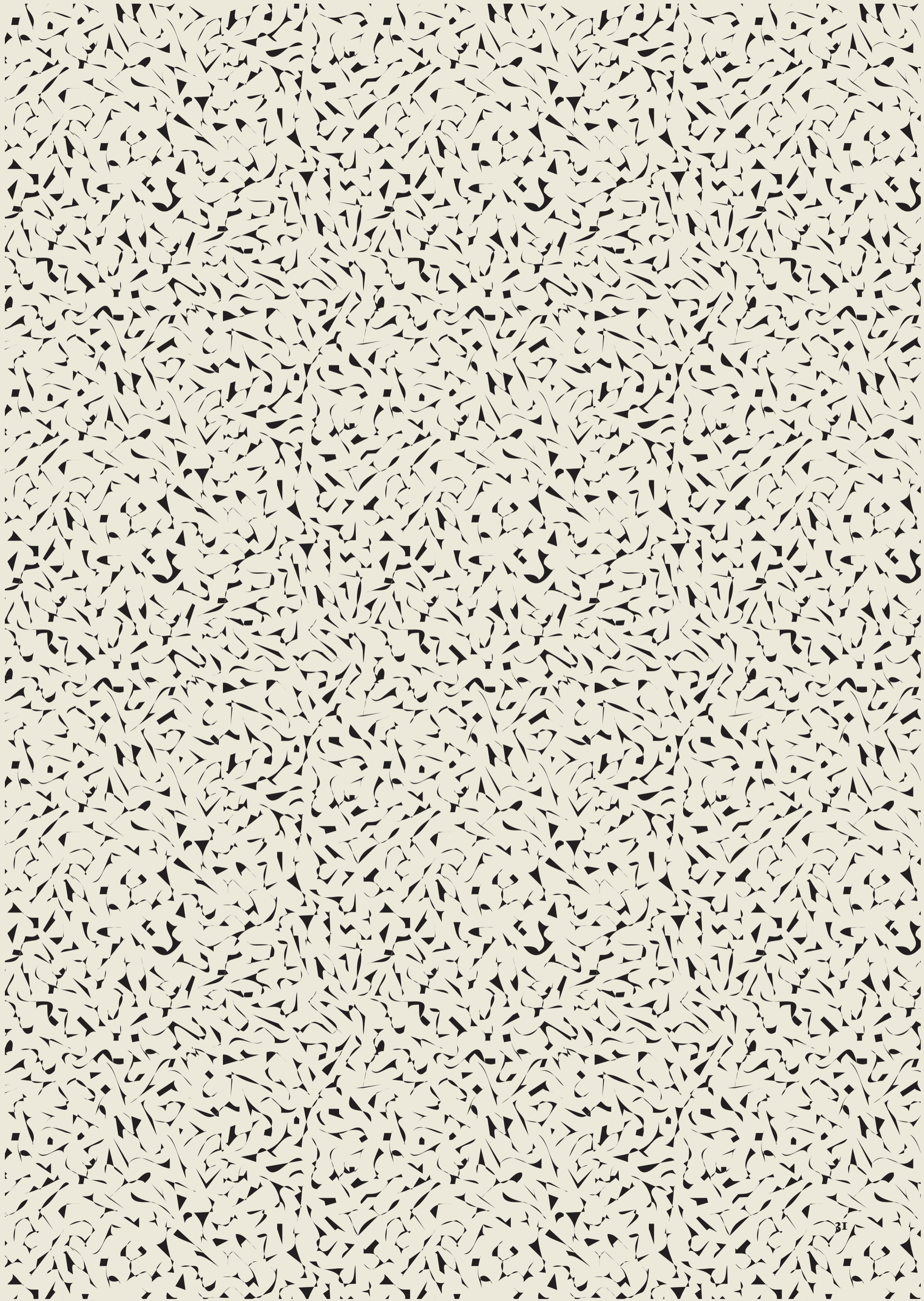
Data-out → Outer circle color



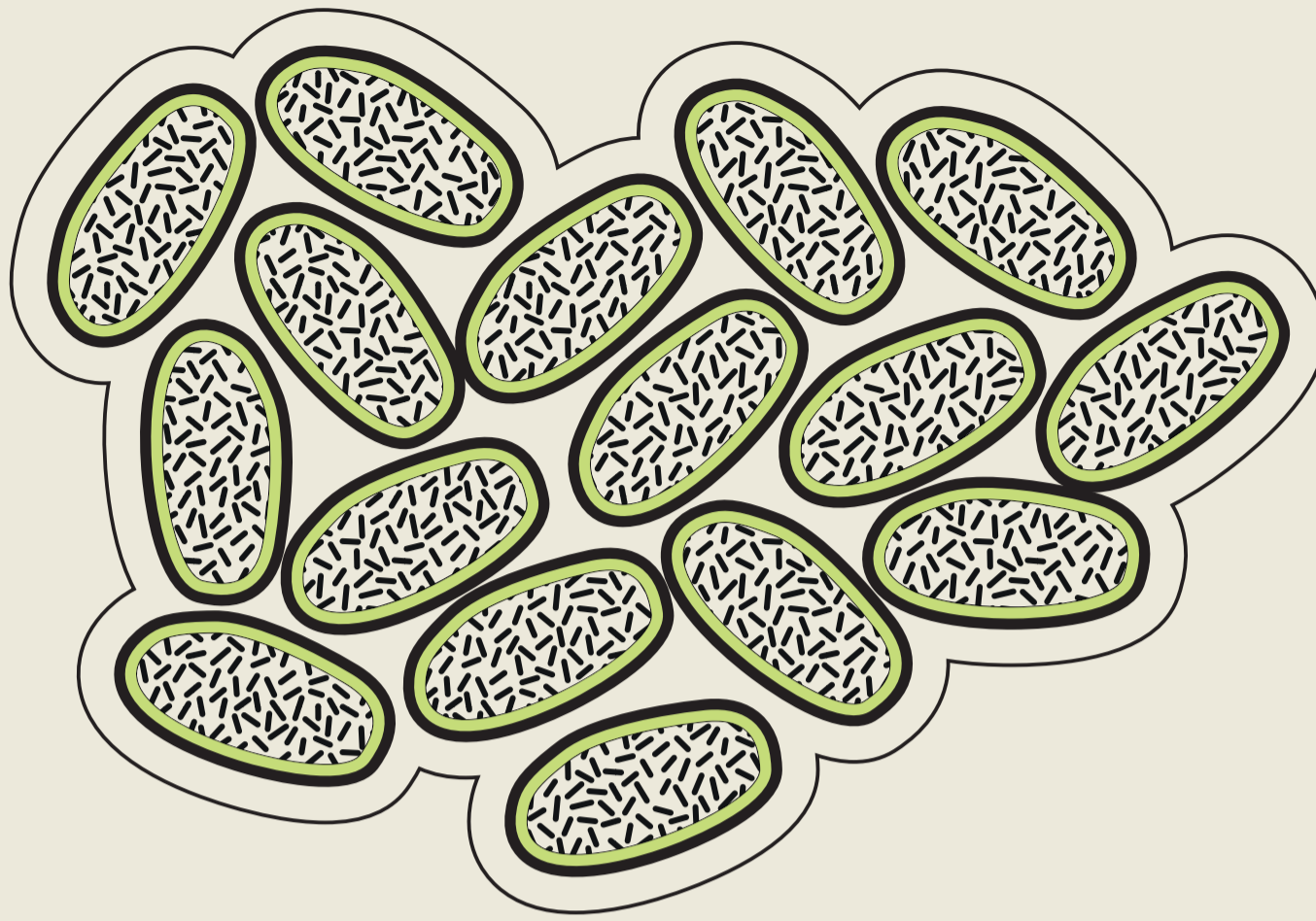
Y E S
D A T A O U T



N O
D A T A O U T

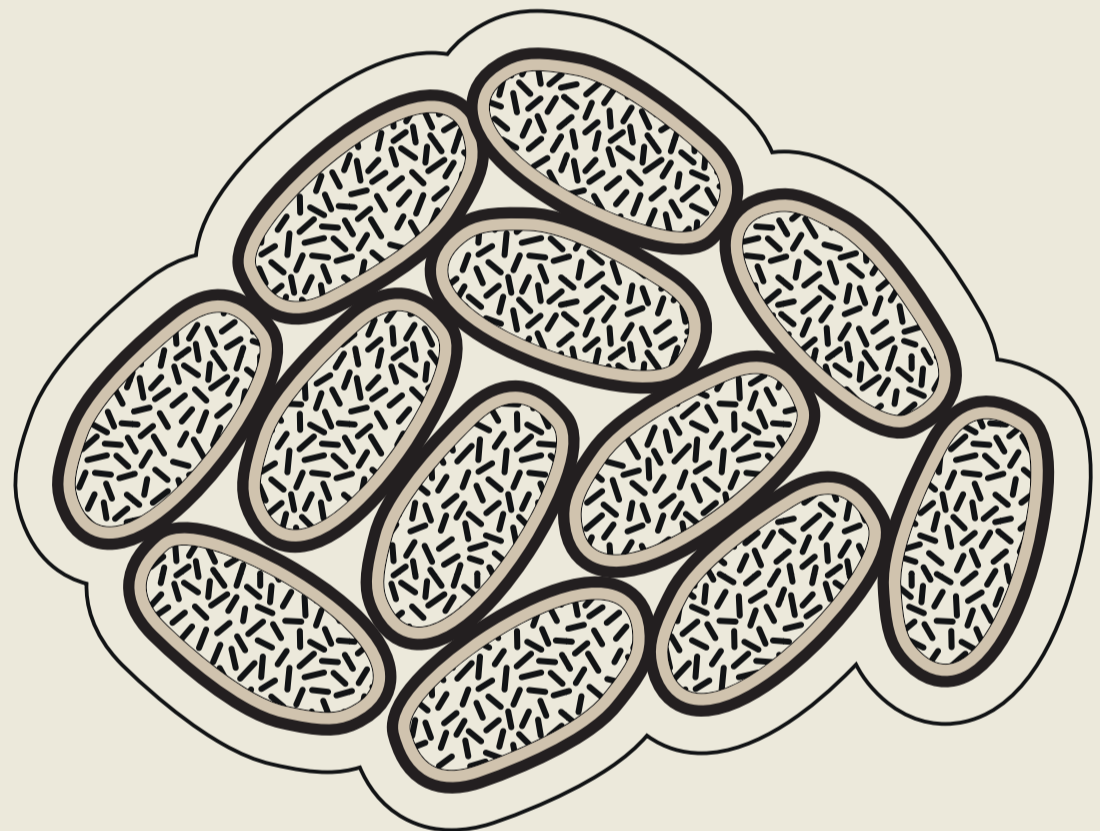


Ad



DOUBLECLICK

Google's DoubleClick products provide ad management and ad serving solutions to companies that buy, create or sell online advertising.



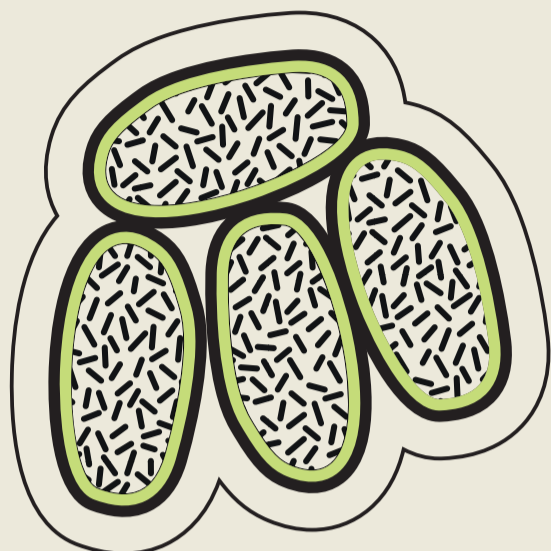
TRAFFICJUNKY

TrafficJunky uses a flexible CPM model, which is based on a self-serve live bidding platform. You select sites and ad placements you would like to advertise on and bid for traffic.

QUANTCAST

Quantcast measures and organizes the world's audiences in real-time so advertisers can buy, sell and connect with the people who matter most to them. Used by the top 10 media agencies, the world's largest brands and more than half of the top advertising supported publishers, Quantcast connects the planning, buying, and media fulfillment processes, delivering the marketplace's most consistent and accountable audiences

Ad



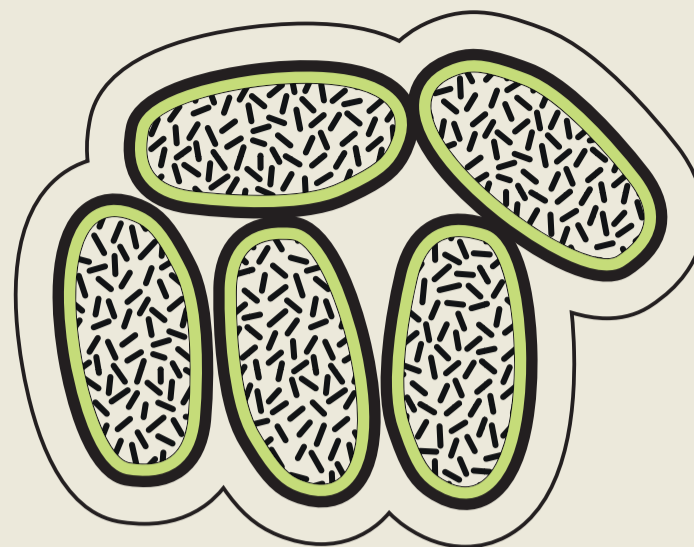
MICROSOFT ATLAS

We can help your business get the attention it deserves. We'll create the best advertising solution for you, and work with your small or medium-sized budget to do it. Expand your online presence today.



DOUBLECLICKDART

Google's DoubleClick products provide ad management and ad serving solutions to companies that buy, create or sell online advertising.



EBAY STATS

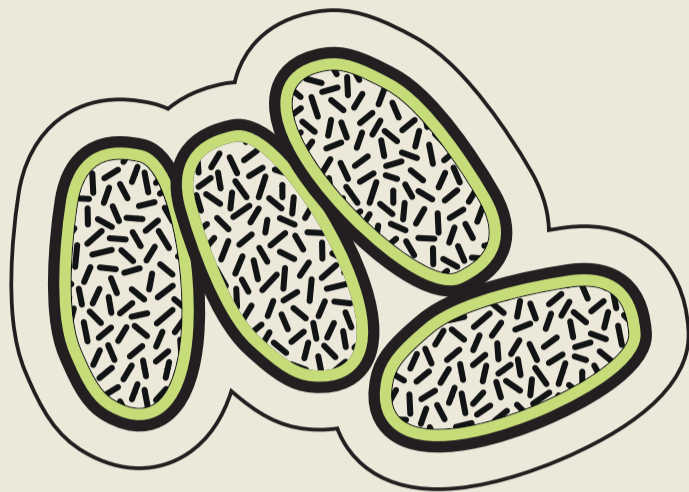
The eBay Partner Network is eBay's in-house affiliate program which operates in 13 countries globally. eBay Partner Network has some of the most generous affiliate commissions on the web. With Quality Click Pricing, the amount paid for each click varies depending on the quality of the traffic sent to eBay.



LEADBACK

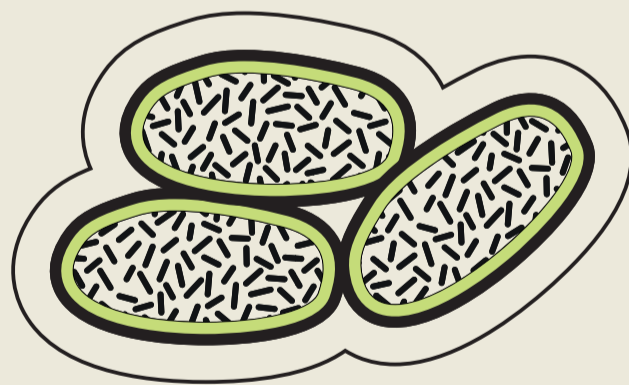
AOL Advertising provides advertisers, agencies and publishers with the most powerful, comprehensive and efficient online advertising tools available anywhere. Our industry-leading network, Advertising.com, helps advertisers reach highly-targeted audiences at scale, and helps publishers maximize revenue.

Ad



ADVERTISING.COM

AOL Advertising provides advertisers, agencies and publishers with the most powerful, comprehensive and efficient online advertising tools available anywhere. Our industry-leading network, Advertising.com, helps advertisers reach highly-targeted audiences at scale, and helps publishers maximize revenue.



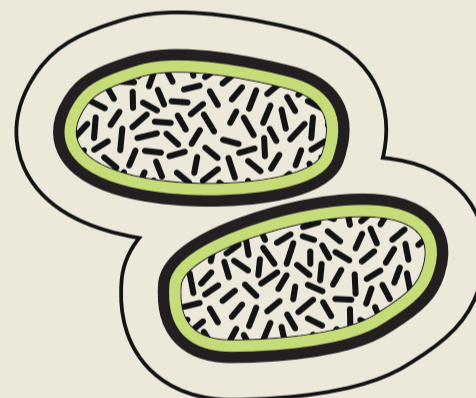
ADULT WEBMASTER EMPIRE

Adult Webmaster Empire (AWE), the web's premiere live webcam affiliate program.



DOUBLEPIMP

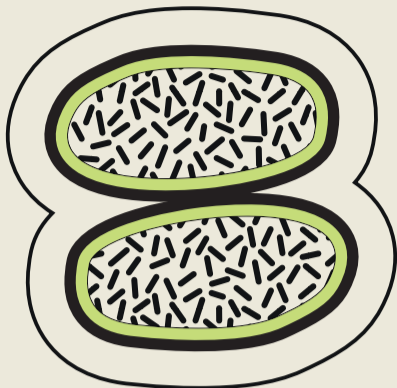
Doublepimp creates additional revenue sources for publishers and advertisers looking to squeeze every penny out of their traffic. Our network is 100% high-impact off page ad units such as overlays, popunders, message windows, and toolbars. These units provide the strongest branding and ROI conversion.



5MIN MEDIA

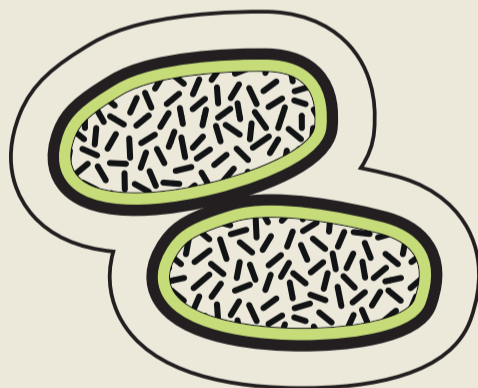
5min Media is the leading video syndication platform for lifestyle, knowledge and informative videos, and the largest independent video property in the US. Reinventing the cable network online, 5min reaches engaged and targeted audiences of passionate consumers through its network of 100s of lifestyle and niche websites.

Tracker



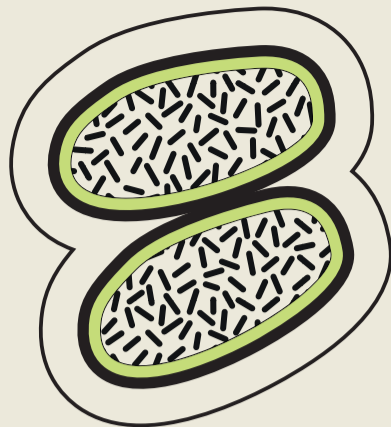
VALUECLICK MEDIAPLEX

Mediaplex provides cross-channel advertising technology solutions and services that enable marketers to achieve one-to-one messaging, greater efficiencies and a competitive edge through insightful reporting and analytics. Our team of industry experts focuses on putting the customer first, providing advanced technology solutions alongside consulting services for the greatest return on their marketing spend.



CBS INTERACTIVE

CBS Interactive builds media experiences based on the things people love most. Across brands like CBS.com, GameSpot, CNET, TV.com and CBSSports.com, we deliver highly engaging editorial programming that combines original, independent and user-generated content.



CRITEO

Criteo's advanced technology enables online e-commerce sites to re-engage with potential customers who have left their website via dynamic banners containing the most relevant product specific recommendations that are generated in real-time for each individual.



TRIBAL FUSION

Tribal Fusion is a digital marketing solutions company that drives superior results at all levels of the purchase funnel. Our company is built around dedicated vertical teams that leverage their industry-specific knowledge of 'what works' to create fully customized advertising solutions to help companies capitalize on opportunities at every level of the consumer decision process. From custom 'brand channels' to increase awareness to vertical-specific targeting tools to drive immediate purchase, Tribal Fusion offers smarter solutions that drive greater results.



ZEDO

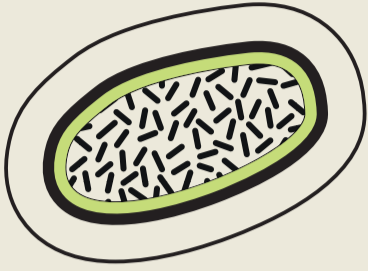
ZEDO, Inc. is a digital ad solutions company that offers products and services for a publisher's premium, self service, and remnant inventory. Products include ad network optimization, innovative rich media formats for publishers' direct sales teams, full featured ad serving, behavioral targeting data built into the ad server, an exchange-like platform for publishers to sell behavioral and DMA targeted inventory at high CPMs, a self service platform to allow advertisers to buy directly from a publisher, outsourced ad ops and more.



GOOGLE ADSENSE

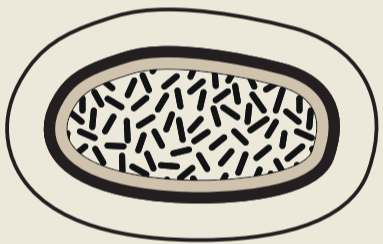
Google AdSense is a free program that empowers online publishers to earn revenue by displaying relevant ads on a wide variety of online content, including site search results, websites, mobile webpages and apps.

Tracker



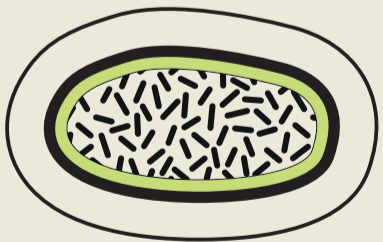
DOUBLECLICK SPOTLIGHT

Google's DoubleClick products provide ad management and ad serving solutions to companies that buy, create or sell online advertising.



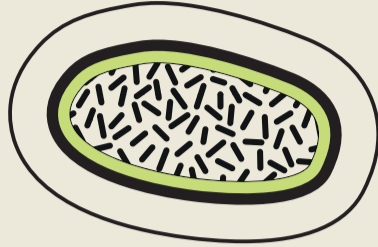
ADF.LY

We are a free URL shortener service with a twist..get paid to share your links on the Internet! Just register for an account and start shrinking. You get paid for every person that visits your URLs.



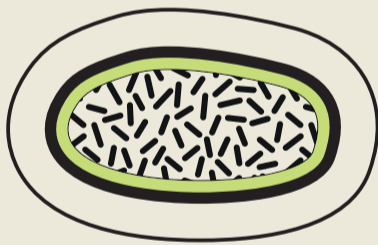
ADZERK

The best solution for managing your online ad sales and delivery.



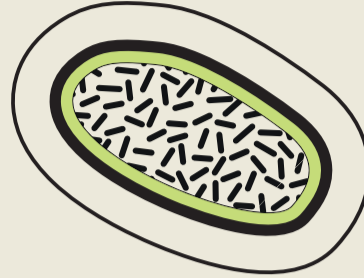
CONDUIT

Conduit is the fastest-growing network of web and mobile app publishers, with over 260,000 members and 250 million end users. The Conduit Network empowers publishers from global brands to independent developers to create, exchange and distribute apps, and to collaborate through business partnerships.



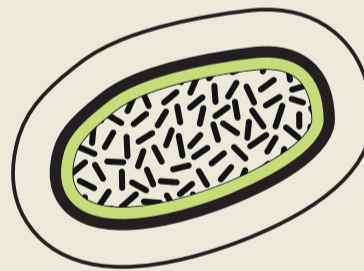
ADVERTISE.COM

Advertise.com is a premier platform for keyword PPC, display and Remarketing, enabling advertisers to reach a wider audience through our exclusive network of search engines and publisher properties.



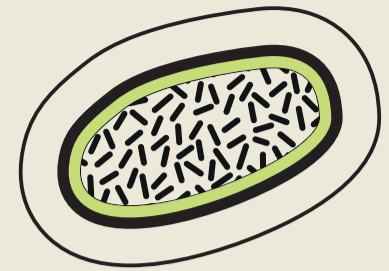
OPENX

OpenX is the world's leading independent provider of digital advertising technology that enables businesses to manage and maximize their ad revenue. OpenX products, including OpenX Enterprise and OpenX Market, provide a comprehensive revenue serving platform by combining ad serving with a unique ad exchange.



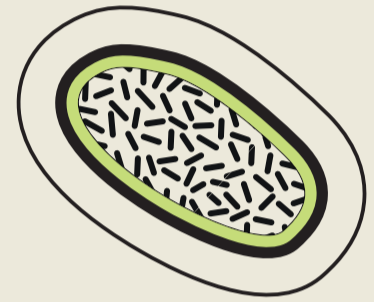
GOOGLE ADWORDS CONVERSION

No matter what your budget, you can display your ads on Google and our advertising network. Pay only if people click your ads.



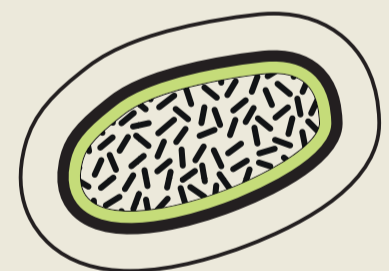
ADLANTIS

AdLantis is an online ad management system to deliver, manage, and optimize your ads.



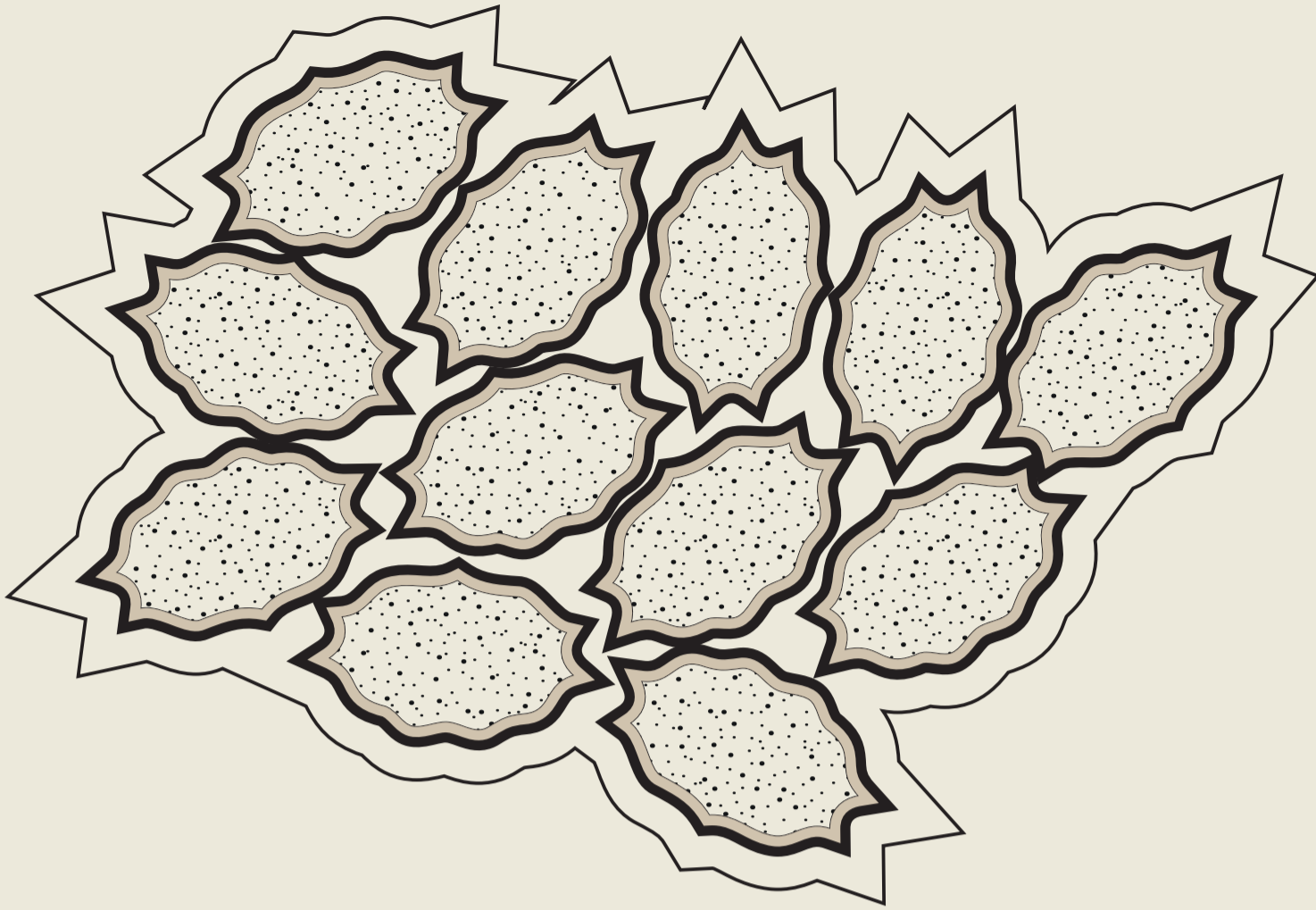
SMART AD SERVER

Smart AdServer develops and markets premium ad serving solutions for media agencies and publishers to manage display campaigns for Web, mobile and iPad/tablets.



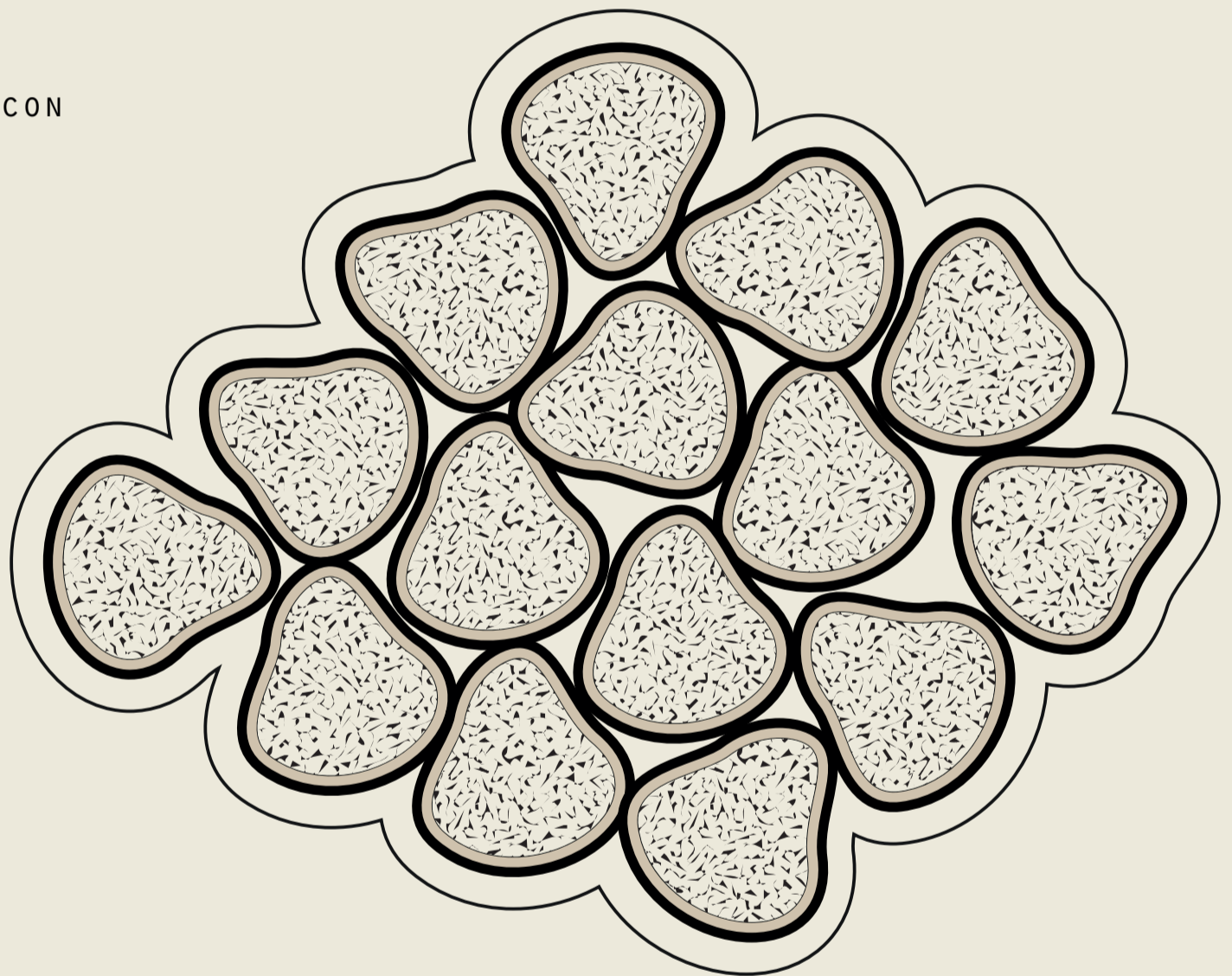
MICROAD

MicroAd offers clients with several ad delivery options, including behavioral targeting and retargeting. CyberAgent, Japan's largest Internet advertising agency in terms of revenues is the parent company of MicroAd, providing a tremendously strong sales channel through which most clients have been won.

Tracker

SCORECARD RESEARCH BEACON

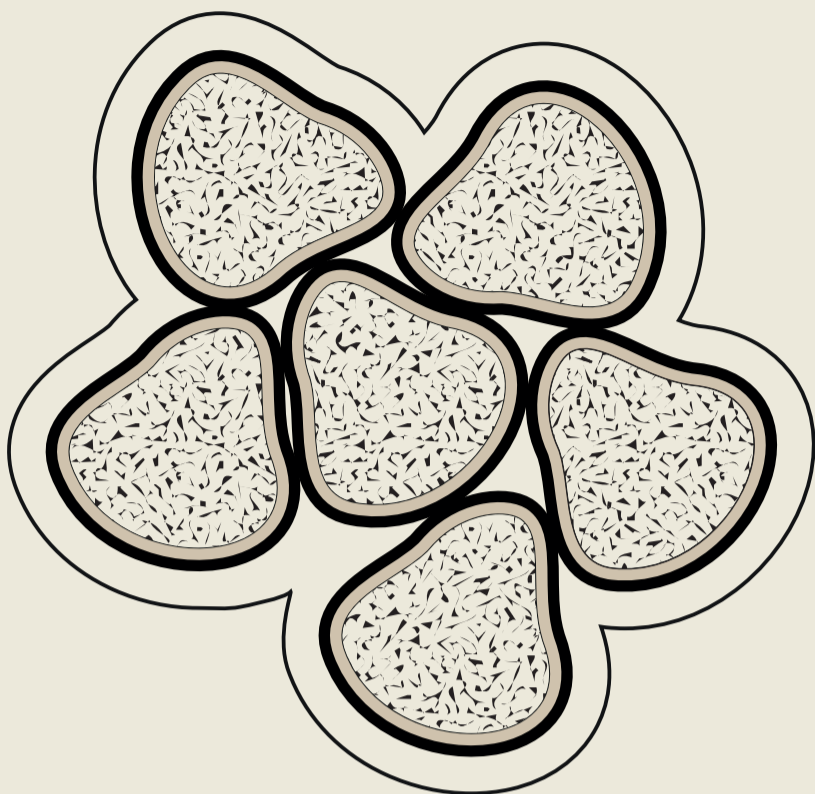
"Full Circle Studies is a leader in the Internet market research industry. Companies across a wide variety of industries turn to Full Circle Studies to conduct research to better understand the likes and dislikes of consumers. The reports that are compiled from these projects help these companies to improve their products and services."



OMNITURE

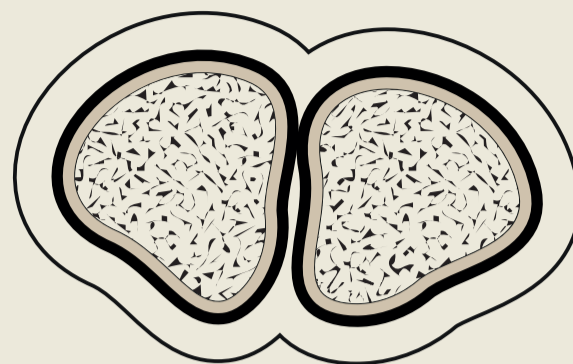
Omniure Insight is now Adobe Insight, powered by Omniure. Adobe Insight enables organizations to quickly analyze large volumes of rapidly evolving data in real-time. Adobe Insight is part of the Adobe Digital Marketing Suite of applications for online business optimizations.

Tracker



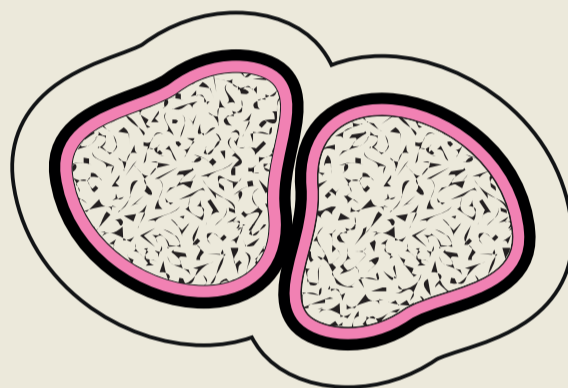
DYNAMIC LOGIC

Millward Brown is one of the world's leading research agencies and is expert in effective advertising, marketing communications, media and brand equity research.



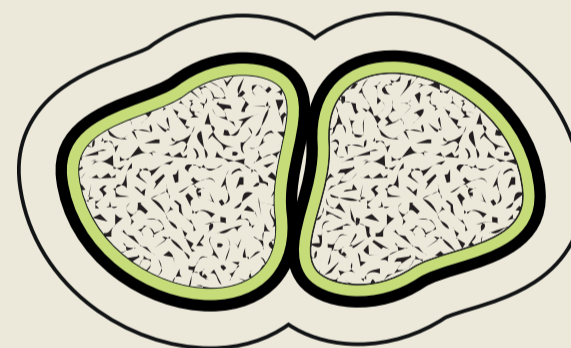
RAMBLER

[Rambler is] a portal that offers users a variety of quality services in the field of navigation, communications, media and entertainment on the Internet.



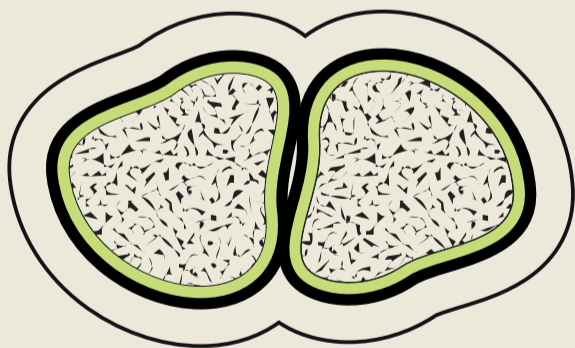
VALUED OPINIONS

Our research survey websites, including this site (www.valuedopinions.com), provide survey respondents with the opportunity to express opinions and attitudes about the most current issues. We conduct research that involves people from around the world, and we aggregate and anonymize such research (stripped of any personally identifiable information) before sending research results to our clients.



ACERNO

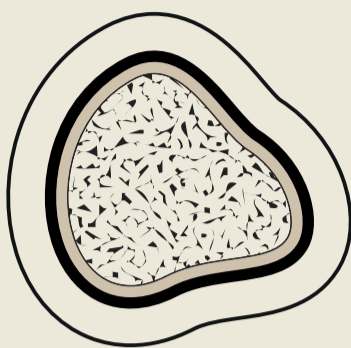
Akamai has created a digital operating environment for the Web. Our global platform of thousands of specially-equipped servers helps the Internet withstand the crush of daily requests for rich, dynamic, and interactive content, transactions, and applications.



VISUAL WEBSITE OPTIMIZER

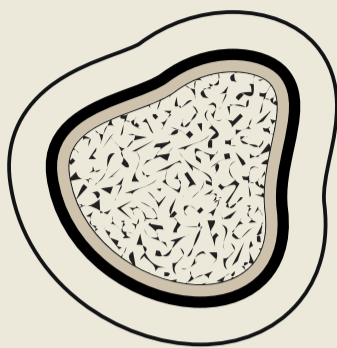
Visual Website Optimizer is an easy to use A/B testing tool featuring point-and-click test designer and WYSIWYG editor for creating variations.

Tracker



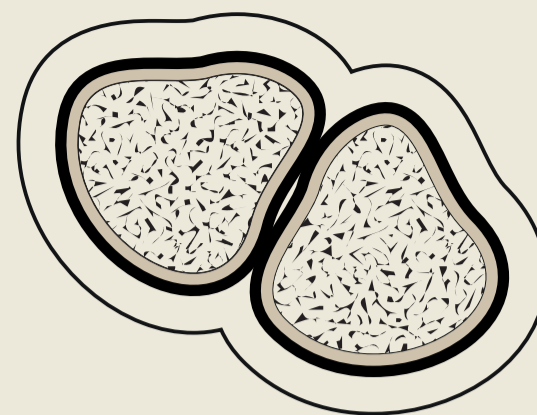
WEBTRENDS

Our diverse products and services portfolio includes real-time web analytics, unified digital analytics across mobile, social and web, optimization + testing, audience segmentation, targeting + scoring, mobile and social apps creation and Facebook Page management, professional consulting services and general digital marketing wizardry.



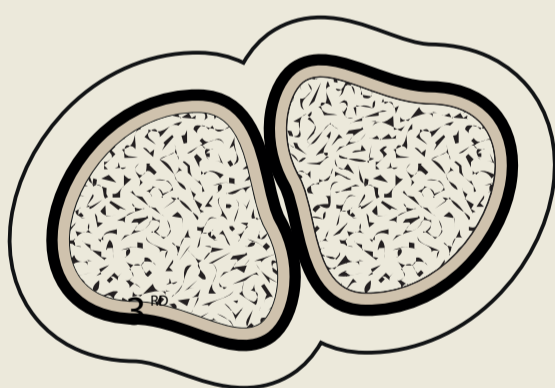
ASK.COM

At Ask.com, our mission is to answer questions - personal and professional, easy and difficult, silly and serious- with the best information from the web and from real people - all in one place.



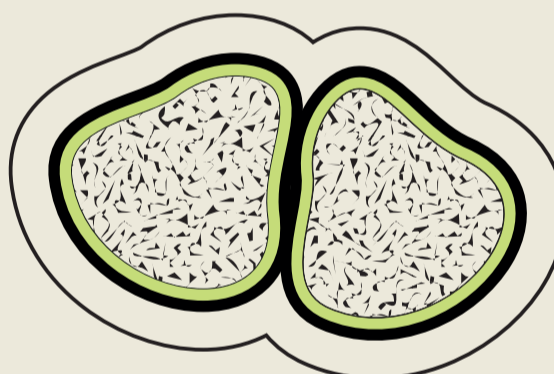
VOICEFIVE

VoiceFive is a leader in the Internet market research industry. Companies across a wide variety of industries turn to VoiceFive to conduct research to better understand the likes and dislikes of consumers. The reports that are compiled from these projects help these companies to improve their products and services. Companies employ VoiceFive to carry out a variety of projects including attitude and usage studies, discrete choice analyses and recall studies.



WORDPRESS STATS

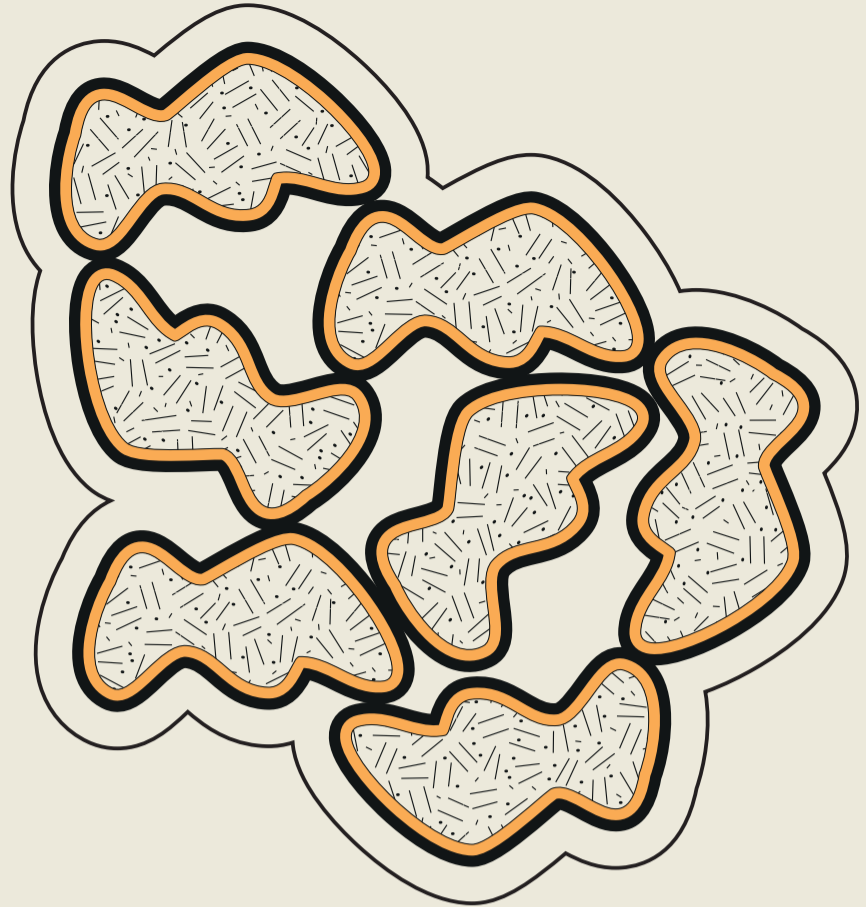
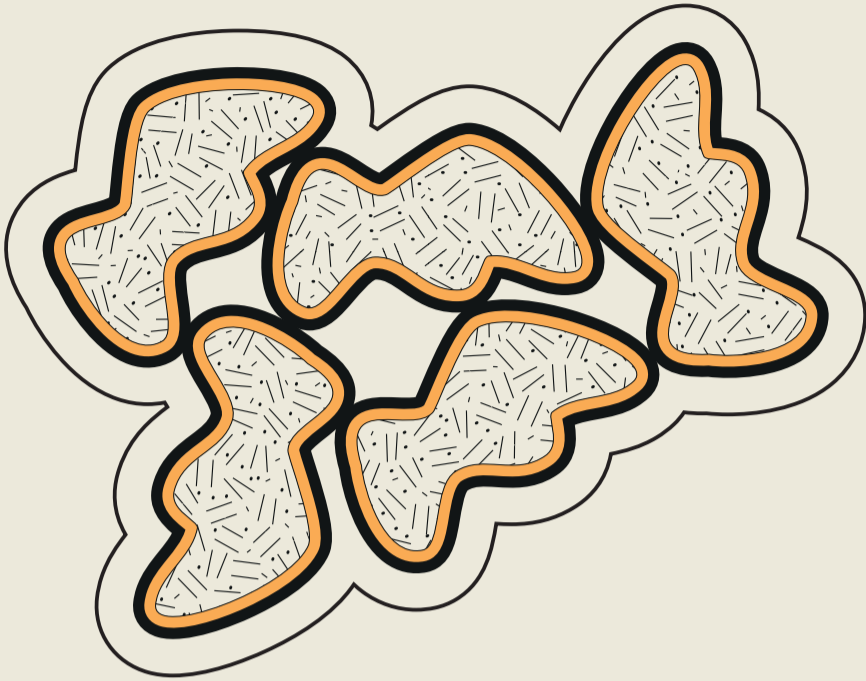
We created WordPress.com, a hosted version of the open source package where you can start a blog in seconds without any technical knowledge.



MEDIAMIND

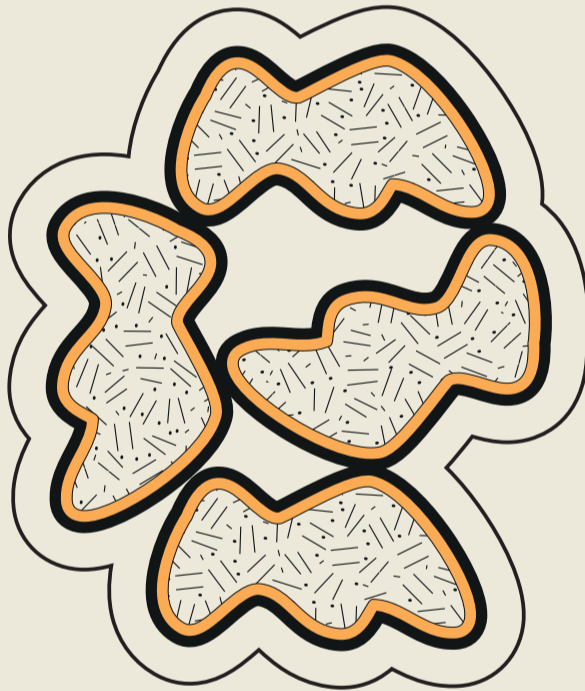
MediaMind, a division of DG, is a global provider of digital advertising solutions that optimize the use of media, creative and data for enhanced campaign performance. Our unique platform incorporates display ad serving, search, rich media, video, dynamic ads, mobile and emerging media; providing marketers with a cross-channel view of ad campaigns. Our leading MediaMind Rich Media and Video capabilities service the most innovative and inspiring global digital campaigns.

Widget



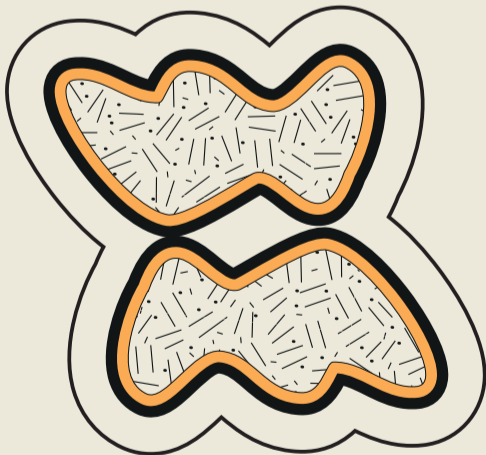
FACEBOOK SOCIAL PLUGINS

Facebook gives people the power to share and makes the world more open and connected. Millions of people use Facebook everyday to keep up with friends, upload an unlimited number of photos, share links and videos, and learn more about the people they meet.



TWITTER BUTTON

Twitter is a real-time information network that connects you to the latest information about what you find interesting...At the heart of Twitter are small bursts of information called Tweets. Each Tweet is 140 characters in length.



LIVE INTERNET

The largest blogging service in Russia, began its work in July 2002. Service provides a wide range of personal settings, and allows the user to maintain an electronic version of the online diary, to share their experiences, thoughts and opinions with other users, upload photos, favorite music, videos, discuss and comment on entries of other users, exchange private messages, create communities of interest and much more.

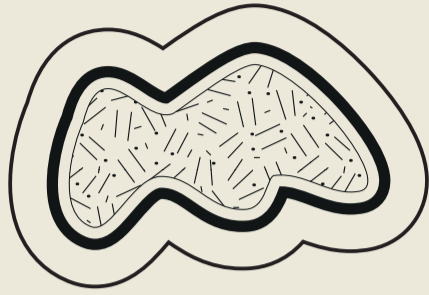
GOOGLE +

Google's mission is to organize the world's information and make it universally accessible and useful.

TUMBLR

Tumblr lets you effortlessly share anything. Post text, photos, quotes, links, music, and videos, from your browser, phone, desktop, email, or wherever you happen to be.

Widget



FACEBOOK CONNECT

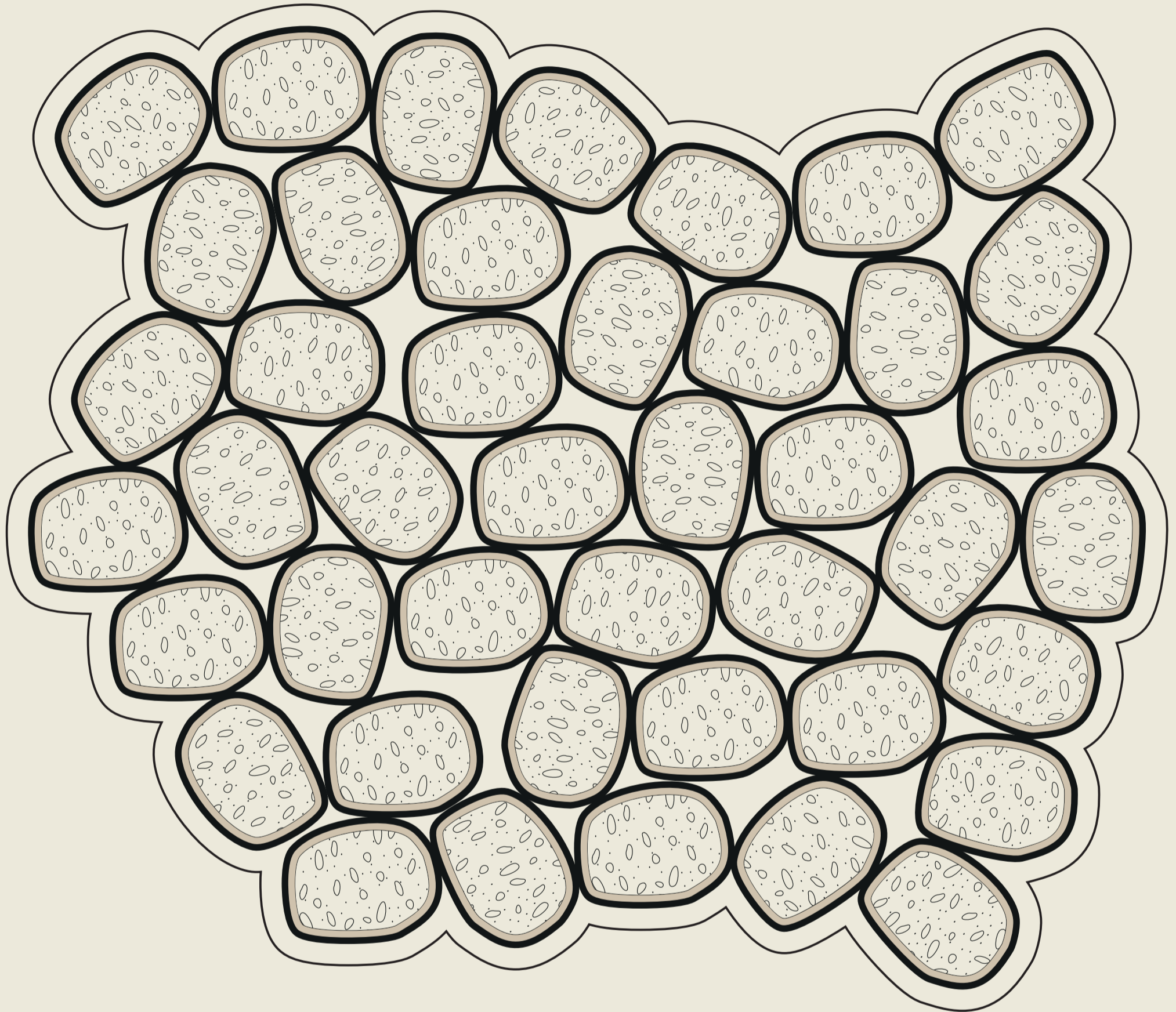
Facebook gives people the power to share and makes the world more open and connected. Millions of people use Facebook everyday to keep up with friends, upload an unlimited number of photos, share links and videos, and learn more about the people they meet.



DIGG WIDGET

Digg is a place for people to discover and share content from anywhere on the web. From the biggest online destinations to the most obscure blog, Digg surfaces the best stuff as voted on by our community.

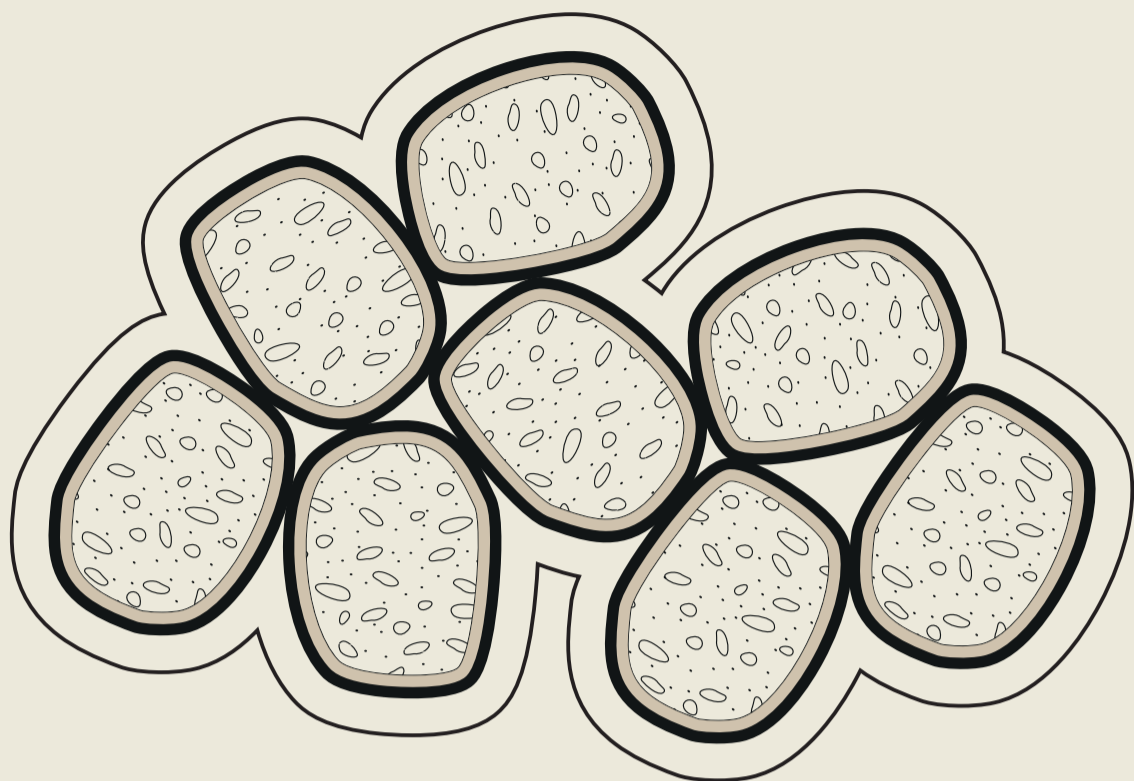
Analytics



GOOGLE ANALYTICS

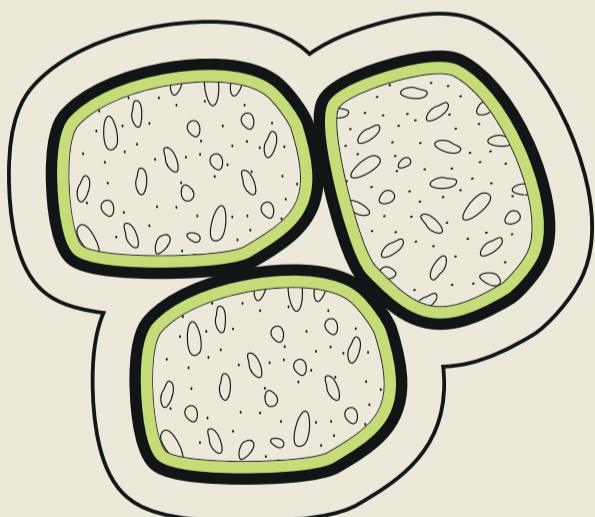
Google Analytics gives you insights into your website traffic and marketing effectiveness. We help you buy the right keywords, target your best markets, and engage and convert more customers.

Analytics



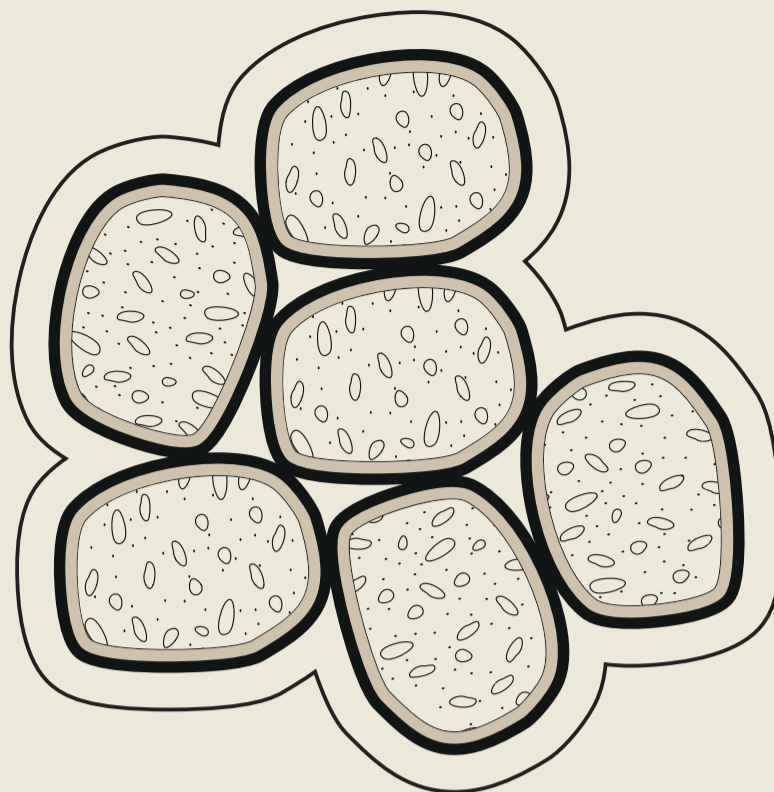
NETRATINGS SITECENSUS

SiteCensus is a browser-based audience measurement tool that provides in-depth tracking and analysis of your Web site users, site performance and other critical measurement data. Examine your site performance, usage trends, content and product placement, visitor loyalty, search engine performance and visitor behavior.



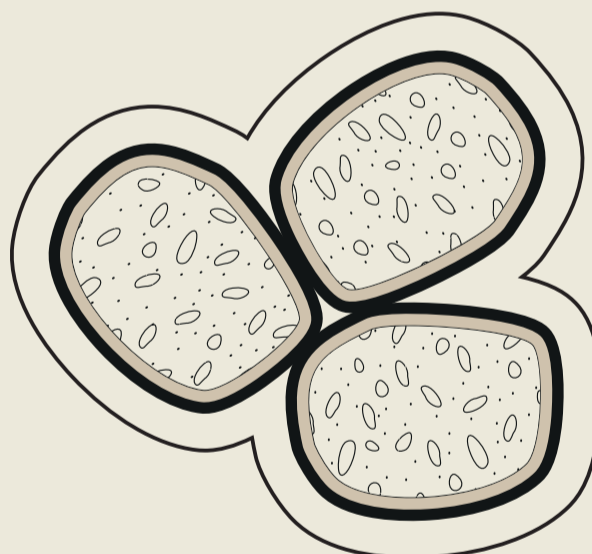
ADPLAN

AdPlan is a mobile advertising 'effectiveness measurement' platform. AdPlan is powered by the Platform ID technology provider.



TNS

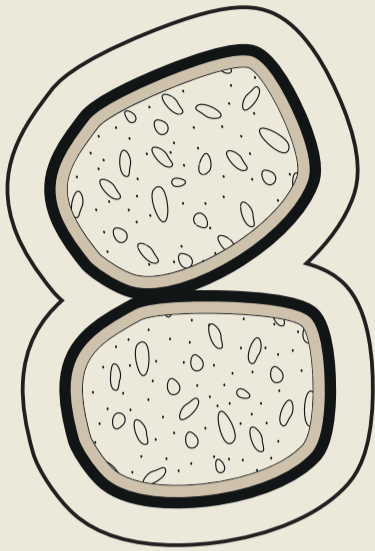
TNS is the world's largest Custom Market Research specialists. We provide quality marketing information delivered by Global Industry Sector expert consultants, innovative Market Research Expertise across the product life-cycle, in 80 countries.



YAHOO ANALYTICS

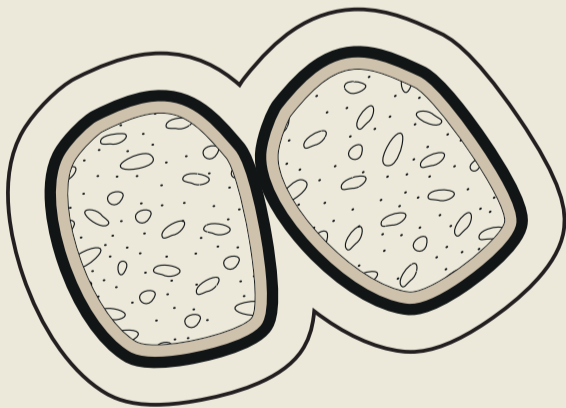
Yahoo! Web Analytics is a highly customizable, enterprise-level website analytics system designed to help website businesses increase sales and visitor satisfaction, reduce marketing costs and gain new insight on online customers.

Analytics



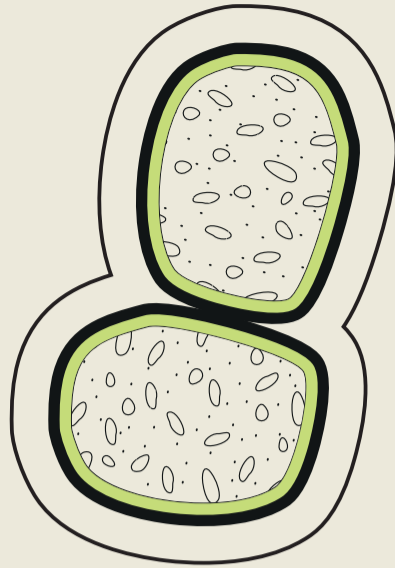
FORESEE

As a pioneer in customer experience analytics, ForeSee continuously measures satisfaction across customer touch points and delivers critical insights on where to prioritize improvements for maximum impact. Because ForeSee's superior technology and proven methodology connect the customer experience to the bottom line, executives and managers are able to drive future success by confidently optimizing the efforts that will achieve business and brand objectives.



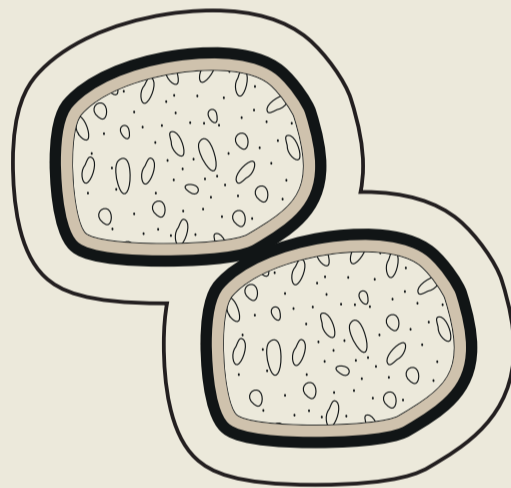
INFONLINE

INFOnline is a provider of Internet audience measurement services in Germany.



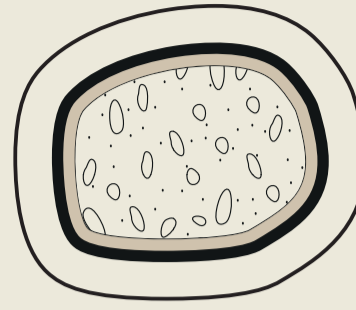
VISUAL REVENUE

Visual Revenue Inc. provides a predictive analytics solution that helps online media organizations set a better front page.



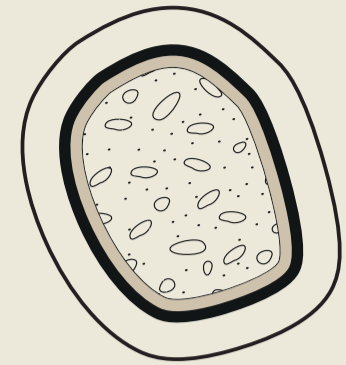
CLICKTALE

ClickTale is the industry leader in Customer Experience Analytics (CEA), providing businesses with revolutionary insights into their customers' online behavior. Over 70,000 businesses rely on ClickTale to optimize website performance, improve usability and dramatically increase conversion rates.



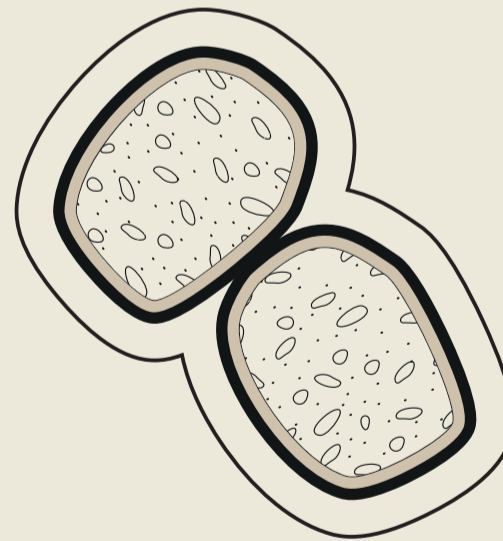
CNZZ

www.cnzz.com [is an] investment by the internationally renowned venture capitalists IDG network technology services company. China's Internet is the most influential free traffic statistics service provider, focused on the Internet, various types of site to provide professional and authoritative, independent third-party data for statistical analysis.



YANDEX.METRIX

Yandex.Metrica is a tool that measures website traffic, analyzes user behaviour and ad efficiency. Metrica works just like any other web counter: a code embedded on the pages of your website registers and gathers detailed information about each user visit. Advertisers using Yandex.Direct and Yandex.Market can find out exactly what happens on their website once a user clicks on their ad. Metrica also offers a wide range of tools for analysing various cross sections of the website audience, which can be used for fine-tuning ad campaigns. You can even watch every click, scroll and keystroke performed on your website as if you were looking through the eyes of the visitor.



CROWD SCIENCE

Crowd Science is revolutionizing audience targeting and segmentation by providing innovative solutions and technologies to online publishers and brands.

Akamai

DESCRIPTION

Akamai is the CDN market leader in terms of revenues and has been around since 1998. Akamai's EdgePlatform is one of the world's largest distributed computing platforms. 95,000 servers in 1,900 networks across 71 countries handle a significant portion of the world's Web traffic.

NETWORK MAP

Akamai has more than 1000 POPs, located all over the world including continents like South-America, Africa and Middle-East.

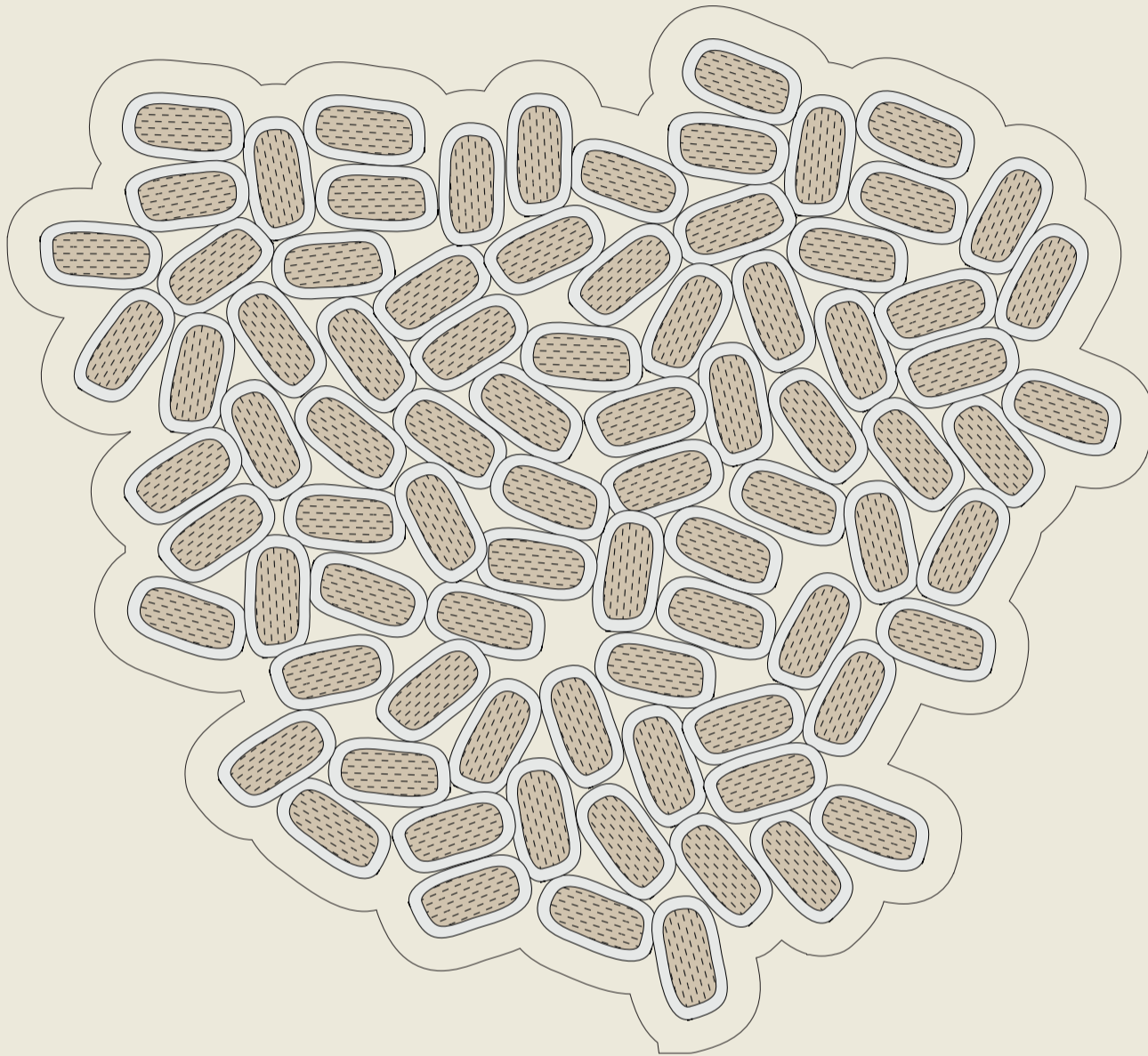
LOCATION

Cambridge, Massachusetts, US

FINGERPRINTS

*.akamai.net;
.akamaiedge.net*

Google



DESCRIPTION

Google is an American multinational corporation which provides Internet-related products and services, including Internet search, cloud computing, software and advertising technologies.

NETWORK MAP

Google Inc. currently owns and operates 6 data centers across the U.S., plus one in Finland and another in Belgium. On September 28, 2011 the company has announced to build data centers in Singapore, Hong Kong and Taiwan, which will be operational in one to two years.

LOCATION

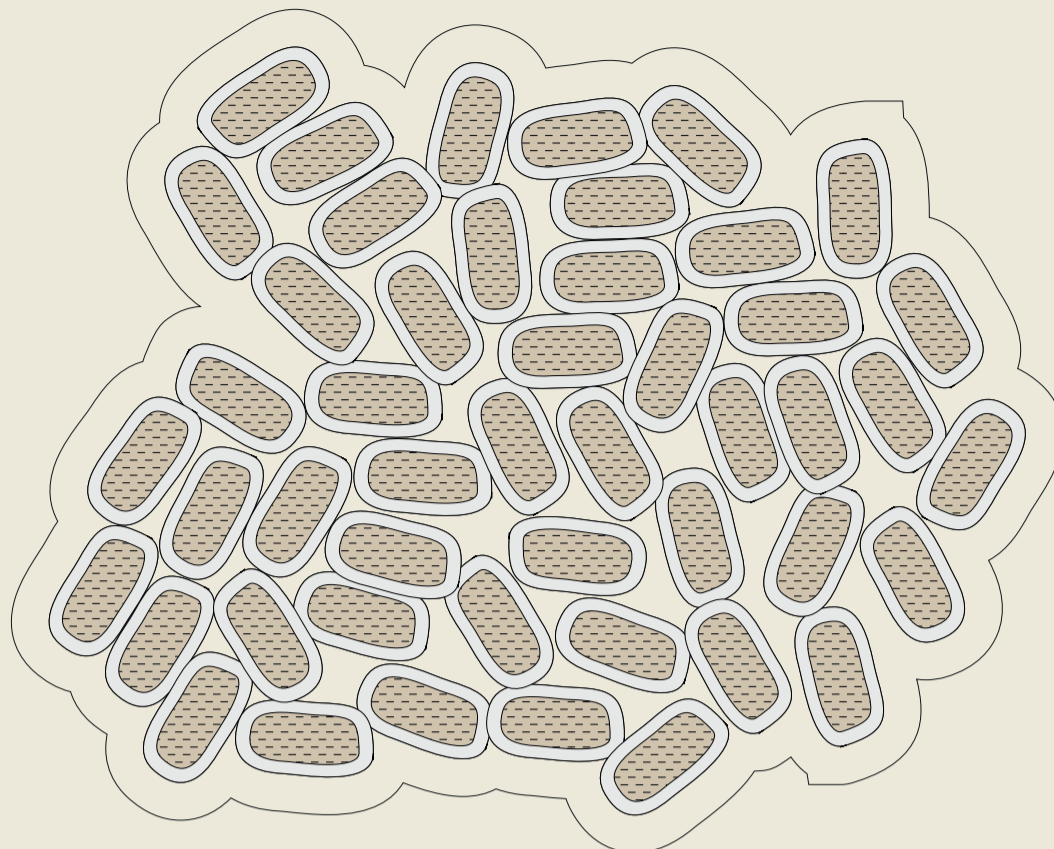
*Mountain View,
California, United States*

FINGERPRINTS

.google.;
googlesyndication.;
youtube.;
.googleusercontent.com

CDN NAME

CDNetworks



DESCRIPTION

Founded in 2000, CDNetworks has offices in the US, Korea, China, Europe, and Japan. It is the only multinational Content Delivery Network provider with POPs in 6 continents. CDNetworks has a strong focus on emerging markets like Brazil, Russia and India.

NETWORK MAP

CDNetworks has a strong POP coverage on all continents, including 25 in China and 2 in Africa. Also, they have 10 POPs in Russia.

LOCATION

Seoul, Korea

FINGERPRINTS

*.panthercdn.com;
.cdngc.net; .gccdn.net;
.gccdn.cn*

CDN NAME

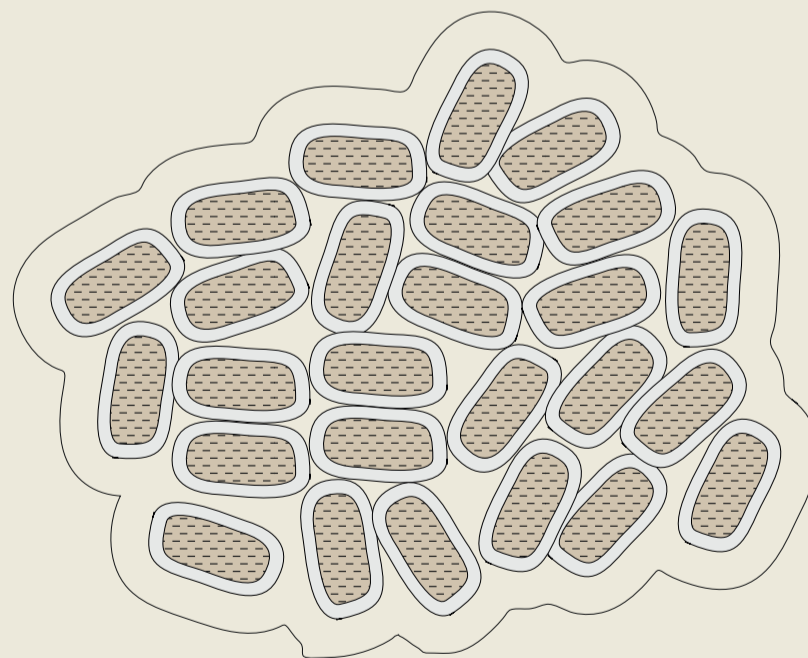
EdgeCast

DESCRIPTION

EdgeCast was founded in 2006 and operates out of Santa Monica, CA in the US. EdgeCast claims to be constantly rated amongst the fastest CDNs and participates in the Global Internet Speedup initiative. EdgeCast has 100+ employees and more than 3,000 customers including Wordpress.com, Break.com and Tumblr.

NETWORK MAP

EdgeCast identified a smaller number of strategic global locations near primary Internet Exchange Points, and built supercharged points of presence, called SuperPOPs. These SuperPOPs have massive amounts of computing and caching power and are directly connected to all the major backbone networks.

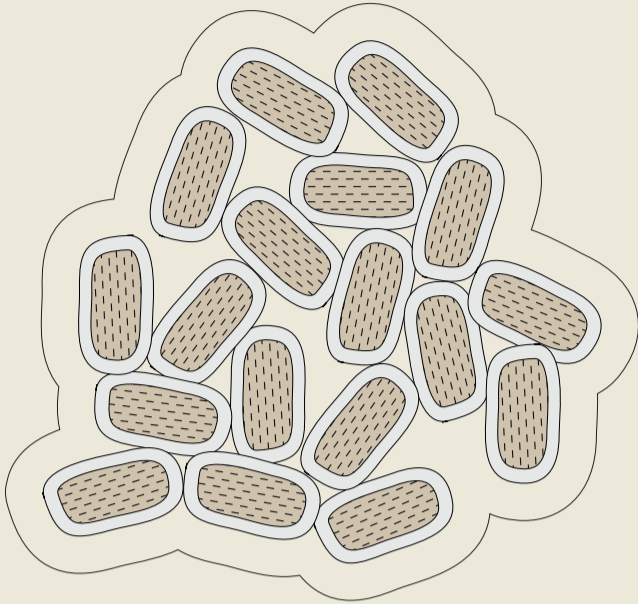


LOCATION

*Santa Monica,
California*

CDN NAME

SWIFT cdn



DESCRIPTION

SwiftServe is a new generation Content Delivery Network (CDN) designed specifically to meet the rich media demands.

NETWORK MAP

P2P?

LOCATION

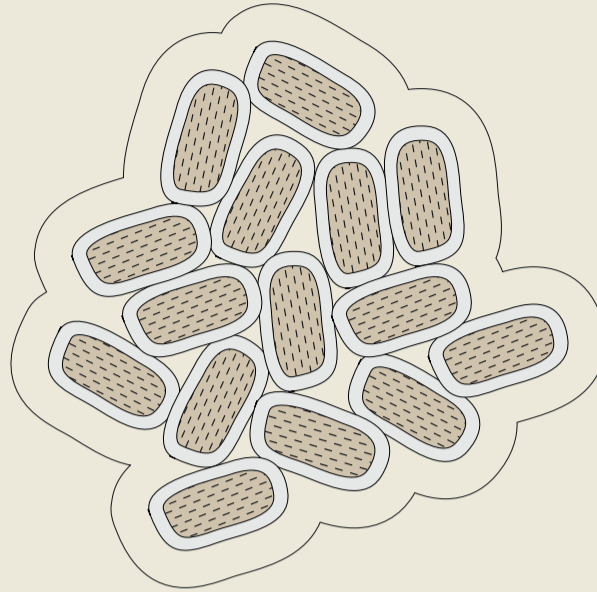
Singapore

FINGERPRINTS

*.akamai.net;
.akamaiedge.net*

CDN NAME

IXDNS.com

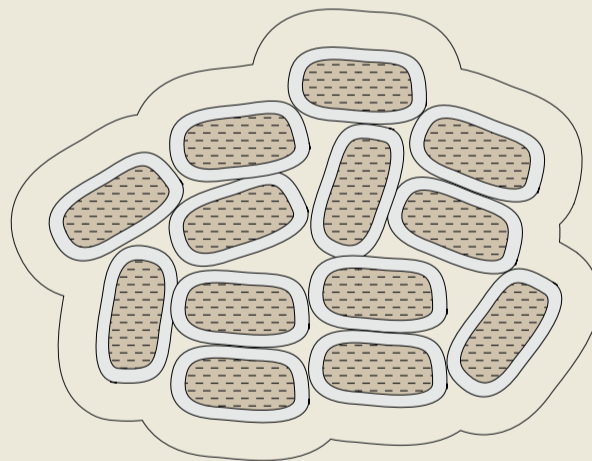


DESCRIPTION

This URL belongs to Tucows.com and points to ChinaNet servers.

CDN NAME

*Amazon
Cloudfront*



DESCRIPTION

Amazon CloudFront is a web service CDN for delivery of static and streaming content. It integrates well with other Amazon Web Services but can be used stand-alone too, via an extensive API.

NETWORK MAP

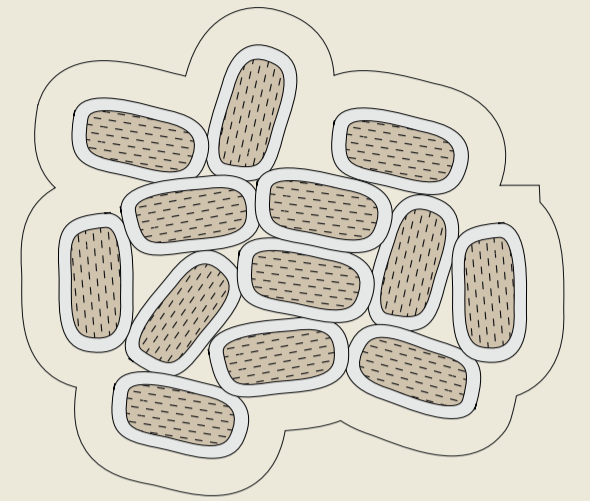
<http://aws.amazon.com/cloudfront/#details>

FINGERPRINTS

.cloudfront.net

CDN NAME

level3



DESCRIPTION

Level3 is a big company that has been providing a wide range of content delivery services for over 20 years. Level3 owns a Tier 1 backbone and has over 3.6 Tbps of peering capacity. The Level3 CDN services allow for token-based authentication, traffic prioritization, resource popularity tracking, full usage logs with auto-download tools, and more.

NETWORK MAP

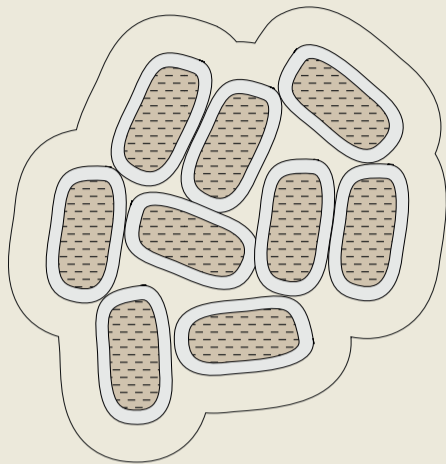
Level3 is the network. They own the backbone and have POPs all over the world.

LOCATION

Broomfield, Colorado

CDN NAME

ChinaCache

**DESCRIPTION**

ChinaCache was founded in Beijing in 1998 and since then has become a top player in the CDN market and the leader in China, due to the many POPs and local presense. ChinaCache has quite a few POPs outside China, incl. Korea, Indonesia, Japan, US and Europe, and offers a range of content and application delivery services.

NETWORK MAP

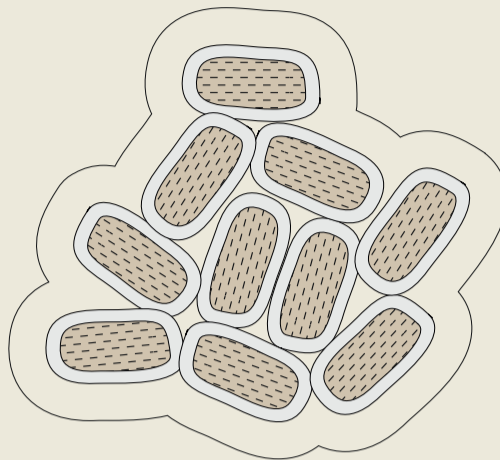
ChinaCache has 600 POPs in 146 cities across the globe, with a focus on mainland China: 11000 servers in 127 locations. The POPs in Seoul, Kuala Lumpur and Jakarta give ChinaCache an edge in specific parts of Asia. ChinaCache owns a Tier 1 network with 1 Tbps capacity and peers with all major China ISPs and Tier 1 global providers.

LOCATION

Beijing, China

CDN NAME

Taobao

**DESCRIPTION**

Taobao is owned by Alibaba Group. They use a private CDN.

LOCATION

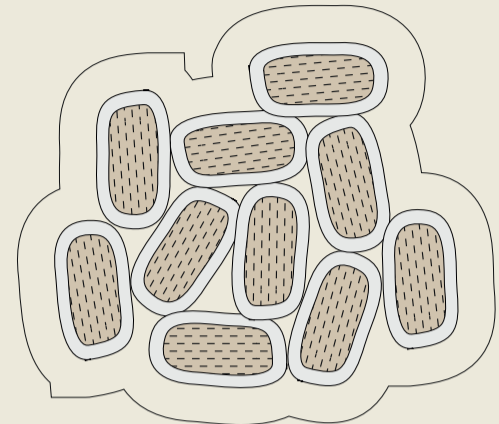
China

FINGERPRINTS

.gslb.taobao.com

CDN NAME

Windows Azure

**DESCRIPTION**

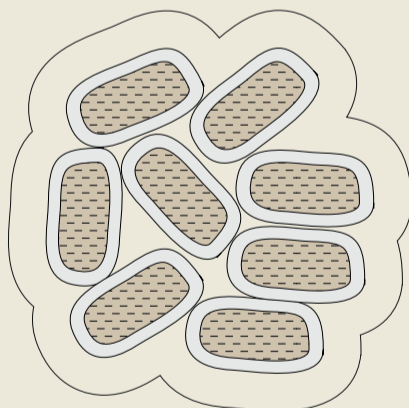
Windows Azure is a Microsoft cloud computing platform used to build, host and scale web applications through Microsoft data centers.

LOCATION

Redmond, Washington, U.S.

CDN NAME

Alimama

**DESCRIPTION**

Alimama is part of the Alibaba group.

LOCATION

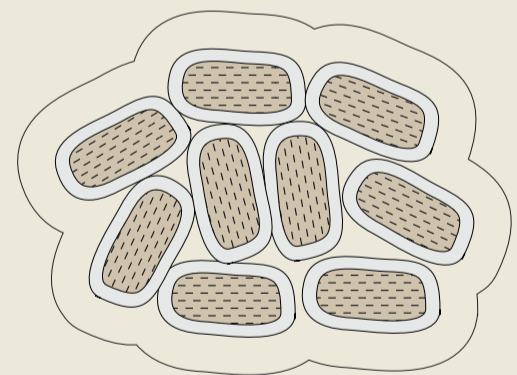
China

FINGERPRINTS

.gslb.tbcache.com

CDN NAME

Yahoo!

**DESCRIPTION**

Yahoo CDN provides common javascript and CSS files to websites and is also used to host their own files.

LOCATION

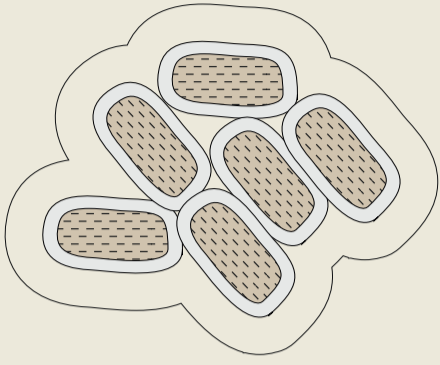
Sunnyvale, California, U.S.

FINGERPRINTS

.ay1.b.yahoo.com; .yimg.

CDN NAME

Limelight



DESCRIPTION

Limelight entered the CDN industry in 2001 and today has the second biggest infrastructure of all CDNs with 18000+ servers. The Limelight network relies on Adaptive Intelligence software and network innovations to make real-time decisions about each request for your content. Customers include Disney, BBC, EA and Adobe.

NETWORK MAP

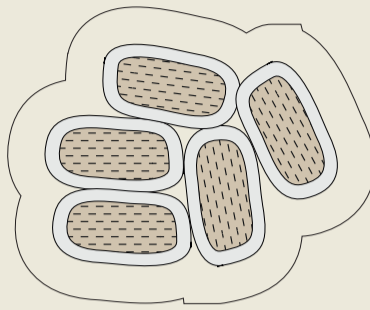
Limelight has POPs in all the main continents including South America, Africa, Asia, Australia and for China acceleration, they partner with Prime Networks for additional POPs in China. Limelight has 18000+ servers in over 80 locations and peers with over 900 ISPs globally.

LOCATION

Tempe, Arizona

CDN NAME

Reflected networks



NETWORK MAP

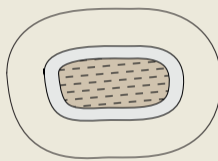
Over 25 localized nodes throughout the Americas, Europe, Africa and Asia.

LOCATION

Chicago, IL, USA

CDN NAME

MaxCdn



DESCRIPTION

MaxCDN is a low-threshold, static file serving CDN. You can sign up online, be up & running in 5 minutes and they offer a 30-day, money-back guarantee. With an aggressive \$39.95-for-your-first-TB offering, MaxCDN is appealing to those looking for a first experience with a CDN.

NETWORK MAP

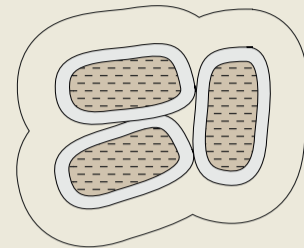
With 9 edge locations in play (and more coming soon), MaxCDN loads your files from original server to end users at their nearest geographic location. Each POP location boasts multiple 10 gigabit capacity. MaxCDN has over 500 peering partners worldwide and direct reach into over 90 countries .

LOCATION

Los Angeles, CA

CDN NAME

Cotendo



DESCRIPTION

Cotendo is a high-end CDN platform run by intelligent software, with innovative technology and services like Adaptive Image Compression, SEO Analytics and Site Assure (origin monitoring and automated failover system). Cotendo offers a full suite of acceleration services including Dynamic Site Acceleration and Advanced DNS. Cotendo was acquired by Akamai in 2012.

NETWORK MAP

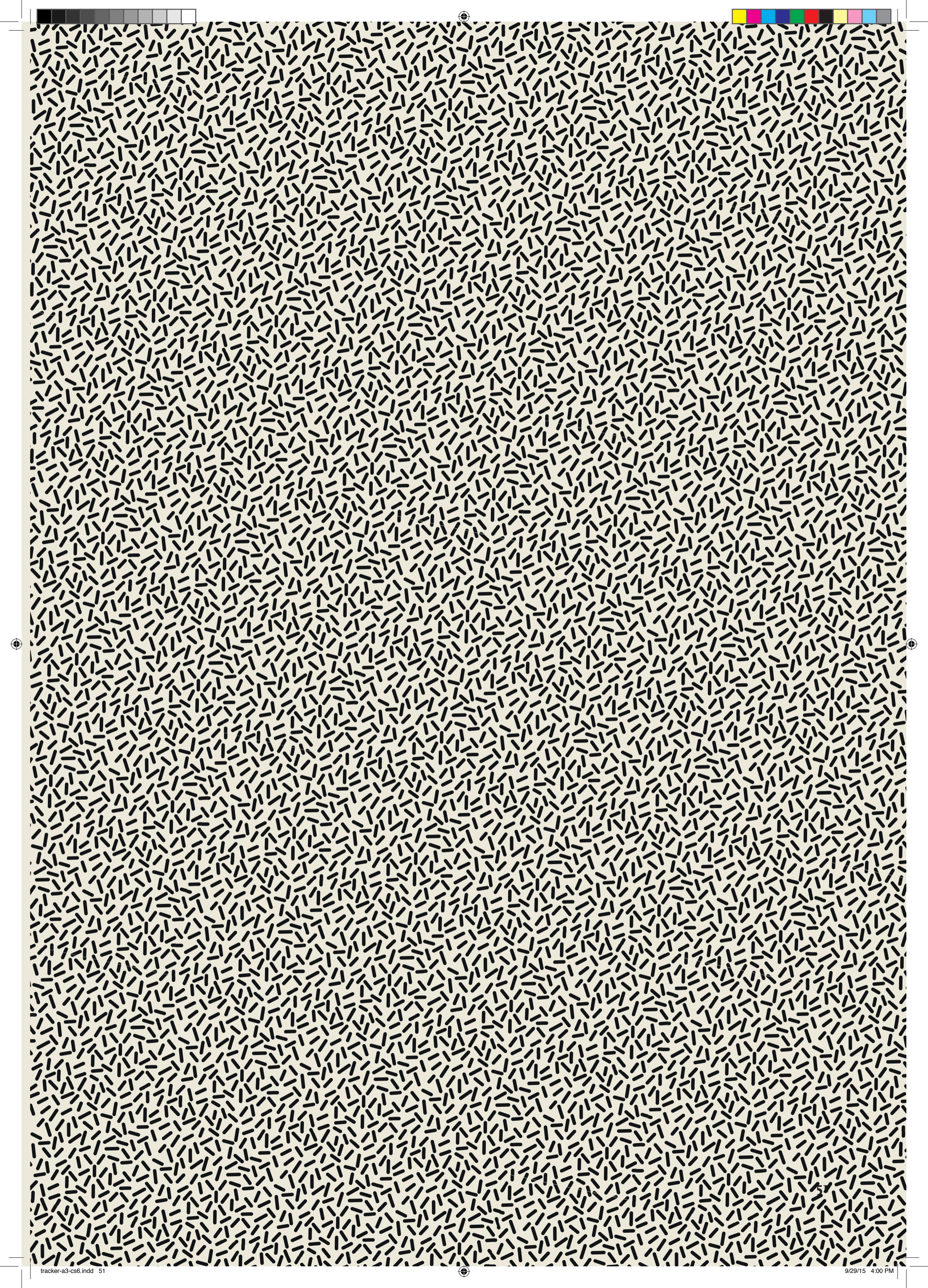
Cotendo has 35 POPs world-wide. They have a strong presence in Europe, South-America, India and China. POP locations include Spain, Italy, Sweden, Argentina, Brazil, Israel and Qatar.

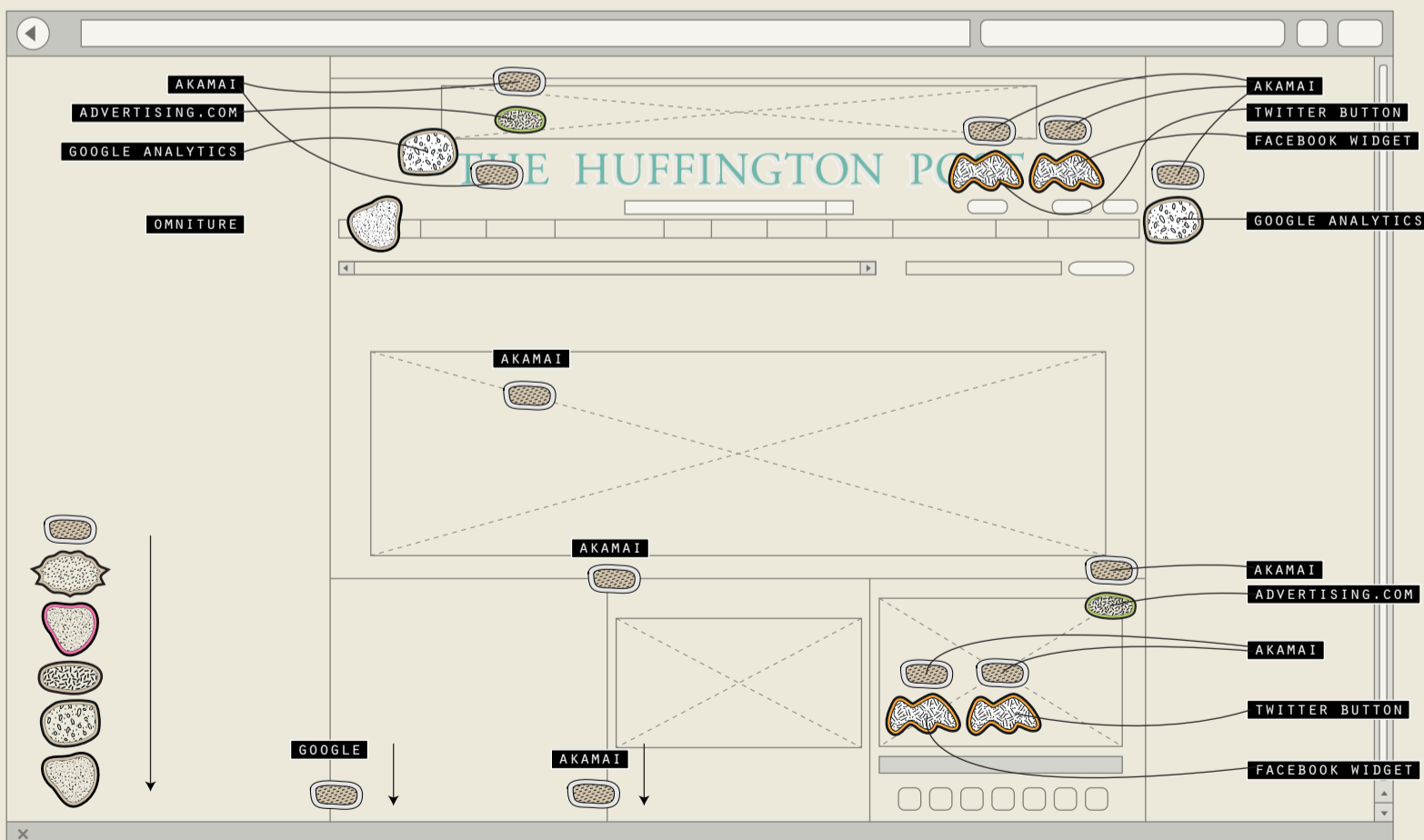
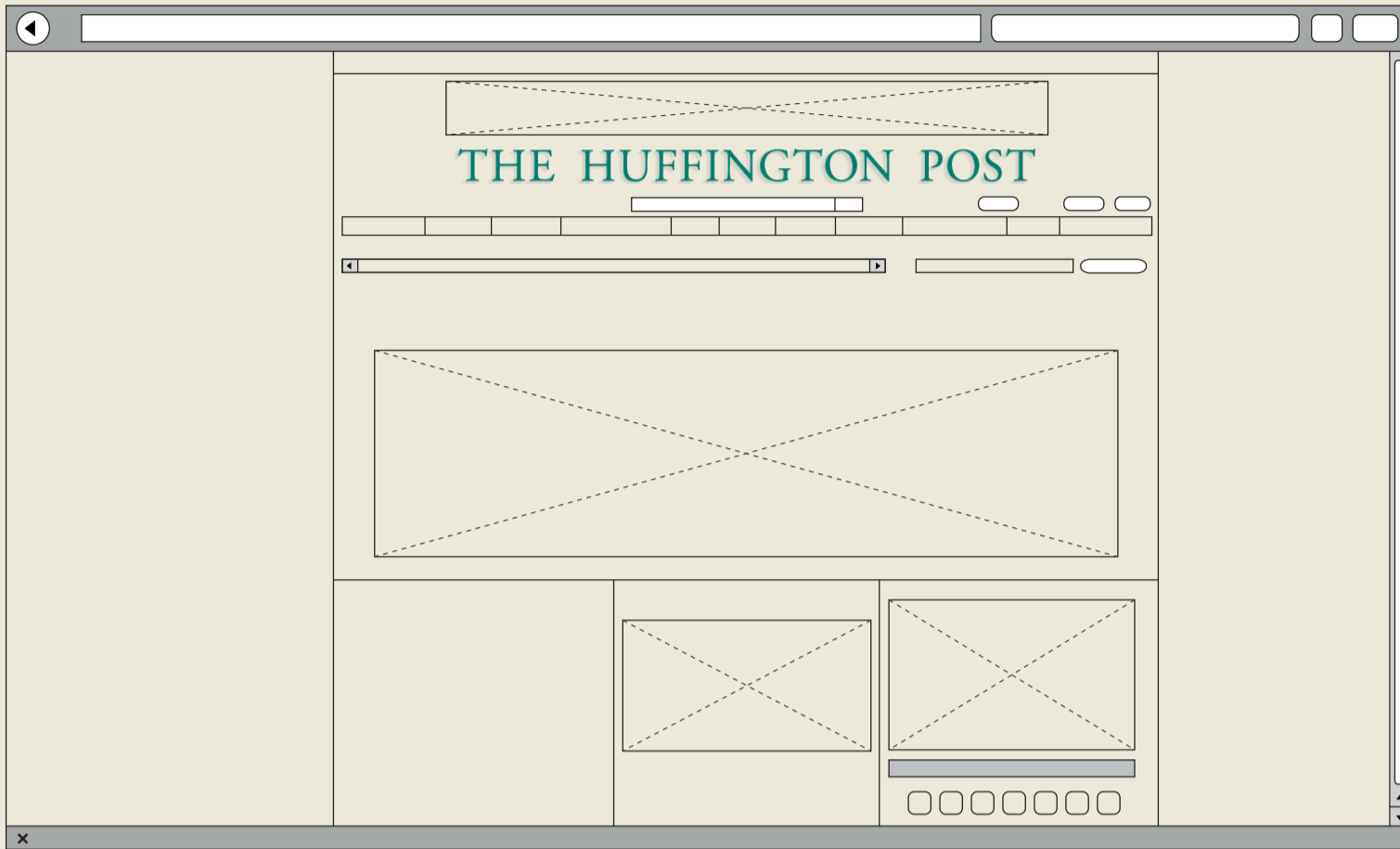
LOCATION

Sunnyvale, CA

FINGERPRINTS

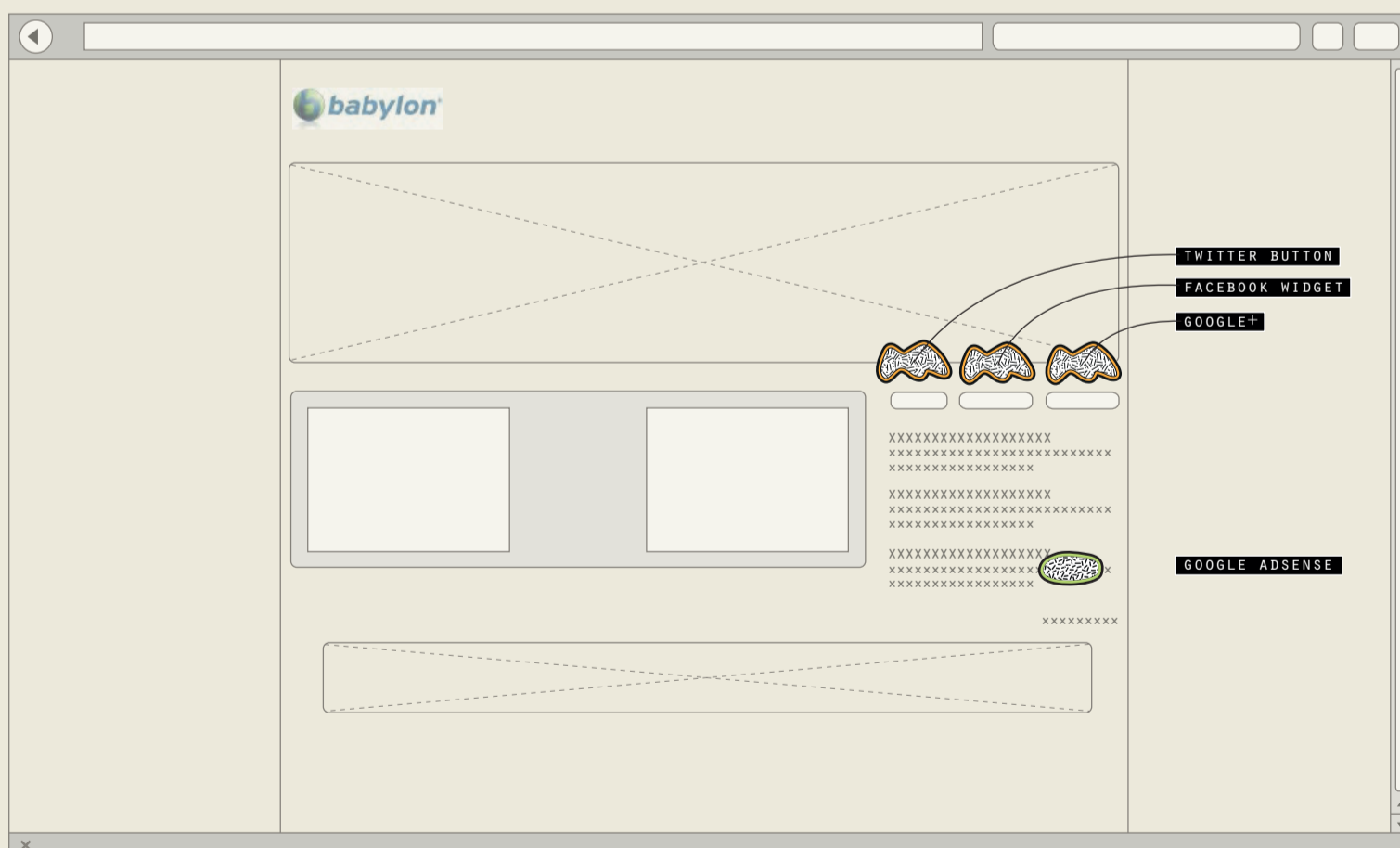
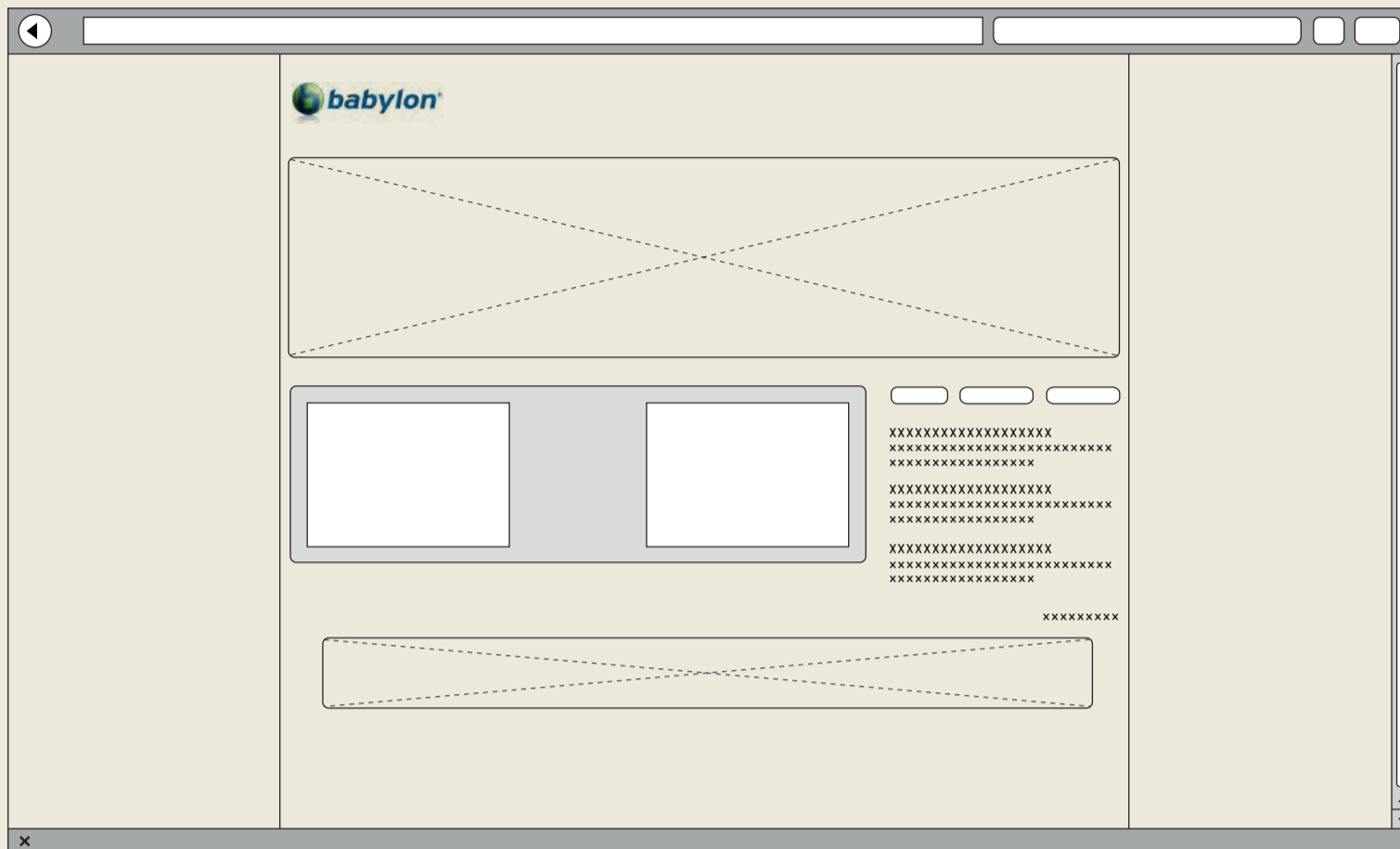
.cotcdn.net



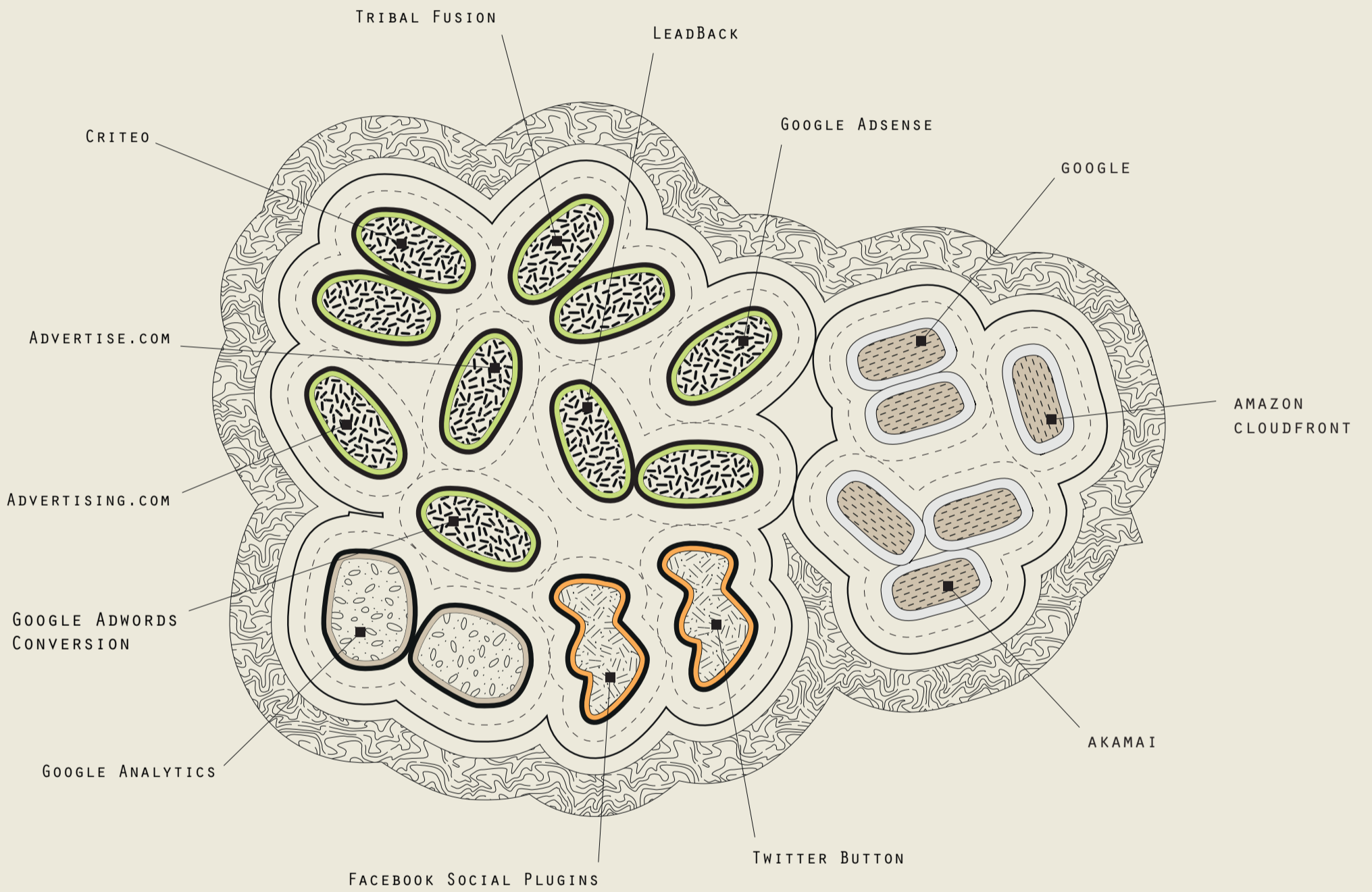


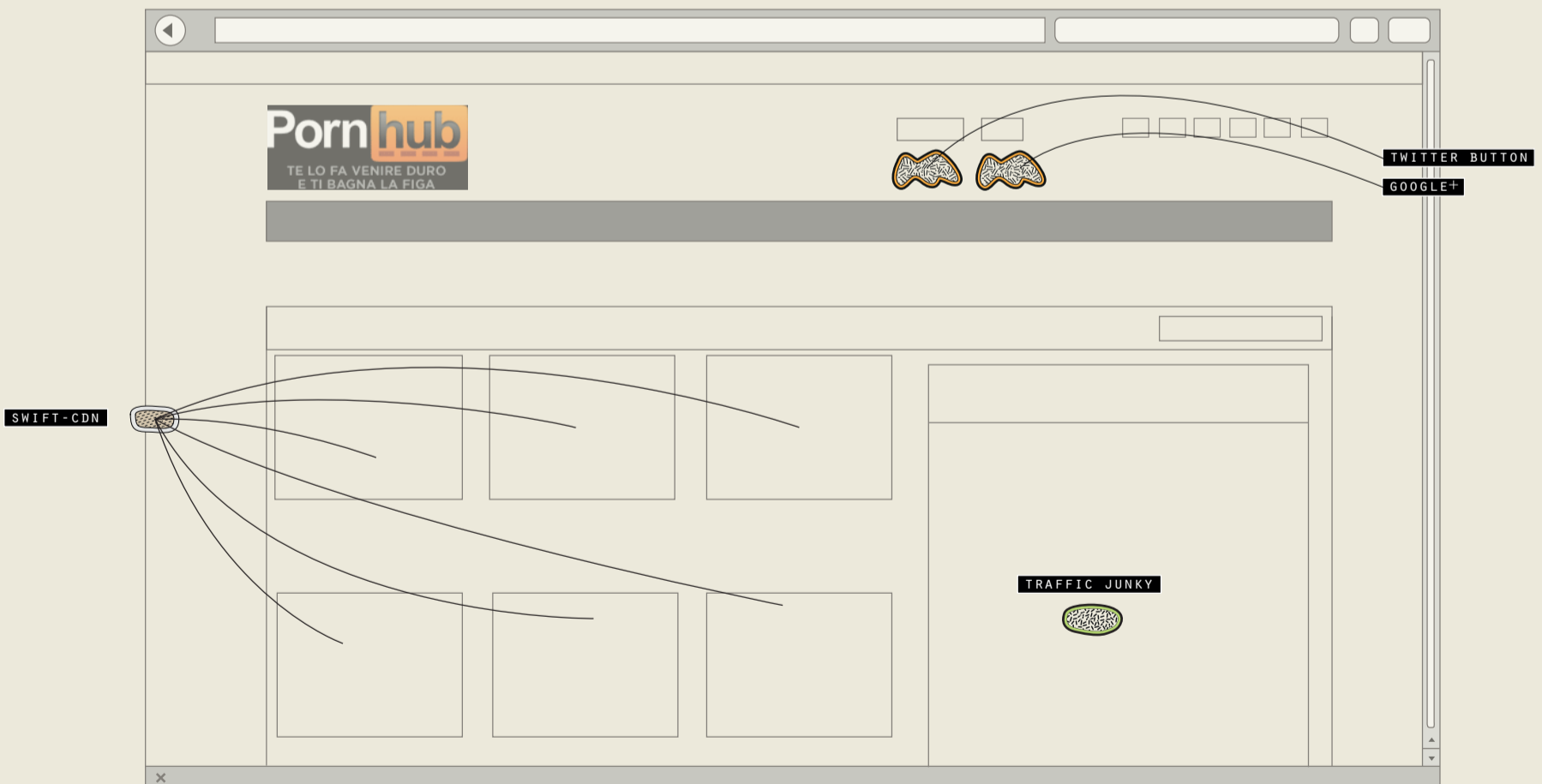
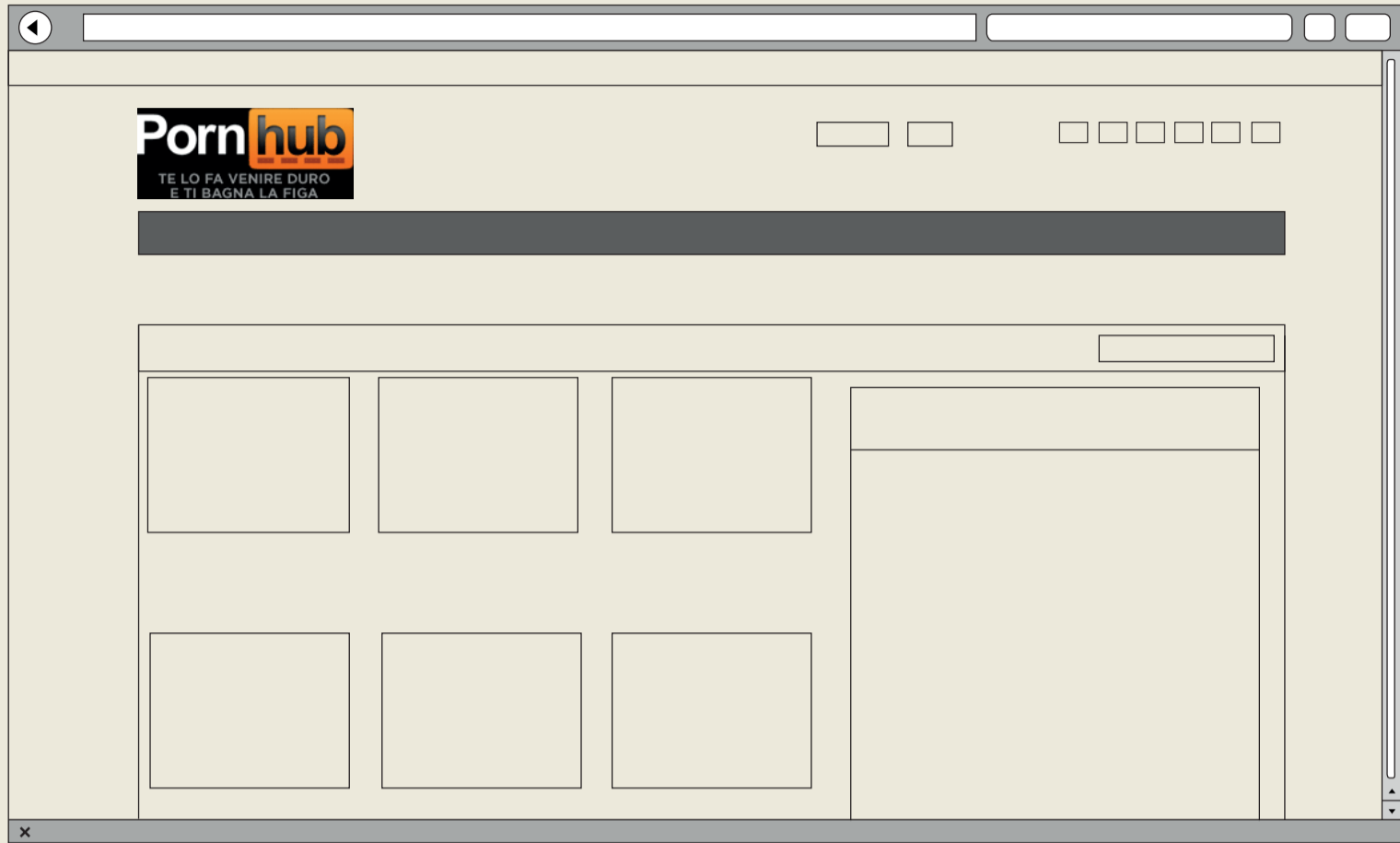
The Huffington Post





Babylon





Pornhub



FINDINGS

~ Distinguishing Between CDNs and Trackers

After identifying and classifying fingerprints of trackers and the cloud, we focused on the dataflows enabled by both, following the objective to map out the data ecologies embedded in such interfaces to the cloud and to explore the distinctions between trackers, widgets and CDNs. Special interest was taken in how such trackers and CDNs enable data in- and out- flows, which data is being collected and which is implemented back into websites, and which devices are involved in such processes, as documented in the glossaries.

However, we found that Ghostery also detects the fingerprints of tracking devices that not only collect data, but also feed back data to websites, most notably through advertising, through platform

related widgets which implement content from social media platforms and through few consumer survey devices. The majority of data inflows, however, are enabled through the content delivery networks, which can feed in both static media content (images, videos), but also dynamic one (CSS, scripts).

Bringing together both data collection and data in-flows, such trackers indicate that a too discrete delineation between content delivery networks and tracking devices might not hold up. First, as a series of trackers also feed in data into web-pages. Secondly, we found that a series of trackers, most notably those providing data-inflows co-occur with fingerprints to cloud services. Especially the case of www.huffingtonpost.com shows that the

same element on the page both contains fingerprints of the AOL advertising network and a fingerprint of Akamai. Further, social media buttons, like Facebook Like or Twitter button come with fingerprints of Akamai, indicating a relation between the widgets and content delivery networks. The unfolding ecology of data flows allows to draw conclusions about the relationship between the website and the cloud. Webpages, this study indicates, are not discrete entities, but function as frames or containers for data collection and data inflows. The website emerges as assemblage with multiple interfaces to the cloud, bringing together and being defined by a series of tracking and content delivery agents.



