ITEGA

INFORMATION TRUST EXCHANGE
WORKING PAPERS

ITEGA fix for advertising: Audience Profile Books

ver. 1.0 by Don Marti, 04-27-16 (LINKED TO LAUNCH PLAN)

base file: ITE fix for advertising: Audience Profile Books-04-27-16v1
https://docs.google.com/document/d/1APHGo8zrOFPKeFuoK_oURvk4GqJFTxVCfm-ZMEcds-A/edit

---

Goals of the Information Trust Exchange Governing Association include to help users regain control of their privacy and identity; help publishers to improve the relevance and value of advertising through deeper knowledge about their users' interests; and move toward a platform where a "fast pass for news" is possible without a dependence on tech platforms.

This document describes the function of the "Audience Profile Book" (APB) -- a feature of the ITEGA system design. A publisher compiles interest attributes of their users into an APB which contains no personal information about individuals. Rather it places individuals into interest cohorts. These cohorts are then shared with an ITEGA common service, a User Data Exchange. Like a railroad switchyard, anonymous interest cohorts from one publisher can be matched and aggregated with those of other publishers. Only the original publisher can identify a unique user within a cohort.

PROTYPE IMPLEMENTATION: The User Data Exchange (UDEX.org)

At October, 2017, a prototype implementation of the APB specification has been built by vendor Clickshare Service Corp. For a Q-and-A detailing the UDEX service please go to THIS LINK.

ITEGA fix for advertising: Audience Profile Books

By Don Marti
(dmarti@mozilla.com)
The author works for Mozilla Inc. as a participation strategist on "open source" projects. This document does not necessarily reflect the views of Mozilla.

It's time to fix advertising on the internet and the Information Trust Exchange Governing Association is organizing to do so.

Today, much online ad money goes to intermediaries, some goes to low-quality sites, and about a third of the money vanishes into fraud. Unlike in print, a high-quality news site is in direct competition with low quality sites and fraud sites for ad money.

The current web advertising model depends on tracking the same user across multiple sites (either "anonymously" or using PII). For reasons of privacy and resulting regulation, per-user targeting across the web is unsustainable and likely to become less available to advertisers in the future. **Here are four reasons why it is also bad for publishers:**

- **Targeted advertising fails to support news:** News sites are unable to pay the bills with web ad revenue. If current trends continue, news ends when print does. Publishers are seeking higher-value models. Finding them is not optional.

- **Competition to reach users:** Sites that produce content for a local audience or community of practice are now in direct competition with targeted ads, which can appear on low-value sites, to reach the same audience.

- **Fraud:** Today's adtech ecosystem makes fraud relatively easy and anti-fraud relatively difficult. Advertisers are effectively funding billions of dollars worth of artificial intelligence research, to create increasingly human-like ad-fraud bots. The presence of fraud in the system drives down the price of all web advertising, even ads on high-quality sites. Publishers, advertisers and copyright holders, not adtech intermediaries, bear the costs of fraud.

- **User choices:** Most users are willing to accept some advertising, but aversions to some targeting practices are widely held. Protection from strongly disliked practices, such as price discrimination and targeting by medical condition, is already a competitive advantage for the Apple Safari browser, and other browsers feel pressure to innovate.

Advertising right now is how a lot of news organizations hope to be able to pay the bills, as print advertising is going away. Advertising is certainly not there yet. The same content on the web

brings in roughly 10% of the value of that content in print. Yet advertising is something that people are relying on being able to do on the web, and make it work somehow.

Web advertising as it is practiced today is not consistent with the kind of user control over their own information that is important for a sustainable web. So information about users can effectively follow them from site to site resulting in unpredictable and unwanted privacy and security concerns.

Web advertising as it works today facilitates putting quality sites into direct competition with low-quality sites and with outright fraud. If you wanted to get an ad in front of people in Alameda, Calif., you could buy advertising in the *Alameda Sun* or you could go to some geo-targeted ad tech intermediary that claims to reach the 94501 zip code and buy advertising there.

We have the opportunity to build a new advertising system that:

- Works with and supports user privacy principles champion by two web-centric non-profits, the Mozilla Foundation and the Electronic Frontier Foundation.
- Doesn't break economic *signaling*, as targeted ads do
- Has immediate effects on current buzzword problems such as #adfraud and #fakenews
- Is small enough to implement in a reasonable amount of time.

Publisher strong points are domain expertise and reputation. A new system must use what publishers are good at, instead of trying to out-Facebook Facebook. Companies that practice problematic forms of user tracking tend to reduce the total value of web advertising much more than they actually profit.

From the user point of view:

- Users want free ad-supported sites
- Users express dislike for some common tracking practices (66% of adult Americans said they do not want marketers to "tailor advertisements to their interests")
- News and cultural works built at companies are an important part of the web's value, and we really don't want to see that type of site move off of the open web and into some kind of silo that is controlled by a single company.

Introducing the *Audience Profile Book*

Here we introduce the *Audience Profile Book. (APB)* This is a set of data -- an aggregation of individual user attributes related to demographics and interests. The data for an individual user is assigned a unique alias which makes the data anonymous. Only the entity assembling the APB can map the unique alphanumeric alias to an individual user's name or other PII. However, attributes of these anonymous (yet unique) users can be sorted into "interest cohorts" suitable

for content personalization or advertising targeting. These APBs may be used individually or aggregated with other APBs from other publishers, to creating enlarged pools of users with similar interest attributes. The ITE can then work with *member* ad networks and agencies to provide access to cohorts of like-interest users across parts or all of the ITE -- sorted by interests or geography.

Adverts get passed through to users via ITEGA, so the advertiser/agency doesn't know individual users, but is assured of delivery to real people who are contained within one or more interest cohorts. Logged reports of ad views come from ITEGA-member vendors as a trusted third party; these two aspects should reduce most fraud and make higher CPMs more defensible. **Also, since the process of collecting and collating user attributes into APB cohorts happens at the network level, it eliminates the need for third-party cookies on user machines -- eliminating privacy compromises and sluggish end-user browser behavior.** The publisher is responsible for seeing that ads contracted for delivery to a particular cohort (of which the publisher's APB has a piece) actually are viewed by the publisher's users within that cohort. **Thus only the publisher has to "know" the digital address of a unique user.**

The ITE knows individual users by a unique ID provided by the user's home-base publisher (this is necessary to be able to bill ads (and, futurely content views or subscriptions), but does not have any way of mapping that unique ID to a real person at the home-base publisher level. Again, only the publisher can do that.

The architecture is designed so that user "persona" data is managed by individual publishers, or a collaborating group of publishers (such as TrustX or Pangea) in a first-party relationship with their users/subscribers, but that data can be shared on a session basis -- with a user ID that persists for at least a week or longer with advertisers and agencies able to use it as part of real-time-bidding (RTB) processes. The user data would be stored temporarily at a ITEGA-member Data Aggregator/User Data Exchange *(potentially Mozilla in alpha)* or ITE-contracted data-service provider (DSP) to whom all advertisers, agencies, networks or exchanges would make calls for the data as they are engaged in RTB. In this way the network would not be relying on the reliability of thousands of publisher's servers for real-time user-data services; it would be dependent on one, or perhaps a few, providers of user identity data.

Key points:

1. Trustworthy, fraud-free, non-PII, current knowledge of user interests
2. Plays nicely with reputable privacy-protection services (DNT, etc.)
3. Creates new opportunity for publishers to add value to user relationship as privacy helper ("InfoValet")
4. Advertisers can have a single view of aggregated user cohorts (without PII), through that home-base publisher. Audiences are addressable but individuals are not.
5. This means everyone has a common contact source of user data -- Audience Profile Books (APBs) from which to figure their bids for ad space
6. The system does not preclude the continuation of the current cookie-matching world but over time it should be marginalized for its inefficiency and bad user experience.

7. Only the APBs are updated and combined by the Data Aggregator in real time, no individual user data PII or otherwise, is included. ("Big Brother is blind.")
8. Distributing the user data-management functions reduces the potential for identity theft on a massive, systematic scale and allows for the application of local laws and customs regarding privacy.

How it works for a user

1. User chooses a "Home Base" to manage their personal
    a) A publisher
    b) An identity service provider (such as RespectNetwork)
    c) Their browser (such as Firefox)
    d) An affinity group or tech company
2. Home Base must be a member of the ITEGA
3. Home Base gathers minimal info – email address – and creates profile that can be gradually filled out with user permission
4. Home Base assigns a permanent ID to user in ITE standard format (example FooBar123@ite.HomeBase.com)
5. ID is never shared to anyone other than the ITE network services
6. For federation authentication purposes (billing, service class) but not for advertising, ITE may create an encrypted session ID for a user (a "token ID") that can be shared with other publishers and which times out.

VISITING THIRD-PARTY SERVICES -- NON-ADVERTISING

1. When user visits other ITEGA-member services, those services check with an ITEGA-member authentication service to see if the user is an ITEGA-member user. A session ID is either created or retrieved.
2. This happens by looking for a non-cookie ITE unique identifier on the user's browser
3. If so, they get the anonymous ID either at the third-party site or by running a JavaScript function to get the ID stored on the browser. The third-party site can create additional "attributes" (interests, preferences) and contribute them to the ITE linked to the anonymous ID. This propagates the new attribute(s) to the user's HomeBase core profile. Sites might attempt to set ITE attributes for a non-ITE user, or change attributes that have been locked by the user or their Home Base, but these should fail harmlessly.
4. Complete temporary profile is at the ITE, but the core (master) profile is at the Home Base publisher (or identity-service provider).
5. User can modify user-generated or controlled attributes. User can choose to share attributes with sites (like "share my location", under user control, not like a cookie, behind the scenes)
6. Changes propagate (like DropBox-type cloud service) to the publisher's chosen ITE cloud service (possibly using Firefox Sync in alpha)
7. A few attributes are "locked" by contributor and can only be made private or deleted by user; cannot be changed (i.e., subscription terms or skill levels earned on

special-interest sites)

How it could work for advertisers – A simple example

The advertiser

A national retailer, "Outdoor Equipment Centers," has an internal customer "first-party" database. Many of their existing hiking, biking, jogging skiing and recreation-focused customers are readers of an extreme sports website or magazine. They want to reach more of those readers – who are not yet customers -- with a branding message so they will think of Outdoor Equipment Center when getting ready to buy gear.

Outdoor Equipment Centers offers attributes to its customers based on purchase history and attendance at special events. For example, the store might say: "Would you like to add to your user profile the fact that you have purchased a tent and backpack and that you attended a wilderness first-responder training session at the store?

User data sources

Specialty websites, magazines

The extreme sports website or magazine has received permission from their users/readers to collect information about what stories they read and what trips or special promotions they've signed up for. In addition, the users have given the site or magazine permission to share their interests on an *aggregated basis* in order to receive relevant offers, advertising or specialized content.

Some users of the sports web site have taken a survey in which they share their preferred activities and stores. The sports web site requests some of the attributes originally shared by OEC customers.

In addition, Outdoor Equipment Center's marketing department has determined that parents of high-school athletes are among their largest cohort of customers and wants to reach them.

Geographic-specific publishers

Daily newspapers who are part of the Information Trust Exchange have received permission from some of their users to record stories they read, and the paper has assembled an "Audience Profile Book" which includes a large number of readers of high-school sport stories. By cross-matching with its circulation and online user databases, it is able to include in a sports-gear purchasing intender cohort those users who are reading high-school sports pages and who appear likely to be active adults.

What happens next?

The Audience Profile Book

The three Audience Profile Books (APB) of the extreme sports website, the magazine and the daily newspaper are shared to and consolidated by an ITEGA-member Data Aggregator. This assembles likely sports-gear buyers into a consolidated cohort of users which meet the target requirements of Outdoor Equipment Centers. In this example OEC can use both . . .

- OEC-originated attributes shared in the sports site's reader survey
- high school sports interest attribute collected by the local newspaper site

. . . to plan an effective ad buy.

OEC can then deliver ads via ITEGA-member services to those sites – without giving OEC any user-specific identity information or any way of targeting individuals. **OSC, (or an agency working for it) can perform its own analysis on all of the APB cohorts -- and take a lot of time to do so -- because this is done before at the time of ad purchase, not at the time of ad serving.** This presents an alternative to real-time bidding.

Offering and assigning attributes as identifiers

The Audience Profile Book can include "attributes" assigned to users in a permissioned process, that may be automatically offered based on reading habits or selected directly. For example, a publisher could say: "Here is an identifier or some other piece of data that applies to you." The user could accept or reject that attribute. Ond other sites want to learn those identifiers if the user is willing to share them. (Users may also visit a self-serve site, or install a browser extension that lets them pick arbitrary attributes not signed by an authoritative issuer.)

A publisher could also present to a user a survey and the user could agree to respond. The publisher explicitly asks the user to share an attribute as part of the survey, and the user agrees. The site can choose when to request an attribute, and the user can say yes, no, or never share attributes with this site. Sort of like the way that "share your location" works today.

SEE: http://www.w3.org/TR/geolocation-API/#privacy_for_uas

Sites will have to balance user data collection with the risk of driving users away. The decision on when to request attributes is up to each site. Some sites may offer an incentive to assign or request an attribute. Testing of alternate offers and workflows is important. The publisher aggregates attributes from many users into its Audience Profile Book.

The advertiser relies on the ITE common attribute hub ("Audience Profile Store"), or obtains and parses multiple Audience Profile Books to assemble the desired audience, and buys advertising space on one or more publisher sites.

All of this can be implemented by SaaS (Software as a Service) providers working for the publisher and advertiser. All of these providers can be "Do Not Track" compliant.

Key points of information flow

- Sites can offer attributes to users, along with incentives for accepting.
- Sites can ask users to share attributes, and offer incentives for sharing.
- Sites can aggregate attributes about users into an Audience Profile Book.
- Advertisers and third parties can collect and parse Audience Profile Books.

User experience has to be flexible -- some sites will have to be very careful about asking users to accept attributes, and others will be able to make an easier workflow. A children's health site for parents will have a different UX from a web comics site.

Notes on system requirements

- Can start with a suggested list of attributes, but niche publishers will want weird user attributes. Free-form user info must be possible. (See an example of demonstration attributes).

- Every user attribute can have its own home base (*The Alameda Sun* can give me an attribute that says I probably live nearby, and the Linux Foundation can issue me an attribute about what I do for a living.) They may be storable encrypted, unreadable on server side, in a cloud service.

- A single publisher might have multiple section audiences in its Audience Profile Book (travel section readers, live music news readers). The APB will need to let publishers break out readers by section. (“arts and music” might be a section at one publication, but another might have want to have sections for “music” and “arts and literature”)

- Publishers should be able to choose how to split it up. An advertiser doesn’t get to reach an individual, or a cohort of readers, though, just placement on that section’s pages. A new reader will see the same ad on a given arts story that a frequent arts reader and survey taker does.

Publishers are positioned to make the ITE work -- by transforming advertising

The big internet technology platforms have lots of user data. Not just a bigger head start, but more experts and more money, and the power to get more data faster. You can't win a Big Data race with the people who invented Big Data and can do it more efficiently, with higher scale, than anybody.

What do newspapers have that the giant Internet companies don't? Reputation.

But reputation for honestly reporting the news does not scale in the same way that Big Data does. The winnable long-term plan is to transform the web advertising medium. Right now we have a data-driven game based on following users around, delivering less and less valuable advertising, and making it up on volume. That's a game where large Internet companies engaged in adtech/adfraud have the advantage.

In a reputation-based game, creepy Internet billionaires are at a disadvantage, and newspapers can win.

A complete breakfast -- championing tracking protection

Any next-generation system needs to be developed in combination with advancing Surveillance / Tracking Protection technology, in applying protection to more users, and with collecting data on the buying habits of protected users. No advertiser or agency will bargain with users or publishers for fairly shared information if they can just take it. Any marketing spend has to be sold internally through a complex chain of agencies, intermediaries, marketing, and top management at an advertiser company. Every decision has to be something that an individual marketer can sell to the next step in the chain. Current surveillance marketing is designed to be appealing to the current buying process. We have to cut off non-permissioned user data collection to get attention for data sharing.

High-value advertising must be an integrated part of a technical transition that also includes closing off the options for low-value advertising in the same medium. For example, publishers can warn users when they're vulnerable to third-party tracking, and encourage them to get protected. This will reduce the number of "bad" ads that those users see on any site.

	Client side	**Server side**
Make low-value ads harder	Layer 1: block connections to untrustworthy trackers Layer 2: don’t persist cookies and unsafe state Layer 3: clean up problematic state (layers)	Safe ad blocker warnings: don’t block privacy tools as ad blockers. Surveillance warnings: Inform users about alternatives so that they can configure in-browser tracking protection. Reverse tracking walls: offer bonus content to protected users.

Make high-value ads easier	Attribute assignment with user control Cross-site attribute sharing with user control	Information Trust Exchange attribute sharing Safe web analytics Future: federated paywalls

Publishers can't out-Facebook Facebook to offer creepier and more targeted ad placements. However, publishers can find common aspects of user-privacy-driven tracking protection improvements and publisher-business-driven tracking protection improvements and make them a priority.

Realistically, we can assume that advertisers and agencies will ignore the new ITEGA system until they see that it's a way to reach a significant audience that they can't reach in other ways.

Regulatory re-balancing of data risks

User identity theft and other security risks are negative externalities of data-collection practices that enable, among other things, ad targeting. Regulators in some jurisdictions are likely to try to shift these costs to the firms that collect the data.

Ad blocking seems to be mainly a balance between the hassle of blocking (dealing with broken sites and anti-adblock warnings) and the hassle of not blocking (slow page load times, annoying ads). However, ad blocking did not go mainstream until retargeting showed users how ads were attempting to reach them, and were not just something on the site. The economic signal of non-targeted advertising is a way to shift the ad blocking balance back toward advertising.

A new advertising system will let the web lose the directly user-targeted ad while continuing to provide the information that advertisers need in order to measure ad performance and continue supporting high-quality sites.

RELATED LINKS:

The Mozilla Manifesto
https://www.mozilla.org/en-US/about/manifesto/

List of protection layers
http://blog.aloodo.org/posts/protection-layers/

Tracking warnings for publishers
http://zgp.org/targeted-advertising-considered-harmful/#solution-tracking-protection-for-publishers

Ad blocking: why now?
https://digitalcontentnext.org/blog/2015/07/06/ad-blocking-why-now/

A new acronym for quality ads on the web
http://blog.aloodo.org/posts/new-acronym/

Does the Audience Profile Book strategy work for brand advertisers?

Considerations:

- Standard attributes and cohort-assignment for audience segments
- A near-guarantee that, by design, ads will reach humans, not bots
- Verified delivery and auditing by Information Trust Exchange
- Aggregation of APBs across topicals and geographics
- Ability to research APBs in non-real time to craft buys

- NO ability to identify individual targetted users (just where they are and what they are interested in or intend)
- Confirmation of delivery frequency by cohort/segment (not individual user)