Information Trust Exchange Governing Association
http://www.itega.org

QUESTIONS AND ANSWERS:

itega-udex-prototype-q-and-a-CONFIDENTIAL-12-28-17 (long version)

Prototyping a special-purpose DMP for for sharing anonymous user profiles

---

The Information Trust Governing Association (ITEGA) is collaborating with two private vendors, Taxonometrics Inc. and Clickshare Service Corp., to operate a prototype anonymous User Data Exchange ("UDEX") for privacy-by-design serving of relevant advertising and content. In this Q-and-A format, Clickshare's Project Director, Alexander Wise, explains the service. Clickshare has asked that this preliminary document be treated confidentially and not forwarded or reposted beyond an initial recipient.

DIAGRAM:
http://newshare.com/itega-launch/itega-architecture-DIAGRAM-v3-10-27-17.pdf
ALSO SEE APPENDIX A: ID lexicon

Q: What is the User Data Exchange (UDEX) project?

A: The UDEX project demonstrates a specialized data-management platform we call an anonymous user-data exchange, or UDEX, which operates between between identity service providers (IdSPs) and advertising stakeholders such as Demand Side Platforms (DSPs) in order to keep customer details anonymized, but still let ad stakeholders deliver effectively targeted advertising to those end-users. In our prototyping case, the IdSPs are online news publishers, which have traditionally made most of their revenue off selling advertisements. Online, though, the ads need to be targeted to the right eyeballs, so the publishers essentially release some of the end-users' data along with the ad space on their pages. The idea of the UDEX is to let the publishers manage the Personal Identifying Information (PII) of their users, while still allowing the ad companies to use certain information relevant to their interest in selling ads. The publishers still get ad revenue, and the ad companies are still able to target ads to individuals, but the end-users get to remain anonymous; their personal details are not divulged.

Q: As you discussed this with ITEGA, what emerged as the consensus for the need you are addressing?

A: We are starting from the notion that ad servers exist in the internet marketplace, and the companies that run them are fed by behind-the-scenes tracking of individuals' internet activity. Bits and pieces of information about you are picked up based on what you are doing — your IP address, maybe your email address, picking up a bit about your physical address, what you purchased, what you've googled, and what you appear interested in based on sites you've visited. Advertising data platforms will build profiles on individuals and share these data among themselves in order to discover correlations, and then can build more complete profiles. These profiles can get pretty detailed. They use this information to target ads to you about stuff you're hopefully interested in. That's a major piece of it.

I believe I should have some inherent control over my information. We recognized that there is this balancing act between three sides of a triangle:

- ADS CARRY THE FREIGHT: As long as the internet infrastructure costs money to operate, there is a need to monetize online services, unless the service providers are going to actively charge me for them. End-users, myself included, appreciate when there are services that are free — I remember when micropayments were the solution to all the problems, and people were going to charge me a fraction of a cent for a web page, and that was going to pay for it. That hasn't happened yet. Advertising became how content is paid for. And now it is paying for less and less as advertising migrates to Google and Facebook because they have both more customer data and millions more end-users than online newspapers. That means the ads are worth less, so more and more of them need to be shown, and they need to be more effectively targeted.

- SOME TARGETING IS GOOD: From a marketer's perspective, and indeed from my perspective as a customer, targeting ads is in many ways a good thing. I would much rather be presented ads that are of potential interest to me than having stuff randomly thrown at me by the ad companies in the hope that something sticks. Knowing something about me for that purpose is a good thing, but there should be bounds to this.

- PRIVACY/DATA CONTROL: The third element, of course, is privacy. In the advertising economy, I, as the product whose attention is being sold, would like to have some control over what information the advertising networks are collecting about me. What details do they have? I'd like to be able to have some say over that.

The crux of it is being able to control my notion of identity online, but we can see the obvious need for the revenue the publishers get from the ad companies, and the utility of targeting ads effectively on the part the ad companies.

Q: This sounds a bit contradictory. How can you balance data privacy while still giving data to the advertising companies?

A: The trick is in creating identities that can be used to isolate information so that the ad companies can't correlate data that isn't already attached to that identity. The idea is to stop this business of sharing information between networks to get a more complete profile of individual end-users and replace it with only the information the ad companies require to effectively target their ads, and allow end-users to effectively choose what information they want to share about themselves.

We want to create a certain level of anonymity enforceable by the public using GDPR and the authority of ITEGA's membership, governance and rule-making. Our solution is for that is to create the UDEX Project.

Q: So what have you built to achieve a level of anonymity without breaking useful advertising or the ability to "personalize" content?

What we had to build was software that keeps track of identity profiles anonymously through the generation and use of unique IDs while also noting the interests and demographic data of those profiles so that we had general information for the ad companies. The IDs, generated by the UDEX, are only accessible to certain entities, and they come in three forms: Primary IDs, Durable IDs, and Temporary IDs. *(For definitions, see Appendix A)* The home IdSP (the entity that has the authentication relationship with the subscriber) will send a request to the UDEX to create the Primary IDs of each of their subscribers. This ID stays with the IdSP. Durable IDs are generated by the UDEX by request from the home IdSP to be given out to other trusted entities, like another publication. They correlate with the Primary ID at the home publication. Temporary IDs are also generated at the request of the IdSP or other trusted entities, and that is what's passed on to the ad companies. Neither the ad companies nor other IdSPs apart from the home publisher will have a method for knowing to which individual profile the ID correlates. The subscribing end-user and their home-base publisher can grant a UDEX the limited right to use the information about a customer but that right can be withdrawn, and ITEGA's business rules will forbid the UDEX from revealing that information to anyone else. The UDEX would basically stand between the ad companies and the IdSP, and would tell the ad companies if there are any end-users who fit a certain ad targeting profile without giving up any personal identifying information. The ad companies get the information they need, the publishers get to sell ad space, and the end-users get to remain anonymous. Everybody wins.

Q: So you are saying that there is a service, sanctioned by ITEGA, which gets permission to extract the demographics and information interests from individual profiles to pass on to ad networks? That service – the anonymous User Data Exchange – assigns a unique key to that interest data, but only the user's home

service provider has the key to find the information needed to connect it to a named individual.

A: That's correct. There is some technical uncertainty still as to whether the UDEX should ever need to associate a named individual with an anonymous record of interests and demographic data. We want to make that technically impossible, but if that is itself impossible to achieve in a practical system, then ITEGA will, through membership rules enforceable on a member UDEX, prohibit the UDEX from data-sharing in any way that could permit re-identification of end-user PII with anonymous interest and demographic attributes.

The idea is to delegate some of the responsibility for managing individual customer data down the chain to the customer themselves, or to their identity service provider (IdSP), in this case, the news publisher. So what I built, in an effort to support this, is a repository where the information will be stored.

On the one side, the various web content services – multimedia publishers – will be collecting and -- if permitted by the end-user via GDPR-compliant queries -- sharing and using individuals' data to make decisions about serving content and the like based on that individual's preferences.

On the other side, a mechanism exists for managing the cloud of aliases that serve as anonymity-granting keys, which can be shared with advertising networks who do not need to know who you are as an individual and, indeed, don't want the legal or business responsibility of touching any personally identifying information.

Q: Now how does the individual express their preferences for the use of their profile information, whether anonymously stored or not?

A: I can answer your question in two dimensions. The first is from the individual's' point of view. In this early demonstration we're operating, it is currently very simple as a starting point. As a public user, I can express through my identity service provider what information I'm willing to share, and what information I am not willing to share. Currently, that is binary: share or not. Going forward, we will need a richer interpretation of what that is. In the future, there could be levels of sharing: I'm willing to share to everybody, to some people, and to nobody. Right now, we don't have that but ultimately, we want to build something powerful enough to express a layered customer intent that is unobtrusive enough people will actually use it.

The other dimension is there is a notion of visibility of tracked information to the network of data users, such as advertisers or other content providers. An individual might allow their email address to be stored within a User Data Exchange server. If it is marked public by the customer, it means the other stakeholders who have access to your internal information can see it. But ad networks will not have access to an email address, because it is unique to you and that would tend to make you identifiable as an individual. However, the User Data Exchange server, though not giving up your personally identifiable information, would still impart enough general information that tells the ad companies that specific ads would be appropriate and lucrative to show.

Q: I'm getting a picture of a service which stores demographic and interest information about individuals, but which by rule defends the idea that data consumers, like advertisers, who are not authorized by the targeted customer will only receive data about those end-users that can't be linked to them individually by name or uniquely identify attributes.

A: That's correct. There is a collection of data that lives within the service, and there is a mechanism by which ad-ecosystem stakeholders can ask about these data. And what we think will be the use case is that the UDEX will group segments of anonymous individuals by their common demographic or interest fields.

You can say things like: "Tell me about the segments of anonymous individual profiles in the system that includes zip code 01002." That's a simple one. What we've built can give you an answer to that question and give you back a scoring of the number of anonymous profiles of people in that zip code. It can also segment anonymous profiles of individuals across multiple attributes. You could ask for people age 18-24 living in the 01002 zipcode and are interested in outdoor recreation, for example.

Using these segments, an advertiser can match anonymous profiles to relevant advertising. However, in order to preserve the anonymity of the profiles, these profiles are represented by temporary aliases.

Q: You just mentioned the idea of temporary aliases. What's that about?

A: If the UDEX gives out the same ID to an anonymous profile indefinitely, or the same ID to multiple advertising sources, over time it could be possible to associate activities and make algorithmic guesses about who the underlying person is. Or, to begin to associate interests that the end user does not wish associated with their profile.

So, we address this problem in two ways, first by giving different IDs to each DSP or other ad-placement service, and second by making the IDs that are given to advertising sources only valid long enough to allow reasonable ad-serving "frequency capping." This allows them to follow a customer so that they can show a sequence of ads over several pages, but means that at the end of the identifier's lifetime any information that the advertising sources are disconnected.

So the only entity that can always link the customer's profile with a real person is that person's home identity service provider. Again, we postulate that can be a news organization, but in theory it could be a bank or an Internet Service Provider, an affinity group or a new business class of internet Identity Service Providers (IdSP).

Q: What is contained in these identity lookup aliases?

A: An alias has four parts. The first part is the content provider's ID, the second part is the ad server's ID, then there is a unique, *temporary* ID that differentiates that customer from every other customer but cannot be linked back to the actual person and so therefore obfuscates any relationship to personally identifiable information. Finally, there is an expiration date after which the ID will no longer be valid. So if a given content provider is called FOO, then all of the

identity look-up aliases issued by them or for them are going to have FOO in them somewhere. The UDEX issues a *different* unique temporary alias to each advertising service so that ad networks are unable to cross-match the aliases they received from ITEGA services to discern a unique individual. We call this a third-generation alias (the one given to advertising services) because it is generated by the UDEX from a second-generation alias, which was in turn generated from a persistent ID possessed only by the customer's home identity service provider (IdSP).

Q: Thanks. Now back to this question of scoring. What's that about?

A: The ad network can ask in its queries for a scoring of anonymous profiles across attributes. They can either ask for an assessment of the quality of a segmenting scheme within the population of end-users in a profile server, or they can submit an alias they received from a publisher website via UDEX for a customer who is in the middle of requesting a web page, and ask, "Does this anonymous profile score high on any of the attributes my advertisers are looking for?"

Q: Wow, that kind of cool. How many attributes can an ad server ask for in real time?

A: We have been thinking about three, but the code we've written will support 60 or more. A segment could be sports car purchasers, and that segment may be made up of an interest of automobiles and an income in the $70,000 plus range. So that would be two attributes that describe that segment. But of course, there could be a whole collection of attributes that together defines a sports car purchaser. We're just not sure at a production scale whether upwards of 60 attributes would be feasible in terms of latency and server wait times. The more attributes, the more data is involved, and that means the communication among servers to determine which ad is best to serve takes longer. It could affect page load times, which may be annoying to the page visitor. Our expectation is that if this is the *only* personalization and matching that has to take place – as opposed to the current environment involving potentially many dozens of ad server calls on just one page – then things will be noticeably faster than the current environment. However, in this prototype, no effort has been expended yet on speed optimization.

Q: I think I understand they idea of aliases and how they relate to profiles of end-user attributes. Walk me through an example of how things work.

A: Let's describe a hypothetical that we've built for. Assume a series of steps as follows:

1. Customer FooBar agrees in a GDPR-compliant way to allow their local newspaper or other news service to gradually assemble a profile of them that contains some demographic and content interest information, and perhaps other identity segments related to buying habits as well. Let's call those elements "profile attributes".

2. The newspaper works with an ITEGA-member technology provider to send those attributes to an ITEGA-member anonymizing User Data Exchange.

3. The UDEX stores the customer's profile attributes. It also assigns a new semi-permanent alias to those attributes. The alias will be maintained so long as the end user's identity service provider (and IdSP, in this case, the news organization) authorizes.

4. The UDEX sends back to the IdSP a set of *temporary* aliases to the same customer profile. Each alias is different but links to the same data. One alias is assigned to each of the ad-ecosystem stakeholders the publisher -IdSP deals with. This could be a fairly large number of aliases, each one unique.

5. At some later point, say within two weeks, the customer goes to a web page where an ITEGA-member ad service wishes to serve an ad. The customer's identity service provider (news organization) pulls the alias appropriate to the customer's profile and the requesting ad network, and sends it to that ad network.

6. The advertising-ecosystem user can then do one of two things with the alias:

    a) In the first instance, it can check its own database and see if it has previously received and stored that alias from an ITEGA anonymizing User Data Exchange. If it has, it has likely associated it with some anonymous profile attributes. It checks if it wants to serve an advertisement to an individual with those attributes.

    b) In the second instance, the ad network can submit the alias to one or more User Data Exchanges, asking for a score of the anonymous individual's level of interest in a collection of the demographic or interest segments sought by its advertisers. As noted above, in the prototype demonstration this is limited to a request for three attributes.

Q: So what you have explained is a system in which advertisers can find out in real time whether a person requesting a particular web page is an attractive target to be served a particular ad. But the profile driving that decision is controlled by the end user and their trusted home identity service provider (IdSP), in this demonstration, a news organization. And the system's intention is to make it impossible for an individual's anonymized attributes to be shared for more than about two weeks (only to facilitate "frequency capping"), and even then only by a single ad network.

A: That's all correct.

Q: You mentioned that this project was also involved in a network accessible to end-users via a single sign-on feature. Could you please elaborate on this?

A: In our prototype demonstration, let's say you want to look at some protected content at the Rutland Herald, a daily newspaper in Vermont, which you know is in your local newspaper network. But it's not your home base paper, not where you bought your

subscription. This doesn't matter. For a moment, let's say that the Worcester Sun is your home base news organization, but you're not logged in yet. The network will know because you have a subscription and a cookie was previously dropped on your computer indicating that you're a network subscriber. So you click on that protected article at Rutland, you're redirected to the Worcester Sun to log in, which you do. Then you get redirected back to the Rutland content. In the prototype, before you see the page you requested at Rutland, you get served an interstitial ad, and that ad will be served based upon any interest segment your profile belongs to that the advertiser has requested through the process we've just been talking about. Note that the advertiser never knows who you are or even whose customer you are. They just know that you have expressed a level of interest in their target market via clicking on something or reading a page to know that showing you an ad relevant to those interests is probably not a wasted impression. And, they also know that you are a *real person* because an ITEGA-member service provider has you as a registered customer in a long-term trust relationship, which may well include a billing relationship.

So the key points here are that it is the Worcester Sun that has the Primary ID for its customer. A Durable ID is created for the Rutland Herald, and the Rutland Herald requests the UDEX create a Temporary ID for use with an advertiser. So advertisers are always at least two steps removed from being able to identify a specific profile.

Q: Suppose someone has an account at more than one ITEGA service provider – say a newspaper and an affinity group. Would the UDEX technology allow those to be associated with each other, or is that a policy question to be decided?

A: It's definitely a policy question to be decided with due respect for GDPR. Right now the UDEX stores a bunch of Personal Identifying Information for pragmatic purposes. Since a policy goal of ITEGA might be to allow collection of knowledge about individuals across multiple domains, in a privacy-by-design-and-by-rule manner, you can at present in the prototype match PII up so you can recognize the customer that comes from Worcester and also, say, from YouStream(™) as being the same ITEGA network user. The advantage to storing PII on the UDEX means it would have richer information about the end-user's interests than a single home base. However, this also means that the UDEX is holding PII, and sharing it beyond the original provider would likely require an explicit authorization for specific use to comply with GDPR. This is a policy question for the ITEGA governance process.

Q: How complete is the demonstration service?

A: At Jan. 1, 2018, we are showing the process of creating profiles, assigning Temporary IDs to profiles which have been stripped of PII, and exchanging those IDs among and between DSPs, other advertising stakeholders, publishers, IdSPs and the UDEX. We can execute the decision by an ad server to send an ad into a particular position on a content page. But we're not working with real ads, so in a live demonstration we'll just see in the

ad space a message which says words to the effect of: “An advertisement meeting the segment of (whatever the segment name is) has just been placed in this position.”

Q: Talk a bit about how UDEX handles the processing of user “interests” received from either publishers or identity service providers when they have permission from their end user to store topical interests or other profile attributes.

A: In the simplest case, we might treat interests as Booleans – either on or off. Either I'm interested in hiking or I'm not interested in hiking. So when a data consumer (such as an ITEGA member DSP or other advertising stakeholder) sends queries, it is in the form of “Is this anonymous profile interested in hiking?” The UDEX responds that they either are or are not interested in hiking, or rather, the profile either is or is not in the hiking segment. So a data consumer could send a query: “I’m interested in people interested in hiking, but not sports cars,” and the UDEX would respond accordingly.

Initially, we planned to respond to such on-or-off queries. But there are emerging news personalization services, such as YourStream(™), that store intensity of interest, or attributes, in a given topical area based on real-time interactions. This will add a score on how interested a profile might be in certain things. If I offer you 10 articles about hiking and you take one of them, but the rest of them you ignore, then I can say maybe you aren't so interested in hiking, maybe it was something other than hiking that triggered you to click on this link. Maybe it isn't hiking you are interested in, maybe it is the Teton Mountains in Wyoming. So YourStream(™) assigns fractional scores to say "this person is a little bit interested in hiking". That's useful because over time, as the interaction evolves, you can get a clearer notion of a person's level of interest. Or, you might conclude over time that you are wrong. So you have to keep track of that information.

The DSP or advertising stakeholder sends their ad segments to the UDEX some time previous to the actual serving, so when the time comes for an ad to be shown, the UDEX has to make a threshold judgment about when to consider a profile as being in or out of a segment. At that point, when a profile is scored into a segment, somebody who has a higher score will fit better than somebody who does not. If someone scores 90% interested in cars and 20% interested in hiking, and you score them as to the car segment vs. the hiking segment, they are going to score higher on the car, obviously. At that point, the question an ad server might ask is "how well does this customer fit these segments?" and the UDEX sends a score between zero and 100 for each one.

The ad server can also ask things like: “Tell me about the segments of anonymous individual profiles in the system that includes zip code 01002." That's a simple one. What we’ve built can give you an answer to that question and give you back a scoring of the number of anonymous profiles of people in that zip code. It can also segment anonymous profiles of individuals across multiple attributes. For example, you could ask for people aged 18-24 living in the 01002 zipcode who also are interested in outdoor

recreation. All that helps the ad server's decision about which ad to serve. An advertiser can match anonymous profiles to relevant advertising.

Q: What are the privacy considerations at play here?

A: The whole thing is set up to hide the information about individual, identifiable users from advertisers so that single-user tracking is as close to impossible as possible, at least from data collected from within the ITEGA member community. We try and obscure that by instead of being able to ask about that raw data on a specific user, an advertiser can only ask about the segments and ask how well does this user -- who is about to be served an ad -- how does that user fit a given segment or segments.

Now, in our prototype UDEX we are able to store PII, but ITEGA rules will forbid it to pass that anywhere else except back to the user's home base. And its possible to imagine a scenario where even that doesn't need to reach the UDEX server. These are policy decision to be made within ITEGA's emerging governance structure.

Q: When you talk about topical interest and demographic attributes of an anonymous user ID, what is the scope here?

A: While there is no technical limit on how many attributes may be shared, there *is* a limit on how many attributes can be used when defining an interest segment -- a query is made up of one or more attributes, and segment can't have more than 60 or so attributes that it reasons about. If a segment I'm looking for might be sports car purchasers and so a sports-car purchaser segment would be made up of a sports car purchaser has an interest of automobiles, they have an income in the $70,000 plus range, that would be two. There could be some number -- a collection of attributes -- that together defines a sports-car purchaser.

Q: What about ad-placement context?

One piece we haven't talked about is the need for the ad network to know the context for the ad it is considering placing. Also, there can be a distinction between the entity serving a web page on which ads would need to be inserted, and the customer's home base identity service provider. Let me explain both.

First, we include a mechanism for tagging page content types and sending the content type as part of the message that goes to an advertising network carrying the temporary alias to an anonymous profile. We have homework to do to determine whether there are uniform taxonomies of both advertising content types and host web-page content types that we would recommend ITEGA adopt as standards. We assume there are.

Second, the system is set up so that if an ITEGA-member website accepting an ad on to its pages is NOT the customer's home base, it can find out who the customer's home base is and get them to send the appropriate alias information to the ad network. That's a complication in how this all works, but it is something Clickshare has been doing for many years in other contexts.

Q: How do you think ad platforms will handle this system?

A: The ad platforms have an inventory, a collection of ads that can be presented to the customer, and each ad has associated with it a set of profile segments that it satisfies based on the ITEGA-sanctioned ad-type and user-attribute-cohort taxonomies. You would expect these taxonomies to based on identities in a form similar to what YourStream(™) has specified. You might be able to be as granular as typing keywords in to ad network inventory. Ad server segments are written in terms of a mixture of interests and demographics. The profile demographic taxonomy we're prototyping is based on Clickshare's taxonomy of customer demographics. Clickshare has a rich data model that underlies most of its end-users. So the demographic model comes from Clickshare's experience and the identity-interests model comes from YourStream(™)'s work.

Q: So what actually happens at the moment an end user clicks on a web page and that page has a place where an ad network has the opportunity to show an ad?

A: The ad request is tagged with an anonymous identifier for that end user that was earlier assigned by UDEX.

Let's say the double-anonymous ID received by the ad server is "FooBar123". The ad server (Profile Usage Agent in our lingo) submits a request to the UDEX which says, in effect, "I have user FooBar123 who I'm considering serving an ad to right now. I want to know how well they fit into the sports-car purchaser segment, and the Japanese restaurant-goer segment." The UDEX has a scoring API for passing an anonymous user ID and a couple of bits describing interest segments that the advertising agent wants to reach, so, the Profile Usage Agent will get back a ranking on how well the user matches the sports-car purchaser and Japanese restaurant-patron segments.

CONTEXTUAL ADVERTISING OVERRIDE

That could be the end of it. These queries could have nothing to do with the content on the page the user is looking it. But ideally in a quality-ad environment of the sort we want to encourage, they will. So we have also built a mechanism to pass what we call "overrides" – attributes you want to assert a particular value for. Here's an example: When you are looking at a page about hiking, you can assert that independently of what the UDEX has on that user. You might do that because you know the context is that the end user is on a page about hiking and you have an advertising who sells or makes hiking boots.

Q: From a page-load point of view, what might the impact on the user experience be from this approach?

A: In the prototype, UDEX checks each of these for matches individually. So the more segment queries you process, the more expensive it all gets in terms of machine processing. Is there a limit on this? At some point it starts to effect page-load time for the end user. Our expectation, however, is that if this is the *only* personalization and matching which has to take place – as opposed to the current environment involving potentially dozens of ad-server calls on just one page – then things will be noticeably faster than the current environment. However, in this prototype, no effort has been expended yet on speed optimization.

Q: You mentioned above that the ad server needs to receive sort of a "token" from the publisher server which includes a double-anonymous ID for the end-user about to see an ad. How does that token get created and provided?

A: In this prototype, we are working with my company, Clickshare Service Corp., and its CS Net service, which has been operational for two decades now. It is a network service – which requires some server software – that creates and transfers tokens within headers or URL appends. These tokens, depending upon the application can include a key to access anonymous user attributes involve interests, access rights or payment authorization. We're using CS Net because we can at this stage without charge. Our assumption is that ITEGA will want to specify and encourage development of open methods for such data transfer and server code and of course Clickshare will support that.

Q: Can you explain how a user is authenticated and how their data is anonymized and then passed to the data-demographic aggregation and anonymizing server (UDEX server)?

A: In our prototype demonstration, if you click on some protected content at, say, the Rutland [Vt.] Herald, it throws up a login dialogue. You click to login and it bounces you to your "home base" identity service, which we'll call the Worcester Sun. You login there and then you get redirected back to the Rutland content. In the prototype, before you see the page you requested at Rutland, you get served an interstitial ad, and that ad will be served based upon any interest segment you belong to which the advertiser has requested through the process we've just been talking about. Note that the advertiser never knows who you are or even who's user you are. They just know that you have express a level of interest in their target market – or are reading a page about a topic in their target market – to know that showing you an ad relevant to those interests is probably not a wasted impression. And, they also know that you are a *real person* because an ITEGA-member service provider has you as a registered user in a long-term trust relationship which may well include a billing relationship.

Q: So, to go over this once more, how did the ad server get information about user FooBar123's interests?

When the Worcester Sun gets the log-in request it knows you are waiting for a page at the Rutland Herald. When the Worcester Sun approves your login, it also sends via what will be an open API, a set of attribute-field values and creates a UDEX ID linking to Worcester's permanent ID for you (probably your email address), and then creates a secondary ID for the Rutland Herald so these accounts cannot be correlated outside of of the UDEX. The Rutland Herald then creates a THIRD ID for it to send on to the ad server. The ad server now has a third-generation token for the user FooBar123 who is waiting to see the Rutland page advertisement. The ad server creates a query to the UDEX server; the UDEX server responds with answers about which segments FooBar123 fits within. The ad server then decides which ad to serve to FooBar123 – or takes a pass on the opportunity and the Rutland server then moves to the next ad server in its waterfall, where the process repeats.

So the key points here are that it is the Worcester Sun that has a durable, persistent ID for its end user. It creates a second-generation anonymous ID for the Rutland Herald, and the Rutland

Herald creates a third-generation anonymous ID for use with an advertiser. So advertisers are always at least two steps removed from being able to identify a specific user.

Q: In your example, Worcester has a durable ID that links to a real user. But the other parties have IDs that are opaque as to the specific user. But advertisers need to be able to run campaigns in which they know that a given – unique, but anonymous – user has seen X-number of impressions of an ad over time. What about that?

Rutland gives a temporary profile service ID to the ad server and those IDs have a lifetime during which they are stable – currently we are figuring on 14 days, but that is configurable balancing the interests of the advertising industry with the desire to minimize the potential for illicit matching to occur outside ITEGA which would pierce user privacy. Every time Rutland sends an ID for a particular customer to a particular ad server, it will send the same ID during that 14 day period. Eventually it will expire. So the ad server has an ID that it can use to uniquely identify the same customer during a given campaign, but the ID becomes stale and useless for off-network, illicit matching purposes after a set period.

Worcester asks the UDEX for an ID for its customer – call it Sandy -- and the UDEX is capable of correlating those up, based on the same email address. I know the person you are asking about is Sandy, I'm going to give you Worcester Sun Sandy. Then when this whole conversation occurs, Worcester Sun gets asked by the UDEX saying I need an ID to give Rutland. So Worcester says I have Worcester Sandy, please give me Worcester-Rutland Sandy and the UDEX looks inside and says I know Worcester Sandy is actually UDEX Sandy so I will create another alias, Worcester-Rutland Sandy and pass it to Rutland. then Rutland then says I need an ID for Worcester-Rutland Sandy to give to the ad server. So that's Worcester Rutland Ad Sandy.

Q: Suppose a given end user has an account at more than one ITEGA service provider – say a newspaper and an affinity group. Would the UDEX technology allow those to be associated with each other, or is that a policy question to be decided?

A: It's definitely a policy question to be decided. Right now the UDEX stores a bunch of Personal Identifying Information for pragmatic purposes. Since a policy goal of ITEGA might be to allow collection of knowledge about end users across multiple domains, in a privacy-by-design-and-by-rule manner, you can at present in the prototype match PII up so you can recognize the customer that comes from Worcester and also, say, from YourStream(TM) as being the same ITEGA network user. If as a matter of policy it is decided we can't do that, then we can't do that. And the schema that I worked from had demographic information. Depending on how this evolves over time you might put less PII in there. The CS network ID has this idea of a distinguished home and if you're home is always your starting point then conceptually you could in fact silo this stuff up in such a way that only your home would have your PII and everything else would have second- or third-generation anonymous, temporary IDs.

What we are trying to continue to support is a knowledge model that advertisers think they need without also running down the train wreck which end-users are afraid of -- which if what your profile is made up of is your interests, your interests don't change when you are at the NYT vs. the Berkshire Eagle. The UDEX at the moment guards all of that so you get the unified view of

the customer hopefully without the privacy risk that the current approach has so that if you actually have this siloing you lose that and you would have to somehow convince advertisers that is good enough for them to decide to play your way.

Q: So to reiterate, there are three distinct types of IDs within the prototype user-data exchange you have created.

A: That's correct. There are what we just call customer ids, which belong to the a data owner, a customer owners, the home server, the Identity Service Provider (IdSP). In our example above, that is the Worcester Sun. Secondary IDs are handed off from these primaries and they end up with an identity showing what primary they came from and who they were delegated to and because we know where they came from, one of the properties of them, and we can call those durable IDs within the system, they have an indefinite lifetime but the home server can cancel them. Then there are temporary IDs which are the third class of ones which are the only ones that advertisers or their agents ever see and which are void after a fixed amount of time – which we have arbitrarily set in the prototype at 14 days to allow for a reasonable advertising campaign period.

-- CONTINUED ON NEXT PAGE --

APPENDIX A

TYPES OF UDEX IDs

Three types of "aliases" are used within the UDEX system. As a result, only the UDEX can establish relations, and then only for purposes permitted by ITEGA exchange rules.

- Primary ID -- Created only by an Identity Service Provider (IdSP) who has been chosen by an end-user. Links to PII managed by the IdSP on behalf of the user and according to privacy choices made by the user and enforced by ITEGA membership.

- Durable (secondary) ID – Which end-user's IdSP can generate to pass on to a User Data Exchange (UDEX), or to an ITEGA-member remote home, such as YourStream(TM). It might contain access to PII such as an email address. ID's are just IDs. They are indexes to information in the UDEX. In isolation you can't do anything with them.

- Temporary (tertiary) ID -- which can be created from Durable IDs by the UDEX and passed to ITEGA-member Profile Usage Agents (such as ad networks). They are only valid for a limited time, currently, they expire in two weeks. This ID can only be linked by UDEX to a Durable ID, and contains no PII or interest attributes. UDEX issues a different Temporary ID to each Profile Usage Agent, so no off-network identity matching is possible under ITEGA rules or practice.