THE INFORMATION TRUST EXCHANGE

http://www.infotrust.org

**Trust, identity, personalization,
content and user sharing for the news industry**

Functional specification for user-data sharing

(draft v3.0 -- 01-30-16-BD)

This WORKING document proposes functional specifications for exchange of permissioned user data to support customized service of digital content – advertisements, stories or other services. Organizations undertaking one or more proof-of-concept demonstrations or prototyping should refer to the "Request for Proposals: Information Trust Exchange Services" as well as the broader Service Design Specifications. The Consumer Use Case is also instructive. The definition of "Personally Identifiable Information" in this specification is intentionally broad to inspire discussion about what the advertising-industry needs and does not need. An attempt has been made to make terms used consistent with those in the document "Testing a Customer Profile Network" (v0.031 12-29-15) from Clickshare Service Corp.

■ *Bill Densmore*

A. OBJECTIVES

1. Support storage and updating by users of personal attributes and preferences at a most-trusted service provider, such as a publisher or other identity-service provider.

2. Support publisher/agency/advertiser real-time requests for current unique user profile data.

3. Centrally log events for value exchange and third-party verification

4. Avoid use, storage or aggregation of personally identifiable information.

5. Deprecate the use of "cookies"

B. DEFINITIONS

The following terms are used as defined below. Terms in *italics* are taken from and used in the document, "Testing a Customer Profile Network". (citation above).

1. "Advertising Service" – A advertiser, marketer, agency, advertising network or advertising exchange, or any other entity involved in presenting Commercial Messages to Users. If doing so

across the ITE, they must be Members. An Advertising Service may operate as a Profile User Agent.

2. "Attributes" – A set of characteristics stored in a database which are unique to a particular user's account, such as zipcode, service class, subscription start/stop, demographic or preference information.

3. "Commercial Messages" – Advertisements, sponsored content and other information objects created by Advertisers for the purpose of influencing Users, and often presented to Users in collaboration with Publishers.

4. "Cookie" – A persistent HTML cookie storied on an end-user's device.

5. *"Customer Profile Network" – A common platform for the sharing of individual User profile information for the purpose of improving the User experience and managing Privacy.*

6. *"Data Aggregator" – a central point of a Customer Profile Network. An independent entity, such as the Information Trust Exchange, which media sites and other Data Collectors and Profile Usage Agents join.*

7. *"Data Collector" – A media site or other entity that collects information and acts as an identity manager for individual Users.*

8. "Event" – A single, unique completion of an HTTP request for a resource, such as a page, photo, video, advertisement or other content item.

9. "Exchange" – One or more services governed by the Information Trust Governing Association which provides such services as authentication, authorization, transfer of tokens and keys and logging to members of the ITGA. The Exchange functions as a "Data Aggregator."

10. "Key" – An alphaneumeric set of values, typically encrypted, which one Member can provided to another Member for purposes of linking to a unique set of stored Attributes.

11. "Identity Service Provider" – An entity which helps a User to manage their identity attributes in order to obtain ITE network services. An identity service provider acts as a Presenter. It may offer its services "within the cloud" and the User's identity attributes may be stored "in the cloud" on the User's device or within the User's web browser application. The method of storage may affect their trustworthiness.

12. "ITE Data Transfer Protocol" – A standard method for securely exchanging data among and between services within the ITE network, including content servers, authentication servers, identity servers, logging servers and user-management servers.

13. "ITE Mod" – An ITE-compliant modification module to an Apache or other web server to allow it to perform ITE required actions before serving responses to an HTTP "GET" request.

14. "ITE Servers" – Network services operated in behalf of the Information Trust Exchange Governing Association including, but not necessarily limited to:

    a. authenticate.infotrust.org – Stores encrypted session UserIDs and manages obtaining UserIDs from Presenters.

    b. Identity.infotrust.org – Stores user-authorized identity Attributes keyed to UserIDs for use by Presenters and Providers according to ITE rules.

c. Advertising.infotrust.org – Stores identity Attributes keyed to encrypted UserIDs and specifically authorized by users for Advertising messaging; records time-out and are deleted on a schedule set by the end User in collaboration with their Presenter. Also stores advertising-specific log reports for off-line aggregation.

d. Logging.infotrust.org – Stores log reports of HTTP events supplied by Providers for offline aggregation and analysis, including encrypted UserIDs which may only be decrypted by infotrust.org servers or by the server of the user's Presenter.

15. "Log" – A database which contains records of individual accesses to resources and which may include such things as date, time, URL, UserID, Resource Value Class, Markup Ratio or other standard attributes of an Event.

16. "Member" – A business entity which has been admitted into ITGA membership and which may be a publisher, advertiser, agency, author, producer, identity-service provider, technical service provider or other exchange participant or operator.

17. "Personal Identifying Information" (PII)– Any representation of information that permits the identity of an individual to whom the information applies to be reasonably inferred by either direct or indirect means. Further, PII is defined as information: (i) that directly identifies an individual (e.g., name, address, social security number or other identifying number or code, telephone number, email address, etc.) or (ii) by which an agency intends to identify specific individuals in conjunction with other data elements, i.e., indirect identification. (These data elements may include a combination of gender, race, birth date, geographic indicator, and other descriptors). Additionally, information permitting the physical or online contacting of a specific individual is the same as personally identifiable information. This information can be maintained in either paper, electronic or other media. (source: U.S. Dept. of Labor). Also, information which can be used to distinguish or trace an individual's identity, such as their name, social security number, biometric records, etc. alone, or when combined with other personal or identifying information which is linked or linkable to a specific individual, such as date and place of birth, mother's maiden name. Also any **information** that can be used to distinguish one person from another and can be used for de-anonymizing anonymous data can be considered **PII**. (Also see, Wikipedia entry).

18. "Presenter" (or "Home Base") – An entity which manages user accounts for the purpose of providing services, which might include subscriptions, resource-access control or custom services tailored to a unique user, such as personalized information, including advertisements. *Also known as a "Data Collector."* A Presenter may be a conventional publisher, broadcaster, digital-information provider or an just an Identity Service Provider.

19. *"Profile User Agent" – An advertiser or other entity that wishes to use profile information acquired from a Data Collector, which in turn manages it in behalf of one or more Users.*

20. "Provider" (or Publisher) – A provider of information services, which also provides positions within their web services for the placement of advertisements by third parties for viewing by a user. *Also functions as a Profile Usage Agent.*

21. "Session" – A continuous period of time during which a unique Token is assigned to a User by the Exchange in order to facilitate customization of services by Members. It is anticipated that ITGA rules will forbid cross-Session linking of Tokens and Attributes for the purpose storing a permanent unique identification of an individual, so as to constitute PII.

22. "Token" – An alphaneumeric set of values, which may or may not be encrypted, which is provided by one member to another by a common transfer Protocol. A token may contain a Key.

23. "Transfer Protocols" – One or more set of technical and business rules for the communicating of Tokens, Keys, Log reports, Attributes and other required requests, responses and services among and between Members of the ITGA.

24. "User" – An individual consumer end user

25. Information Trust Governing Association (ITGA) – A non-profit, public-benefit collaboration of Members (see below) organized to facility trust, identity, privacy and information commerce.

C. OPERATING EXAMPLE

It one intention of the ITGA (in formation) to facilitate the delivery of Commercial Messages to Users in order to facilitate and sustain one business model for delivering news and other information in the public interest.). Personally Identifiable Information (PII)[1] should be opaque to all parts of the system except the user's identity service provider ("home base.") Therefore, a design goal is to facilitates commercial-messaging exchange without necessity to share any personal identifying information (PII). The delivery methods should, at a minimum priority, be respectful of User privacy, choice and time. The following example is intended to meet that intention, and priority.

1. A User using HTTP protocol on an open network transmits a URL to a web or mobile service provider (Provider or Presenter) which invokes a GET request to receive a resource in HTML (or variants).

2. If the GET request is to a page which is part of the ITE (content or ads are offered by providers who want to personalize or sell services by obtaining User Attributes), a modification in the web server withholds momentarily serving the page.

3. The server's ITE Mod looks at the submitted URL to see if it has a query-string appended which includes an encrypted ITE Member/User ID.

    a. If it does, the server sends a request to identity.infotrust.org, obtains user-authorized identity attributes for the user, makes whatever personalization, authorization, advertising or content sales adjustments required, and then serves the page, passing along the encrypted ITE Member/User ID within the response header or in some other ITE-standardized, HTTP-compliant fashion.

    b. If it does not, the server redirects the GET request to authenticate.infotrust.org. Infotrust.org checks to seek if there is an Infotrust-issued domain cookie on the requesting browser.

        i. If there is, it reads the cookie, decryptes the Member/UserID it contains, and queries identity.Infotrust.org to see if session attributes exist for that UserID.

[1] -- Personally identifiable information (PII) is any data that could potentially identify a specific individual. Any information that can be used to distinguish one person from another and can be used for de-anonymizing anonymous data can be considered PII. As defined by the U.S. government, PII is ""information which can be used to distinguish or trace an individual's identity, such as their name, social security number, biometric records, etc. alone, or when combined with other personal or identifying information which is linked or linkable to a specific individual, such as date and place of birth, mother's maiden name, etc." See also NIST Guide to Protecting the Confidentiality of Personally Identifyable Information (April 2010)

1. If they do, authenticate.infotrust.org forwards user-authorized attributes to the content-providing HTTP server, which interprets them to personalize the response to the User.

2. If no session attributes for the User exist at identity.infotrust.org, then authenticate.infotrust.org returns the ProviderID from the decrypted UserID to the content-providing server, and also redirects the User to a log-in screen for the User's Service Provider. The User's Provider "logs in" its user then uploads user-authorized attributes to identity.infotrust.org. The process then loops back to "i.", above.

ii. If there is not, authenticate.infotrust.org first creates and tags a temporary session identifier to the requested URL. It then presents a login request to the user's browser, with a dialogue that allows the user to indicate (a) If they have a home-base ITE service provider and (b) If not, inviting them to sign up via the website they are visiting. After the signup process is complete, the process loops back to 3(a), above.

4. In the process of completing 3(a) the website that received the original GET request makes calls to each Advertising Service selected by the content Publisher to select (by RTB or other processes) Commercial Messages as part of the HTTP resource gotten by the User's browser.

   a. Each Advertising Service receives a GET request with the appended ITE-complaint query string which includes the unique UserID for the person logged into the device and (presumably) reading the page as well as the temporary session identifier.

   b. The Advertising Service, using ITE Mod on its webserver, next takes the ITE User ID and session identifier and, using ITE Protocol, asks identity.infotrust.org for the temporarily stored attributes of the User. The UserID is in an encrypted format which can be decrypted by identity.infotrust.org but not by the Advertising Service.

   c. The Exchange examines the temporary session identifier, determines that it is valid and current, and returns any Attributes of the User that the User has prospectively authorized for release.

   d. The Advertising Service using the temporary session identifier, now matches the User Attributes received from the Exchange with the Advertising position on the original URL, determines which ad to serve from inventory based on the Attributes, and returns that HTML code and images to the User's browser.

5. The Advertising Service, using ITE Protocol, then sends a Log report of the service of the ad to the advertising.infotrust.org, containing standard HTTP log fields, but including the User's ID, the session identifier, any Page Value Class for the advertisement, if the ITE is optionally expected to aggregate and clear a payment from The Advertising Service to either the user's Presenter/Service Provider or to the Provider/Publisher. Encrypted MemberIDs for each are included in the Log report.

6. Having served the GET response, and using ITE Data Transfer Protocol, the Provider/Publisher now sends a Log report of the HTTP service event to logging.infotrust.org, containing standard HTTP log fields, but including the User's ID, the session identifier, and the Page Value Class, and Markup Ratio, the Rights Owner if the page is being exchanged for value -- and any other elements determined by the ITE to be useful.

D. PROOF OF CONCEPT DEVELOPMENT

To execute a proof-of concept demonstration of this ITE user-sharing service the following principal efforts are required:

1. Code a modification to Apache and other common web servers which (1) intercepts GET requests and performs actions on them before other services are invoked, (2) Adds an encrypted UserID to outgoing GET responses, and (3) Communications with a designated ITE network server via ITE Data Transfer Protocol.

    a. One version can work for Providers/Publishers only.

    b. One version cal work for Advertising Services only.

    c. One version of the Provider/Publisher mod code can also have additional functions required of a Presenter (manages user accounts and serves User Attributes) as well as Provider/Publisher.

    d. The simplest versions do not have to support Page Value or Markup Ratio services where personalization, not payment is all that is required.

2. Develop and activate an ITE Exchange service which (a) uses ITE Protocol to communicate with the ITE Modules of Provider, Presenter and Advertising Service web services (b) Temporarily stores User IDs and Attributes, safeguarding and deleting them according to ITE business rules (c) Responds to HTTP requests from ITE Members and (d) Accepts and stores Log reports of ITE network events, for offline processing and aggregating for payments, for analysis by a User's home-base Presenter/Provider or for anonymous, aggregated analytics.