



- (51) **International Patent Classification:**
H04N 21/258 (2011.01) *H04N 21/442* (2011.01)
H04N 21/45 (2011.01)

(21) **International Application Number:**
PCT/US2015/067917

(22) **International Filing Date:**
29 December 2015 (29.12.2015)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**
62/098,787 31 December 2014 (31.12.2014) US
14/604,394 23 January 2015 (23.01.2015) US

(71) **Applicant:** THE NIELSEN COMPANY (US), LLC
[US/US]; 85 Broad Street, New York, New York 10004
(US).

(72) **Inventors:** SULLIVAN, Jonathan; 5 Village Way, Apt.
16, Natick, Massachusetts 01760 (US). PEREZ, Albert;
160 Bay Street, Apt. 421, San Francisco, California 94133
(US). SHEPPARD, Michael; 515 Argyle Road #2, Brook-
lyn, New York 11218 (US). TERRAZAS, Alejandro; 96
Frederick Street, Unit C, Santa Cruz, California 95062
(US).

(74) **Agent:** ZIMMERMAN, Michael W.; HANLEY, FLIGHT
& ZIMMERMAN, LLC, 150 S. Wacker Drive, Suite 2200,
Chicago, Illinois 60606 (US).

(81) **Designated States** (*unless otherwise indicated, for every
kind of national protection available*): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG,
MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM,
PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC,
SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** (*unless otherwise indicated, for every
kind of regional protection available*): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ,
TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU,
TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,
DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,
LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, KM, ML, MR, NE, SN, TD, TG).

Published:

Published:

— with international search report (Art. 21(3))

[Continued on next page]

- (54) Title:** METHODS AND APPARATUS TO CORRECT AGE MISATTRIBUTION IN MEDIA IMPRESSIONS

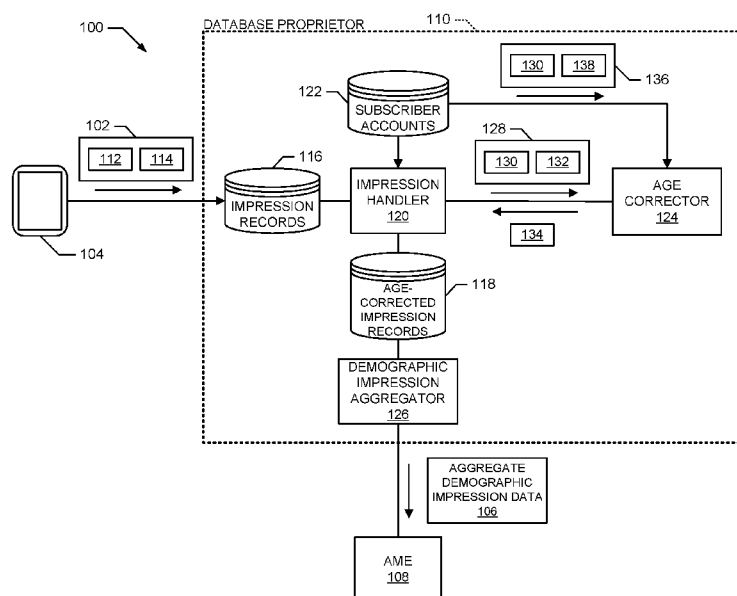


FIG. 1

- (57) Abstract:** This disclosure relates generally to audience measurement, and, more particularly, to correcting age misattribution in media impressions. An example method involves predicting a first age probability density function for an audience member at a first time, the first age probability density function indicative of a first probability that the audience member is in a first age range at the first time. The example method also involves, in response to receiving media monitoring data associated with the subscriber at a second time after the first time, determining a second age probability density function by applying a first aging factor to the first age probability density function for the audience member, the second age probability density function indicative of a second probability that the audience member is in the first age range at the second time, and associating the media monitoring data with the second age probability density function.



-
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

METHODS AND APPARATUS TO CORRECT AGE MISATTRIBUTION IN MEDIA IMPRESSIONS

FIELD OF THE DISCLOSURE

[0001] This patent claims the benefit of U.S. Provisional Patent Application Ser. No. 62/098,787, filed December 31, 2014, which is herein incorporated by reference in its entirety.

FIELD OF THE DISCLOSURE

[0002] This disclosure relates generally to audience measurement, and, more particularly, to correcting age misattribution in media impressions.

BACKGROUND

[0003] Traditionally, audience measurement entities determine audience engagement levels for media based on registered panel members. That is, an audience measurement entity enrolls people who consent to being monitored into a panel. The audience measurement entity then monitors those panel members to determine media (e.g., television programs or radio programs, movies, DVDs, advertisements, streaming media, websites, etc.) exposed to those panel members. In this manner, the audience measurement entity can determine exposure metrics for different media based on the collected media measurement data.

[0004] Techniques for monitoring user access to Internet resources such as web pages, advertisements and/or other Internet-accessible media have evolved significantly over the years. Some known systems perform such monitoring primarily through server logs. In particular, entities serving media on the Internet can use known techniques to log the number of requests received for their media (e.g., content and/or advertisements) at their server.

BRIEF DESCRIPTION OF THE DRAWINGS

[0005] FIG. 1 illustrates an example system to collect impressions 102 of media presented at computing devices and to produce aggregate age-correction demographic impression data.

[0006] FIG. 2 illustrates an example system to collect impressions of media presented at mobile devices and to correct the impressions for age misattribution at the audience measurement entity.

[0007] FIG. 3 illustrates an example implementation of the example age corrector of FIGS. 1 and 2 to correct age misattributions in media impressions.

[0008] FIG. 4 illustrates example age distribution functions used to correct age misattribution.

[0009] FIG. 5 illustrates an example data structure to store predicted age distribution functions for the subscribers of a database proprietor.

[0010] FIG. 6 is an example communication flow diagram of an example manner in which an audience measurement entity and a database proprietor can collect impressions and demographic information based on a client device reporting impressions to the audience measurement entity and the database proprietor.

[0011] FIG. 7 is a flow diagram representative of example machine readable instructions that may be executed to implement the example age corrector of FIGS. 1, 2, and/or 5 to age misattributions in media impressions.

[0012] FIG. 8 is a flow diagram representative of example machine readable instructions that may be executed to implement the example age updater of FIG. 5 to probabilistically age a prediction age distribution function.

[0013] FIG. 9 is a flow diagram representative of example machine readable instructions that may be executed to implement the example demographic impression aggregator of FIG. 1 to aggregate demographic impressions to be sent to the audience measurement entity.

[0014] FIG. 10 is a block diagram of an example processor system structured to execute the example machine readable instructions represented

by FIG. 7, 8, and/or 9 to implement the example age corrector of FIGS. 1, 2, and/or 5.

[0015] FIG. 11 depicts an example system to collect impressions of media presented on mobile devices and to collect user information from distributed database proprietors for associating with the collected impressions.

[0016] FIG. 12 depicts an example classification tree that may be employed to generate an age prediction model that uses subscriber activity metrics to predict real ages of subscribers of a database proprietor.

DETAILED DESCRIPTION

[0017] Examples disclosed herein may be used to correct age-based demographic group misattribution errors in impressions collected by database proprietors. As used herein, an impression is an instance of a person's exposure to media (e.g., content, advertising, etc.). When an impression is logged to track an audience member's exposure to particular media, the impression may be associated with demographics of the person corresponding to the impression. This is referred to as attributing demographic data to an impression, or attributing an impression to demographic data. As used herein, a demographic impression is an impression with attributed demographic data. In this manner, media consumption behaviors of audiences and/or media exposure across different demographic groups can be measured. Example demographic groups include demographic groups defined by a range of ages (sometimes referred to herein as "age-based demographic groups" and/or "demographic buckets"). For example, demographic buckets may be defined for ages 8-12, 13-17, 18-24, 25-34, 35-44, 45-54, 55-64, and 65+. However, age-based demographic group misattribution errors in collected impressions can occur when the actual age of the person corresponding to the impression is different than the reported age of the person. Such misattribution errors decrease the accuracy of audience measurement. To improve accuracies of impression data having age-based demographic group misattribution errors,

examples disclosed herein may be used to predict the actual age of a person, storing that prediction, and calculating an updated age when an impression corresponding to that person is logged.

[0018] An audience measurement entity (AME) measures the composition and size of audiences consuming media to produce ratings. Ratings are used by advertisers and/or marketers to purchase advertising space and/or design advertising campaigns. Additionally, media producers and/or distributors use the ratings to determine how to set prices for advertising space and/or to make programming decisions. As increasing numbers of audience members use computers (e.g., desktop computers, laptop computers, etc.), portable devices (e.g., tablets, smartphones, etc.), gaming consoles (e.g., Xbox One®, Playstation® 4, etc.) and/or online media presentation devices (e.g., Google Chromecast, Roku® Streaming Stick®, etc.) (sometimes collectively referred to herein as “computing devices”) to access media, advertisers and/or marketers are interested in accurately calculated ratings (e.g., online campaign ratings, etc.) for media accessed on these devices.

[0019] To measure audiences on computing devices, an AME may use instructions (e.g., Java, java script, or any other computer language or script) embedded in media as describe below in connection with FIG. 6 to collect information indicating when audience members access media on computing devices. Additionally, one or more user and/or device identifiers (e.g., an international mobile equipment identity (IMEI), a mobile equipment identifier (MEID), a media access control (MAC) address, a web browser unique identifier (e.g., a cookie), an app store identifier, an open source unique device identifier (OpenUDID), an open device identification number (ODIN), a login identifier, a username, an email address, user agent data, third-party service identifiers, web storage data, document object model (DOM) storage data, local shared objects, an automobile vehicle identification number (VIN), etc.) located on a computing device allow a partnered database proprietor (e.g., Facebook, Twitter, Google, Yahoo!, MSN, Apple, Experian, etc.) to identify demographic information (e.g., age, gender, geographic location, race, income level, education level, religion, etc.) for the audience member of the

computing device collected via a user registration process. For example, an audience member may be accessing an episode of “Boardwalk Empire” made available by a media streaming service (e.g. via a website, via an app, etc.). In that instance, in response to instructions executing within the browser or the app, a user/device identifier stored on the computing device is sent to the AME and/or a partner database proprietor to associate the instance of media exposure (e.g., an impression) to corresponding demographic data of the audience member. The database proprietor can then send logged demographic impression data to the AME for use by the AME in generating, for example, media ratings and/or other audience measures. In some examples, the partner database proprietor does not provide individualized demographic data (e.g., user-level demographics) in association with logged impressions. Instead, in some examples, the partnered database proprietor provides aggregate demographic impression data (sometime referred to herein as “aggregate census data”). For example, the aggregate demographic impression data provided by the partner database proprietor may state that a thousand males, ages 18-24, watched the episode of “Boardwalk Empire” in the last seven days via computing devices. The aggregate demographic data from the partner database proprietor does not identify individual persons (e.g., is not user-level data) associated with individual impressions. In some examples, the database proprietor provides anonymized user-level data (e.g., user-level data with personally identifiable information removed) and/or a subset of anonymized user-level data (e.g., gender, demographic bucket, activity metrics, race/ethnicity, etc.) to the AME. In this manner, the database proprietor protects the privacies of its subscribers/users by not revealing their identities to the AME.

[0020] The AME uses this census data to calculate ratings and/or other audience measures for corresponding media. However, during the process of registering with the database proprietor, a subscriber may lie or may otherwise be inaccurate about the subscriber’s age information (e.g., age, birth date, etc.). For example, a potential subscriber below a required age (e.g., thirteen, eighteen, twenty-one, etc.) may be prevented from registering with the

database proprietor unless the potential subscriber enters a false, but eligible, birth date. Often in such examples, the subscriber does not correct the provided age even after the user's true age is above the required age. As such, impressions collected by a database proprietor for such subscribers will potentially be attributed to the wrong age-based demographic group. For example, if a ten-year old person, during the registration process, represents that he a thirteen years old, the impression data associated with that subscriber will be misattributed to a thirteen-year old person instead of a ten-year old person.

[0021] The effect of large-scale misattribution error may create measurement bias error by incorrectly representing the demographic distribution of impressions across a large audience and, therefore, misrepresenting the audience demographics of impressions collected for advertisements and/or other media to which exposure is monitored by the AME. For example, when subscribers initially provide inaccurate age information at registration (e.g., report that they are older than they are, etc.), measured audience demographics may be skewed older than the actual audience demographics. For example, a misattribution error may cause an impression that should be assigned to a certain demographic bucket to be assigned to a different demographic bucket. For example, if the actual age of a subscriber is twenty-two (e.g., in an ages 19-24 demographic bucket), but the reported age of the subscriber is twenty-five (e.g., in an ages 25-34 demographic bucket), the misattribution error would cause an impression corresponding to that subscriber to be assigned to the wrong demographic bucket (e.g., the ages 25-34 demographic bucket instead of the ages 19-24 demographic bucket).

[0022] To correct impression data for age-based demographic group misattribution errors, the AME uses an age prediction model to predict the real age of a database proprietor subscriber. The age prediction model uses activity metrics (e.g., frequency of login, type of computing device used to login, number of connections (e.g., friends, contacts, etc.), privacy settings, etc.) of database proprietor subscribers to assign an age probability density function

(PDF) to the subscribers of the database proprietor. In some examples, through generating the age prediction model, sets of age PDFs are defined by the AME. The age PDFs define probabilities that the real age of a subscriber is within certain demographic buckets (e.g., age-range buckets). For example, an age PDF may state that the probability that the subscriber is in the 2-7 year-old demographic bucket is 0%, the probability that the subscriber is in the 8-12 year-old demographic bucket is 0%, the probability that the subscriber is in the 13-17 year-old demographic bucket is 7%, the probability that the subscriber is in the 18-24 year-old demographic bucket is 18%, the probability that the subscriber is in the 25-34 year-old demographic bucket is 63%, the probability that the subscriber is in the 35-44 year-old demographic bucket is 11%, the probability that the subscriber is in the 45-54 year-old demographic bucket is 1%, the probability that the subscriber is in the 55-64 year-old demographic bucket is 0%, and the probability that the subscriber is in the 65+ year-old demographic bucket is 0%.

[0023] The predicted age of the subscriber predicted by the age prediction model is used as part of the demographic data that is attributed to the impression. However, the predictive value of the age prediction model degrades over time as subscriber behaviors (e.g., as measured by subscriber activity metrics) on which the age prediction model is based change over time. For example, an age prediction model can be developed on day $d=0$, based on subscriber behaviors that are true as of day $d=0$ for different age groups. As subscriber behaviors change for a particular age group over time, the same age prediction model will generate inaccurate results when used at a later time. For example, at day $d=0$, a typical 21-24 year-old male may log into Internet services via a mobile phone for 30% of his total logins. However, at day $d=183$ (6 months), 21-24 year old males might log in to Internet services using mobile phones for 50% of total logins. As such, the age prediction model generated at day $d=0$ based on a device type criterion will generate inaccurate results when used on day $d=183$ due to how log in behaviors have generally changed for subscribers in the 21-24 year-old male age-based demographic group. To counteract the degradation of the age prediction model's predictive

value, the AME may, from time to time, regenerate the model. However, generating the age prediction model requires considerable resources (e.g., time, processing power, bandwidth, memory usage, etc.). Between the time the age prediction model is generated and the time that the age prediction model is used to predict age, the likelihood that some age predictions (e.g., the terminal node assignments) made by the age prediction model may be inaccurate increases over time.

[0024] As disclosed below, to increase accuracies of ages attributed to impressions between the time the age prediction model is generated and the time that the age prediction will be regenerated, the AME and/or the database proprietor predicts the ages of the subscribers of a database subscriber when the age prediction model is initially generated. The age predictions are stored in an age-cache with, for example, a user identifier (UID), an age PDF assigned by the age prediction model, and the date the prediction was made. A UID is used by the database proprietor to identify user activity. For example, a UID may be a device/user identifier, a user name, an alphanumeric code randomly generated when the user registers, an anonymized identifier, an email address, etc. In some examples, the AME and/or the database proprietor, to protect subscriber privacy, do not store activity metrics on which the age prediction is based in the age-cache.

[0025] When impression data corresponding to a subscriber is received, instead of generating a new age PDF through the age prediction model, the predicted age PDF for that subscriber is retrieved from the age-cache and is probabilistically aged. That is, the age PDF is adjusted to account for the probability of the subscriber aging into a new demographic bucket between the date that the age prediction model was used and the date the impression was logged. The adjusted age PDF is then attributed to the corresponding impression.

[0026] The AME may, from time to time, still regenerate the age prediction model. However, the time between regenerating the age prediction model may be extended because the degradation of the model has been ameliorated by probabilistically aging the predicted age PDFs. In such a

manner, computing resources are conserved by decreasing the number of times the age prediction model needs to be generated to determine acceptably accurate ages. In some examples, upon regenerating the age prediction model, the AME and/or database proprietor may use the age prediction model to assign predicted age PDFs to the subscribers of the database proprietor. In some examples, the AME and/or database proprietor may assign age PDFs to all of the subscribers after generating an age prediction model. In some examples, to conserve processing resources, the AME and/or database proprietor may use the age prediction model to assign predicted age PDFs to subscribers that have registered with the database proprietor since the last age prediction model was generated (e.g., new subscribers).

[0027] Disclosed example methods to correct age misattribution in media impressions involve predicting a first age probability density function for an audience member at a first time, the first age probability density function indicative of a first probability that the audience member is in a first age range at the first time. The example methods further involve, in response to receiving media monitoring data associated with the audience member at a second time after the first time, determining a second age probability density function by applying a first aging factor to the first age probability density function for the audience member, the second age probability density function indicative of a second probability that the audience member is in the first age range at the second time, and associating the media monitoring data with the second age probability density function.

[0028] In some example methods involve determining the first aging factor by dividing a first number of days between the first time and the second time by a second number of total days in the first age range.

[0029] Some example methods further involve determining the first aging factor, the first aging factor being a first number of days between the first time and the second time divided by a second number of days in the first age range.

[0030] In some example methods, the first age probability density function is indicative of a third probability that the audience member is in a

second age range at the first time, and the example method further involves, in response to receiving the media monitoring data associated with the audience member at the second time after the first time, determining the second age probability density function by applying a second aging factor to the third probability indicated by the first age probability density function for the audience member to create a fourth probability that the audience member is in the second age range at the second time. In some such example methods, the fourth probability comprises a portion of the first probability and a portion of the third probability.

[0031] In some example methods, the audience member is a subscriber to a database proprietor, and the example methods further involve predicting the first age probability density function for the audience member further using an age prediction model based on subscriber characteristics of the audience member collected by the database proprietor. In some such example methods, the age prediction model is a classification tree and the first age probability density function is a terminal node of the classification tree.

[0032] Disclosed example apparatus include an age predictor to predict a first age probability density function for an audience member at a first time, the first age probability density function indicative of a first probability that the audience member is in a first age range at the first time. The example apparatus further include an age updater to, in response to receiving media monitoring data associated with the audience member at a second time after the first time, determine a second age probability density function by applying a first aging factor to the predicted first age probability density function for the audience member, the second age probability density function indicative of a second probability that the audience member is in the first age range at the second time, and associate the media monitoring data with the second age probability density function.

[0033] In some example apparatus, the age updater is further to determine the first aging factor by dividing a first number of days between the first time and the second time by a second number of total days in the first age range.

[0034] In some example apparatus, the first age probability density function is indicative of a third likelihood that the audience member is in a second age range at the first time, and wherein in response to receiving media monitoring data associated with the audience member at the second time after the first time, the age updater is further to determine the second age probability density function by applying a second aging factor to the third probability of the first age probability density function for the audience member to create a fourth probability that the audience member is in the second age range at the second time. In some such apparatus, the fourth probability comprises a portion of the first probability and a portion of the third probability.

[0035] In some example apparatus, the audience member is a subscriber to a database proprietor, and the age predictor is to predict the first age probability density function for the audience member fusing an age prediction model based on subscriber characteristics of the audience member collected by the database proprietor. In some such apparatus, the age prediction model is a classification tree and the first age probability density function is a terminal node of the classification tree.

[0036] FIG. 1 illustrates an example system 100 to collect impression data 102 of media presented at computing devices 104 (e.g., desktop computers, laptop computers, tablets, smartphones, gaming consoles, Google Chromecast, Roku® Streaming Stick®, etc.) and to produce aggregate age-corrected demographic impression data 106 to be used by an audience measurement entity (AME) 108. In the illustrated example, the impression data 102 is collected by a database proprietor 110 (e.g., Facebook, Twitter, Google, Yahoo!, MSN, Apple, Experian, etc.). In some examples, the impression data 102 and/or separate impression data is also collected by the AME 108. In some examples, the aggregate demographic impressions 106 are used by the AME 108 to generate ratings for media accessed at computing devices (e.g., the computing device 104.)

[0037] In the illustrated example, after media is accessed on the computing device 104, the computing device 104 reports the impression data

102 to the database proprietor 110. In some examples, the computing device 104 reports impression data 102 for accessed media based on instructions embedded in the media that instruct the computing device 104 (e.g., instruct a web browser or an app in the computing device 104) to send impression data 102 to the database proprietor 110. The example impression data 102 includes an impression request 112 and a device/user identifier 114. The example impression request 112 includes a media identifier (e.g., an identifier that can be used to identify content, an advertisement, and/or any other media) corresponding to the media accessed on the computing device 104. In some examples, the AME 108 modifies and/or encodes the media identifiers so the database proprietor 110 cannot identify the media. In some examples, the impression request 112 also includes a site identifier (e.g., a URL) of a website (e.g., YouTube.com, ABC.com, Hulu.com, etc.) that served the media to the computing device 104 and/or a host website ID (e.g., sbnation.com) of the website that displays or presents the media. In some examples, the impression request 112 includes an impression identifier that may be used to uniquely identify the impression request. In some examples, the device/user identifier 114 is a device identifier (e.g., an international mobile equipment identity (IMEI), a mobile equipment identifier (MEID), a media access control (MAC) address, etc.), a web browser unique identifier (e.g., a cookie), a user identifier (e.g., a user name, a login ID, etc.), an Adobe Flash® client identifier, identification information stored in an HTML5 datastore, a vehicle identification number (VIN) and/or any other identifier that the database proprietor 110 stores in association with demographic information about subscribers corresponding to the computing devices 104. In some examples, the database proprietor 110 uses the device/user identifier 114 to determine if the user of the computing device 104 is a subscriber of the database proprietor 110.

[0038] In the illustrated example of FIG. 1, the database proprietor 110 includes an example impression records database 116, an example age-corrected demographic impression records database 118, an example impression handler 120, an example subscriber accounts database 122, an

example age corrector 124, and an example demographic impression aggregator 126. The example impression records database 116 stores the impression data 102 received from the computing device 104. The example age-corrected demographic impression records database 118 stores age-corrected impression data generated by the example impression handler 120. In some examples, the impression records database 116 and the example age-corrected impression records database 118 may be the same database.

[0039] In the illustrated example, the impression handler 120 generates demographic impression data by matching demographic information from the subscriber accounts database 122 to the impression data 102. In the illustrated example, the impression handler 120 uses the user/device identifier 114 to determine if the impression data 102 is associated with a subscriber corresponding to an account in the subscriber accounts database 122. In some examples, a value (e.g., a user/device identifier) matching the user/device identifier 114 is stored in a corresponding subscriber account record in the subscriber accounts database 122. In some examples, the impression handler 120 may store and/or otherwise access an intermediate table that associates the user/device identifiers 114 with identifiers used to index subscribers (e.g., a subscriber identifier 130) in the subscriber accounts database 122. For example, the impression handler 120 may look up on an intermediate table a cookie value of a cookie used as a user/device identifier 114 to determine the subscriber identifier 130. In some examples, if the database proprietor 110 cannot find a match for the user/device identifier 114 and a user/device identifier in the subscriber accounts database 122, the corresponding impression data 102 is not processed further. The example impression handler 120 retrieves demographic information corresponding to the subscriber identifier 130 from the subscriber accounts database 122.

[0040] In the illustrated example, the example impression handler 120 sends an age-correction request 128 to the example age corrector 124. In the illustrated example, the age-correction request 128 includes a user identifier 130 (UID) (e.g., the user/device identifier 114, the subscriber identifier, etc.) and an impression date 132. The example impression date 132 is a date

corresponding to when the impression data 102 was generated by the computing device 104 (e.g. included with the impression data 102 when the impression data is sent to the database proprietor 110). Based on the UID 130 and the impression date 132, the example age corrector 124 generates an age-corrected probability density function (PDF) 134. In some examples, the age-corrected PDF 134 define probabilities that the real age of the subscriber corresponding to the subscriber identifier 130 is within certain demographic buckets (e.g., demographic groups defined by age ranges). For example, the age-corrected PDF 134 may indicate that the probability of the subscriber being in the 18-21 age range is 11.6%, the probability of the subscriber being in the 22-27 age range is 44.5%, the probability of the subscriber being in the 28-33 age range is 36.7%, and the probability of the subscriber being in the 34-40 age range is 7.2%.

[0041] In the illustrated example, the impression handler 120 generates a demographic impression by attributing the demographic data (e.g., without the subscriber-reported age) retrieved from the subscriber accounts database 122 and the age-corrected PDF 134 to the impression data 102. The example impression handler 120 stores the demographic impressions in the example age-corrected demographic impression records database 118.

[0042] The example demographic impression aggregator 126 creates aggregate demographic impressions 106 using the demographic impressions stored in the example age-corrected demographic impression database 118. By providing aggregate demographic impressions 106, the database proprietor 110 protects the privacies of its subscribers by not revealing their identities or subscriber-level media access activities, to the AME 108. In some examples, the demographic impressions are aggregated by media. For example, the demographic impression aggregator 126 may create aggregate demographic impressions 106 for season two, episode one of the television program titled “Boardwalk Empire.” In some examples, the demographic impressions are aggregated for a certain time period. For example, the demographic impression aggregator 126 may create aggregate demographic impressions 106 for an advertisement for a food product, such as a “Kellogg’s® Eggo®

Cinnamon Toast Waffles” commercial, for the past seven days. The example demographic impression aggregator 126 creates aggregate demographic impressions 106 by demographic group (e.g., gender, income level, race/ethnicity, marital status, etc.). For example, the demographic impression aggregator 126 may create aggregate demographic impressions 106 for unmarried males accessing the television program titled “Being Human.”

[0043] In some examples, the demographic impressions are aggregated as granularly (e.g., aggregated by many demographic categories (e.g., ethnicity, religion, education level, gender, income, etc.)) or as coarsely (e.g., aggregated by only a few demographic categories) as agreed between the database proprietor 110 and the AME 108. For example, the database proprietor 110 and the AME 108 may agree that the database proprietor 110 will provide aggregate demographic impressions 106 for the television program titled “Scandal” categorized by gender, ethnicity, and education level.

[0044] To generate the aggregate demographic impressions 106, the demographic impression aggregator 126 aggregates the age-corrected PDF 134 stored in the age-corrected demographic impression records database 118. In some examples, to aggregate the age-corrected PDFs, the demographic impression aggregator 126 sums the age probability value for a corresponding demographic bucket of the demographic impressions being aggregated and divides the summed probability values by the number of demographic impressions being aggregated. In some such examples, the demographic impression aggregator 126 repeats this process for each demographic bucket. In some examples, the demographic impression aggregator 126 generates an aggregate age-corrected PDF in accordance with Equation 1 below.

$$AAP_j = \frac{\sum_{i=0}^n ADF_j(i)}{n}$$

Equation 1

In Equation 1 above, AAP_j is the aggregated age probability for demographic bucket j , i is a demographic impression, n is the number of demographic impressions to be included in the aggregate demographic impressions 106, and ADF_j is the age-corrected PDF probability value for demographic bucket j . In

some examples, the demographic impression aggregator 126 compiles the aggregated age probabilities (AAP_j) for the demographic buckets into the aggregate age-corrected PDF that is included in the aggregate demographic impression 106.

[0045] Consider the following example for males that accessed season 1, episode 2 of a television show titled “Being Human” in the last seven days. The example corrected-age distribution functions 134 for demographic impressions A through D are given in table 1 below.

	Age 12-14	Age 15-17	Age 18-20	Age 21-24	Age 25-29	Age 30-34
A	0.200	0.350	0.300	0.100	0.025	0.025
B	0.000	0.700	0.150	0.100	0.000	0.050
C	0.000	0.000	1.000	0.000	0.000	0.000
D	0.100	0.350	0.400	0.075	0.075	0.000

EXAMPLE AGE-CORRECTED PROBABILITY DENSITY FUNCTIONS
CORRESPONDING TO EXAMPLE DEMOGRAPHIC IMPRESSIONS A
THROUGH D

Table 1

In the illustrated example, the aggregate age probability for the age range 12-14 demographic bucket is 7.5% $((0.2+0.0+0.0+0.1)/4)$, the aggregate age probability for the age range 15-17 demographic age bucket is 35% $((0.35+0.7+0.0+0.35)/4)$, the aggregate age probability for the age range 18-20 demographic age bucket is 46.3% $((0.3+0.15+1.0+0.4)/4)$, the aggregate age probability for the age range 21-24 demographic age bucket is 6.9% $((0.1+0.1+0.0+0.075)/4)$, the aggregate age probability for the age range 25-29 demographic age bucket is 2.5% $((0.025+0.0+0.0+0.075)/4)$, and the aggregate age probability for the age range 30-34 demographic age bucket is 1.8% $((0.025+0.05+0.0+0.0)/4)$. Based on the example corrected-age distribution functions in Table 2 above, the aggregate corrected-age PDF attributed to the aggregate demographic impressions 106 corresponding to males that accessed season 1, episode 2 of the television program titled “Being Human” in the last seven days is given in Table 2 below.

Age 12-14	Age 15-17	Age 18-20	Age 21-24	Age 25-29	Age 30-34
0.075	0.350	0.463	0.069	0.025	0.018

EXAMPLE AGGREGATE AGE-CORRECTED PROBABILITY DENSITY
FUNCTION

Table 2

[0046] In the illustrated example, from time to time, the age corrector 124 retrieves subscriber data 136 from the subscriber accounts database 122 in order to predict age PDFs for the subscribers of the database proprietor 110. In some examples, the age corrector 124 retrieves subscriber data 136 after a new prediction model is generated. The example subscriber data 136 includes the example UID 130 and example subscriber activity metrics 138. The example subscriber activity metrics 138 are collected and/or calculated based on subscriber behavior and/or demographic factors. The example subscriber activity metrics 138 include self-reported demographic data and behavioral data of a subscriber of the database proprietor 110. The example subscriber activity metrics 138 include login frequency, mobile login activity, privacy settings, post rate, number of contacts, days since registration, whether a cell phone number has been included in registration information, etc. Using the example UID 130 and example subscriber activity metrics 138, the example age corrector 124 predicts and stores age PDFs for the subscribers corresponding to the retrieved subscriber data 136.

[0047] FIG. 2 illustrates an example system 200 to collect impression data 102 of media accessed at computing devices and to correct the impression data 102 for age misattribution at the AME 108. In the illustrated example, the computing device 104 sends impression data 102 to the AME 108. In some examples, the computing device 104 sends impression data 102 to the database proprietor 110. The example database proprietor 110 sends anonymized subscriber demographic data 202 to the AME 108. The example anonymized subscriber demographic data 202 is demographic data from a subscriber database (e.g., the subscriber database 122 of FIG. 1) that has one or more items of personally identifiable information (e.g., name, home address, email

address, date of birth, birthplace, telephone number, national identification number, etc.) removed. In some examples, the anonymized subscriber demographic data 202 also includes an identifier that may be used to match the anonymized subscriber demographic data 202 with impression data 102. In some examples, the database proprietor 110 provides the anonymized subscriber demographic data 202 in response to receiving the impression data 102. In some examples, the database proprietor 110 provides the anonymized subscriber demographic data 202 in response to receiving a request from the AME 108 for the anonymized subscriber demographic data 202. In some examples, the database proprietor 110 provides, from time to time (e.g., at periodic or aperiodic intervals such as every 24 hours or when a particular amount of data has been collected), the anonymized subscriber demographic data 202 in bulk.

[0048] In some examples, the database proprietor 110 provides anonymized identifiers with the anonymized subscriber demographic data 202. In some examples, to facilitate correlating the impression data 102 with the anonymized subscriber demographic data 202, the database proprietor 110 sends subscriber correlation data 205 to the AME 108. In some examples, the subscriber correlation data 205 includes the impression identifier included in the impression request 112 and the corresponding anonymized identifier. In some examples, the subscriber correlation data 205 is sent to the AME 108 in response to the impression data 102 being received by the database proprietor 110. In some examples, the database proprietor 110 sends subscriber correlation data 205 to the AME 108 from time to time (e.g., at periodic or aperiodic intervals such as every 24 hours or when a particular amount of impression data 102 has been collected). In such a manner, the anonymized identifier allows the AME 108 to correlate anonymized subscriber demographic data 202 with impression data 102 without allowing the AME 108 to identify the particular subscriber that caused the impression data 102 to be generated.

[0049] In the illustrated example, the AME 108 includes an example anonymized subscriber database 203, an example impressions collector 204,

an example age corrector 124, an example age adjuster 206, and an example age-adjusted impressions database 208. The example anonymized subscriber database 203 stores the example anonymized subscriber demographic data 202 received from the example database proprietor 110. In the example illustrated in FIG. 2, the impressions collector 204 receives the impression data 102 from the computing device 104. The example impressions collector 204 generates demographic impressions by attributing anonymized subscriber data 202 with impression data 102. In some examples, the impressions collector 204 compares an impression identifier (e.g., the impression identifier 1140 of FIG. 11 below) of the impression request 112 included in the impression data 102 to subscriber correlation data 205 received from the database proprietor 110. In such examples, if a corresponding impression identifier is identified, the impression data 102 is attributed to the anonymized subscriber data 202 corresponding to the anonymized identifier included with the subscriber correlation data 205.

[0050] In the illustrated example, the AME 108 receives subscriber data 136 for one or more subscribers of the database proprietor 110. The example subscriber data 136 includes a user identifier 130 and subscriber activity metrics 138. In some examples, the user identifier 130 is the anonymized identifier included in the anonymized subscriber demographic data 202. In some examples, the subscriber data 136 has personally identifiable information removed. Using the user identifier 130 and the example subscriber activity metrics 138, the example age corrector 124 predicts and stores age PDFs for the subscribers corresponding to the anonymized subscriber data 205. For example, the example age adjuster 206 sends an age-correction request 128 to the example age corrector 124. In the illustrated example, the age-correction request 128 includes a user identifier 130 (e.g., an anonymized subscriber identifier, a user/device identifier 114, etc.) and an impression date 132. Based on the user identifier 130 and the impression date 132, the example age corrector 124 generates an age-corrected PDF 134. The example age adjuster 208 attributes the age-corrected PDF 134

to the demographic impression data and stores the demographic impression data in the age-adjusted impression database 208.

[0051] FIG. 3 illustrates an example implementation of the example age corrector 124 of FIGS. 1 and 2 to correct age misattributions in demographic impression data. The example age corrector 124 generates corrected age PDF(s) 134 of subscribers registered with a database proprietor (e.g., the database proprietor 110 of FIGS. 1 and 2) based on subscriber activity metrics (e.g., frequency of login, type of computing device used to login, number of connections (e.g., friends, contacts, etc.)), privacy settings, etc.). In the illustrated example, when the age corrector 124 receives an age request 128, the age corrector 124 generates the corrected age PDF(s) 134. In some examples, the age requests 128 include one or more user identifiers (UIDs) 130 (e.g., a user/device identifier 114, an anonymized identifier, etc.) with corresponding impression dates 302 on which one or more impressions were logged (e.g., by the database proprietor 110) in association with the UID(s) 130. The example corrected age PDF(s) 134 define probabilities that the real age of a subscriber is within certain demographic buckets as of a corresponding impression date 302.

[0052] In the illustrated example of FIG. 3, the age corrector 124 includes an example age predictor 304, an example age prediction model 306, an example age cache 308, and an example age updater 310. The example age predictor 304 is structured to predict an age PDF 305 for a subscriber of the database proprietor 110 based on subscriber data 136 (FIG. 1) provided by the database proprietor 110. The age PDFs 305 define probabilities that the real age of the subscriber is within certain demographic buckets on the date the prediction is made (e.g., the date on which the age PDF 305 is determined). The example age predictor 304 stores the predicted age PDF(s) in the age cache 308 with the UID 130 and the date that the age PDF was predicted. In some examples, to protect subscriber privacy, the age predictor 304 does not store the subscriber activity metrics 138 provided with the subscriber data 136 in the age cache 308.

[0053] The example age prediction model 306 receives the subscriber activity metrics 138 included with the subscriber data 136 and generates age PDFs. The AME 108 generates the example age prediction model 306 by analyzing subscriber activity metrics of subscribers of the database proprietor 110 that are also enrolled as panelists by the AME 108 (FIG. 1) and the demographic information supplied by the panelist (e.g., when the panelist is enrolled). For example, when a person is enrolled to be a panelist, the person informs the AME 108 of which database proprietors 110 the person subscribes. In that scenario, the AME 108 obtains consent from the enrolled panelists to use the subscriber activity metrics 138 from the database proprietor 110 and the demographic information collected by the AME 108 to contribute to an age prediction model 306. While the self-reported demographic information (e.g., age, etc.) reported to the database proprietor 110 is generally considered inaccurate, the demographic information collected from the panelist (e.g., via a survey, etc.) by the AME 108 is considered highly accurate.

[0054] In some examples, the age prediction model 306 is implemented using a classification tree. In some such examples, using the subscriber activity metrics 138 and demographic data of the panelists, a classification tree model is generated comprising intermediate (e.g., decision) nodes and terminal (e.g., prediction) nodes. The intermediate nodes contain test conditions based on the subscriber activity metrics 138 and demographic data and result in branching paths. For example, an intermediate node may branch in one direction if a login frequency (e.g., the number of times a week the subscriber logs into the database proprietor 110) is greater than five times per week, and may branch in another direction if the login frequency is less than or equal to five times per week. The branching paths may either lead to another intermediate node or a terminal node.

[0055] In such a scenario, the terminal nodes represent a classification (e.g., a predicted age PDF 305) based on the subscriber activity metrics 138 and demographic data of the panelist. After the classification tree is constructed, the panelists' data is application to the classification tree so that

the panelists are classified into the terminal nodes. In some such examples, the age PDF 305 associated with that terminal node is percentage of panelists in the terminal node that fall within a corresponding demographic bucket based on the panelists' real ages. For example, of the panelists in a terminal node, 0% of the panelist may be within the 2-7 year-old demographic bucket, 0% of the panelist may be within the 8-12 year-old demographic bucket, 7% of the panelist may be within the 13-17 year-old demographic bucket, 18% of the panelist may be within the 18-24 year-old demographic bucket, 63% of the panelist may be within the 25-34 year-old demographic bucket, 11% of the panelist may be within the 35-44 year-old demographic bucket, 1% of the panelist may be within the 45-54 year-old demographic bucket, 0% of the panelist may be within the 55-64 year-old demographic bucket, and 0% of the panelist may be within the 65+ year-old demographic bucket. Examples that may be used to generate the age prediction model 306 are disclosed in U.S. Patent Application No. 13/209,292, entitled "Methods and Apparatus to Analyze and Adjust Demographic Information," which is incorporated by reference in its entirety.

[0056] In some examples, when the age prediction model 306 is generated, a table or a database is created that correlates an age PDF identifier (ID) with an age PDF 305. In some such examples, the age prediction model 306 outputs the age PDF identifier (ID) corresponding to the predicted age PDF 305. In some such examples, the age predictor 304 looks up the age PDF ID in a table and/or database to retrieve the predicted age PDF 305 to store in the age cache 308. In some examples, the age predictor 304 stores the age PDF ID in the age cache 212. Alternatively, in some examples, the age prediction model 306 outputs the predicted age PDF 305, which is stored by the age predictor 304 in the age cache 308.

[0057] FIG. 4 shows an example terminal node table 400 showing age PDFs 305a-305c. In the illustrated example, the age PDFs 305a-305c on the terminal node table 400 are identified by age PDF IDs 404 (APDFID 404). The example age PDFs 305a-305c are divided into demographic bucket probabilities 406a-406m. The demographic bucket probabilities 406a-406m

define the probability that the real age of the subscriber falls within the corresponding demographic bucket. When the example age predictor 304 (FIG. 3) uses the example age prediction model 306 (FIG. 3) based on the user subscriber activity metrics 138, the age prediction model 306 outputs one of the age PDFs 305a-305c and/or age PDF IDs 404. For example, if the age prediction model 306 assigns a first subscriber to an age PDF 305b identified by the age PDF ID 404 "A2," the probability that the real age of the first subscriber is 12-14 is 100%. As another example, if the age prediction model 306 assigns a second subscriber to an age PDF 305c identified by the age PDF ID 404 "A3," the probability of the real age of the second subscriber being 2-11 is 6.2%, the probability of the real age of the second subscriber being 12-14 is 32.3%, the probability of the real age of the second subscriber being 15-17 is 60%, and the probability of the real age of the second subscriber being 35-39 is 1.5%.

[0058] FIG. 5 illustrates an example data structure 500 to store age PDF(s) (e.g., the age PDF(s) 305 of FIGS. 3 and 4) predicted by age prediction model 306 (FIG. 3). In the illustrated example, the data structure 500 includes an example user identifier 130, an example age PDF identifier (age PDF ID) 404, and an example prediction date 502. The example user identifier 130 is used by the age updater 310 (FIG. 3) to retrieve the age PDF ID 404 and the prediction data 502 when an age request 128 (FIGS. 1 and 2) is received. The example user identifier 130 may be the user/device identifier 114 (FIG. 1), the anonymized identifier, and/or any other identifying value that remains a consistent identifier for a subscriber between the prediction date 502 and the impression date 302 (FIG. 3). The age PDF ID 404 identifies an age PDF that was predicted for the subscriber based on subscriber activity metrics 138 (FIG. 1) on the prediction date 502.

[0059] In the illustrated example of FIG. 3, the example age updater 310 retrieves the age PDF ID 404 (FIG. 4) and the prediction date 502 (FIG. 5) from the age cache 308 upon receiving an age request 128. In the illustrated example, the age updater 310 probabilistically ages the age PDF retrieved from the age cache 308 to produce an age-corrected PDF 134. To

probabilistically age the age PDF, the age updater 310 calculate aging factor(s) (AF) in accordance with Equation 2 below.

$$AF_j = \frac{NumDays(D_I - D_P)}{NumDays(D_j)}$$

Equation 2

In Equation 2 above, AF_j is an aging factor for the demographic bucket j (e.g., one of the demographic buckets 406a-406m of FIG. 4), D_I is the impression date 302, D_P is the prediction date 502, $NumDays(D_I - D_P)$ is the number of days between the impression date 302 and the prediction date 502, and $NumDays(D_j)$ is the number of days that are spanned by the demographic bucket j . For example, a three-year spanning demographic bucket (e.g., an age 18-20 demographic age bucket, etc.) has 1096 days. For example, if the impression date 302 is December 5, 2014, the prediction date 502 is April 17, 2014, and the span of the demographic bucket is three years, the aging factor would be 0.21 (232/1096) because the number of days between April 17, 2014 and December 5, 2014 is 232. In some examples, where the age ranges of the demographic buckets are the same, the age updater 310 only calculates a single aging factor. For example, if the demographic buckets each span five years (e.g., demographic bucket A spans ages 7-11, demographic bucket B spans ages 12-16, demographic bucket C spans ages 17-21, etc.), the age factors would be identical (e.g., $NumDays(D_j)$ would be 1826).

[0060] To probabilistically age the age PDF corresponding to the age PDF ID 404 retrieved from the age cache 308, the age updater 310 applies the aging factor(s) to the age PDF in accordance with Equation 3 below.

$$ACPDF = APDF \times M$$

Equation 3

In Equation 3 above, $ACPDF$ is the age-corrected PDF 134, $APDF$ is a $1 \times N$ matrix formed by the age PDF retrieved from the age cache 308, where N is the number of demographic buckets in the age PDF, and M is an aging factor matrix calculated in accordance with Equation 4 below.

$$M = \begin{pmatrix} 1 - AF_1 & AF_1 & 0 & \dots & 0 & 0 \\ 0 & 1 - AF_2 & AF_2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 - AF_{j-1} & AF_{j-1} \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

Equation 4

In Equation 4 above, AF_j is the aging factor for demographic bucket j .

[0061] Consider the following example. An age request 128 is received by the age updater 310 with a UID 130 of “8PR6PYRGQC” and an impression date 302 of October 10, 2014. The age updater 310 retrieves a record from the age cache 308 that has an age PDF ID 404 of “A7” and a prediction date 502 of June 9, 2014. The age PDF ID 404 “A7” corresponds to an age PDF as shown in Table 3 below.

Demographic Bucket	13-17	18-21	22-27	28-34	34-40
Probability	0%	15%	53%	32%	0%

EXAMPLE AGE PDF

Table 3

Using Equation 2 above, that age updater 310 calculates the aging factors as shown below in Table 4.

Demographic Bucket	1	2	3	4	5
Aging Factor	0.068	0.085	0.057	0.049	0.049

EXAMPLE AGING FACTORS BASED ON THE EXAMPLE AGE

PDF OF TABLE 3

Table 4

A matrix, M , applying Equation 4 above to the aging factors in Table 4 above is shown in Equation 5 below.

$$M = \begin{pmatrix} 0.932 & 0.068 & 0 & 0 & 0 \\ 0 & 0.915 & 0.085 & 0 & 0 \\ 0 & 0 & 0.943 & 0.057 & 0 \\ 0 & 0 & 0 & 0.951 & 0.049 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Equation 5.

Applying Equation 3 to Equation 5 above and the example age PDF of Table 3 above, the age-corrected PDF 134 determined by the example age updater 310 is shown in Table 5 below.

Demographic Bucket	13-17	18-21	22-27	28-34	34-40
Probability	0%	14%	51%	33%	2%

EXAMPLE AGE-CORRECTED PDF

Table 5

[0062] While an example manner of implementing the example age corrector 124 of FIGS. 1 and/or 2 is illustrated in FIG. 3, one or more of the elements, processes and/or devices illustrated in FIG. 3 may be combined, divided, re-arranged, omitted, eliminated and/or implemented in any other way. Further, the example age predictor 304, the age prediction model 306, the age cache 308, the example age updater 310 and/or, more generally, the example age corrector 124 of FIG. 3 may be implemented by hardware, software, firmware and/or any combination of hardware, software and/or firmware. Thus, for example, any of the example age predictor 304, the age prediction model 306, the age cache 308, the example age updater 310 and/or, more generally, the example age corrector 124 could be implemented by one or more analog or digital circuit(s), logic circuits, programmable processor(s), application specific integrated circuit(s) (ASIC(s)), programmable logic device(s) (PLD(s)) and/or field programmable logic device(s) (FPLD(s)).

When reading any of the apparatus or system claims of this patent to cover a purely software and/or firmware implementation, at least one of the example age predictor 304, the age prediction model 306, the age cache 308, the example age updater 310 and/or, the example age corrector 124 is/are hereby expressly defined to include a tangible computer readable storage device or storage disk such as a memory, a digital versatile disk (DVD), a compact disk (CD), a Blu-ray disk, etc. storing the software and/or firmware. Further still, the example age corrector 124 of FIGS. 1 and 2 may include one or more elements, processes and/or devices in addition to, or instead of, those

illustrated in FIG. 3, and/or may include more than one of any or all of the illustrated elements, processes and devices.

[0063] FIG.6 illustrates an example communication flow diagram of an example manner in which the AME 108 and the database proprietor 110 can collect impressions and demographic information based on a client device 104 reporting impressions to the AME 108 and the database proprietor 110. Examples disclosed herein to predict age may be used in connection with demographic impressions collected using example techniques described below in connection with FIG. 6. FIG. 6 also shows example age correctors 124. In some examples, the age corrector 124 is part of the database proprietor 110 as shown in FIG. 1. In some examples, the age correct 124 is part of AME 108 as shown in FIG. 2. The example chain of events shown in FIG. 6 occurs when the client device 104 accesses media for which the client device 104 reports an impression to the AME 108 and the database proprietor 110. In some examples, the client device 104 reports impressions for accessed media based on instructions (e.g., beacon instructions) embedded in the media that instruct the client device 104 (e.g., instruct a web browser or an app in the client device 104) to send beacon/impression requests 608 to the AME 108 and/or the database proprietor 110. In such examples, the media having the beacon instructions is referred to as tagged media. In other examples, the client device 104 reports impressions for accessed media based on instructions embedded in apps or web browsers that execute on the client device 104 to send beacon/impression requests 608 to the AME 108, and/or the database proprietor 110 for corresponding media accessed via those apps or web browsers. In any case, the beacon/impression requests 608 include device/user identifiers 114 as described further below to allow the corresponding AME 108 and/or database proprietor 110 to associate demographic information with resulting logged impressions.

[0064] In the illustrated example, the client device 104 accesses media 606 that is tagged with beacon instructions 608. The beacon instructions 608 cause the client device 104 to send a beacon/impression request 612 to an AME impressions collector 618 when the client device 104 accesses the media

606. For example, a web browser and/or app of the client device 104 executes the beacon instructions 608 in the media 606 which instruct the browser and/or app to generate and send the beacon/impression request 612. In the illustrated example, the client device 104 sends the beacon/impression request 612 to the AME impression collector 618 using an HTTP (hypertext transfer protocol) request addressed to the URL (uniform resource locator) of the AME impressions collector 618 at, for example, a first internet domain of the AME 108. The beacon/impression request 612 of the illustrated example includes a media identifier 613 (e.g., an identifier that can be used to identify content, an advertisement, and/or any other media) corresponding to the media 606. In some examples, the beacon/impression request 612 also includes a site identifier (e.g., a URL) of the website that served the media 606 to the client device 104 and/or a host website ID (e.g., www.acme.com) of the website that displays or presents the media 606. In the illustrated example, the beacon/impression request 612 includes a device/user identifier 114. In the illustrated example, the device/user identifier 114 that the client device 104 provides in the beacon impression request 612 is an AME ID because it corresponds to an identifier that the AME 108 uses to identify a panelist corresponding to the client device 104. In other examples, the client device 104 may not send the device/user identifier 114 until the client device 104 receives a request for the same from a server of the AME 108 (e.g., in response to, for example, the AME impressions collector 618 receiving the beacon/impression request 612).

[0065] In some examples, the device/user identifier 114 may be a device identifier (e.g., an international mobile equipment identity (IMEI), a mobile equipment identifier (MEID), a media access control (MAC) address, etc.), a web browser unique identifier (e.g., a cookie), a user identifier (e.g., a user name, a login ID, etc.), an Adobe Flash® client identifier, identification information stored in an HTML5 datastore, a vehicle identification number (VIN) and/or any other identifier that the AME 108 stores in association with demographic information about users of the client devices 104. When the AME 108 receives the device/user identifier 114, the AME 108 can obtain

demographic information corresponding to a user of the client device 104 based on the device/user identifier 114 that the AME 108 receives from the client device 104. In some examples, the device/user identifier 114 may be encrypted (e.g., hashed) at the client device 104 so that only an intended final recipient of the device/user identifier 114 can decrypt the hashed identifier 114. For example, if the device/user identifier 114 is a cookie that is set in the client device 104 by the AME 108, the device/user identifier 114 can be hashed so that only the AME 108 can decrypt the device/user identifier 114. If the device/user identifier 114 is an IMEI number, the client device 104 can hash the device/user identifier 114 so that only a wireless carrier (e.g., the database proprietor 110) can decrypt the hashed identifier 114 to recover the IMEI for use in accessing demographic information corresponding to the user of the client device 104. By hashing the device/user identifier 114, an intermediate party (e.g., an intermediate server or entity on the Internet) receiving the beacon request cannot directly identify a user of the client device 104.

[0066] In response to receiving the beacon/impression request 612, the AME impressions collector 618 logs an impression for the media 606 by storing the media identifier 613 contained in the beacon/impression request 612. In the illustrated example of FIG. 6, the AME impressions collector 618 also uses the device/user identifier 114 in the beacon/impression request 612 to identify AME panelist demographic information corresponding to a panelist of the client device 104. That is, the device/user identifier 114 matches a user ID of a panelist member (e.g., a panelist corresponding to a panelist profile maintained and/or stored by the AME 108). In this manner, the AME impressions collector 618 can associate the logged impression with demographic information of a panelist corresponding to the client device 104.

[0067] In some examples, the beacon/impression request 612 may not include the device/user identifier 114 if, for example, the user of the client device 104 is not an AME panelist. In such examples, the AME impressions collector 618 logs impressions regardless of whether the client device 104 provides the device/user identifier 114 in the beacon/impression request 612

(or in response to a request for the identifier 114). When the client device 104 does not provide the device/user identifier 114, the AME impressions collector 618 will still benefit from logging an impression for the media 606 even though it will not have corresponding demographics. For example, the AME 108 may still use the logged impression to generate a total impressions count and/or a frequency of impressions (e.g., an impressions frequency) for the media 606. Additionally or alternatively, the AME 108 may obtain demographics information from the database proprietor 110 for the logged impression if the client device 104 corresponds to a subscriber of the database proprietor 110.

[0068] In the illustrated example of FIG. 6 to compare or supplement panelist demographics (e.g., for accuracy or completeness) of the AME 108 with demographics from one or more database proprietors (e.g., the database proprietor 110), the AME impressions collector 618 returns a beacon response message 622 (e.g., a first beacon response) to the client device 104 including an HTTP “302 Found” re-direct message and a URL of a participating database proprietor 110 at, for example, a second internet domain. In the illustrated example, the HTTP “302 Found” re-direct message in the beacon response 622 instructs the client device 104 to send a second beacon request 626 to the database proprietor 110. In other examples, instead of using an HTTP “302 Found” re-direct message, redirects may be implemented using, for example, an iframe source instruction (e.g., <iframe src = “ ”>) or any other instruction that can instruct a client device to send a subsequent beacon request (e.g., the second beacon request 626) to a participating database proprietor 110. In the illustrated example, the AME impressions collector 618 determines the database proprietor 110 specified in the beacon response 622 using a rule and/or any other suitable type of selection criteria or process. In some examples, the AME impressions collector 618 determines a particular database proprietor to which to redirect a beacon request based on, for example, empirical data indicative of which database proprietor is most likely to have demographic data for a user corresponding to the device/user identifier 114. In some examples, the beacon instructions 608 include a predefined URL

of one or more database proprietors to which the client device 104 should send follow up beacon requests 626. In other examples, the same database proprietor is always identified in the first redirect message (e.g., the beacon response 622).

[0069] In the illustrated example of FIG. 6, the beacon/impression request 626 may include a device/user identifier 114 that is a database proprietor ID because it is used by the database proprietor 114 to identify a subscriber of the client device 104 when logging an impression. In some instances (e.g., in which the database proprietor 110 has not yet set a database proprietor ID in the client device 104), the beacon/impression request 626 does not include the device/user identifier 114. In some examples, the database proprietor ID is not sent until the database proprietor 110 requests the same (e.g., in response to the beacon/impression request 626). When the database proprietor 110 receives the device/user identifier 114, the database proprietor 110 can obtain demographic information corresponding to a user of the client device 104 based on the device/user identifier 114 that the database proprietor 110 receives from the client device 104. In some examples, the device/user identifier 114 may be encrypted (e.g., hashed) at the client device 104 so that only an intended final recipient of the device/user identifier 114 can decrypt the hashed identifier 114. For example, if the device/user identifier 114 is a cookie that is set in the client device 104 by the database proprietor 110, the device/user identifier 114 can be hashed so that only the database proprietor 110 can decrypt the device/user identifier 114. If the device/user identifier 114 is an IMEI number, the client device 104 can hash the device/user identifier 114 so that only a wireless carrier (e.g., the database proprietor 110) can decrypt the hashed identifier 114 to recover the IMEI for use in accessing demographic information corresponding to the user of the client device 104. By hashing the device/user identifier 114, an intermediate party (e.g., an intermediate server or entity on the Internet) receiving the beacon request cannot directly identify a user of the client device 104. For example, if the intended final recipient of the device/user identifier 114 is the database proprietor 110, the AME 108 cannot recover identifier information

when the device/user identifier 114 is hashed by the client device 104 for decrypting only by the intended database proprietor 110.

[0070] In some examples that use cookies as the device/user identifier 114, when a user deletes a database proprietor cookie from the client device 104, the database proprietor 110 sets the same cookie value in the client device 104 the next time the user logs into a service of the database proprietor 110. In such examples, the cookies used by the database proprietor 110 are registration-based cookies, which facilitate setting the same cookie value after a deletion of the cookie value has occurred on the client device 104. In this manner, the database proprietor 110 can collect impressions for the client device 104 based on the same cookie value over time to generate unique audience (UA) sizes while eliminating or substantially reducing the likelihood that a single unique person will be counted as two or more separate unique audience members.

[0071] Although only a single database proprietor 110 is shown in FIG. 6, the impression reporting/collection process of FIG. 6 may be implemented using multiple database proprietors 110. In some such examples, the beacon instructions 608 cause the client device 104 to send beacon/impression requests 626 to numerous database proprietors 110. For example, the beacon instructions 608 may cause the client device 104 to send the beacon/impression requests 626 to the numerous database proprietors 110 in parallel or in daisy chain fashion. In some such examples, the beacon instructions 608 cause the client device 104 to stop sending beacon/impression requests 626 to database proprietors once a database proprietor has recognized the client device 104. In other examples, the beacon instructions 608 cause the client device 104 to send beacon/impression requests 626 to database proprietors 110 so that multiple database proprietors 110 can recognize the client device 104 and log a corresponding impression. In any case, multiple database proprietors 110 are provided the opportunity to log impressions and provide corresponding demographics information if the user of the client device 104 is a subscriber of services of those database proprietors.

[0072] In some examples, prior to sending the beacon response 622 to the client device 101, the AME impressions collector 618 replaces site IDs (e.g., URLs) of media provider(s) that served the media 606 with modified site IDs (e.g., substitute site IDs) which are discernable only by the AME 108 to identify the media provider(s). In some examples, the AME impressions collector 618 may also replace a host website ID (e.g., www.acme.com) with a modified host site ID (e.g., a substitute host site ID) which is discernable only by the AME 108 as corresponding to the host website via which the media 606 is presented. In some examples, the AME impressions collector 618 also replaces the media identifier 613 with a modified media identifier 613 corresponding to the media 606. In this way, the media provider of the media 606, the host website that presents the media 606, and/or the media identifier 613 are obscured from the database proprietor 110, but the database proprietor 110 can still log impressions based on the modified values which can later be deciphered by the AME 108 after the AME 108 receives logged impressions from the database proprietor 110. In some examples, the AME impressions collector 618 does not send site IDs, host site IDs, the media identifier 613 or modified versions thereof in the beacon response 622. In such examples, the client device 104 provides the original, non-modified versions of the media identifier 613, site IDs, host IDs, etc. to the database proprietor 110.

[0073] In the illustrated example, the AME impression collector 618 maintains a modified ID mapping table 628 that maps original site IDs with modified (or substitute) site IDs, original host site IDs with modified host site IDs, and/or maps modified media identifiers to the media identifiers such as the media identifier 613 to obfuscate or hide such information from database proprietors such as the database proprietor 110. Also in the illustrated example, the AME impressions collector 618 encrypts all of the information received in the beacon/impression request 612 and the modified information to prevent any intercepting parties from decoding the information. The AME impressions collector 618 of the illustrated example sends the encrypted information in the beacon response 622 to the client device 104 so that the client device 104 can send the encrypted information to the database

proprietor 110 in the beacon/impression request 626. In the illustrated example, the AME impressions collector 618 uses an encryption that can be decrypted by the database proprietor 110 site specified in the HTTP “302 Found” re-direct message.

[0074] From time to time, the impression data collected by the database proprietor 110 is provided to a database proprietor impressions collector 630 of the AME 108 as, for example, batch (e.g., aggregate) data. As discussed above, some impressions logged by the client device 104 to the database proprietor 110 are misattributed by the database proprietor 110 to a wrong subscriber and, thus, to incorrect demographic information. During a data collecting and merging process to combine demographic and impression data from the AME 108 and the database proprietor 110, demographics of impressions logged by the AME 108 for the client device 104 will not correspond to demographics of impressions logged by the database proprietor 110 because the database proprietor 110 has misattributed some impressions to the incorrect demographic information. Examples disclosed herein may be used to determine corrected age-based demographic groups of impression data provided by the database proprietor 110.

[0075] Additional examples that may be used to implement the beacon instruction processes of FIG. 6 are disclosed in U.S. Patent No. 8,370,489, entitled “Methods and Apparatus to Determine Impressions Using Distributed Demographic Information,” which is hereby incorporated herein by reference in its entirety. In addition, other examples that may be used to implement such beacon instructions are disclosed in U.S. Patent No. 6,108,637, entitled “Content Display Monitor,” which is hereby incorporated herein by reference in its entirety.

[0076] In some examples, the database proprietor 110, before providing aggregated demographics to the AME 108, uses the age corrector 124 to correct age-based demographic group misattributions in the impression data. In some examples, the aggregate impressions data includes an aggregate age PDF. In some examples, when the AME 108 receives anonymized user-level impression data and/or demographic data from the database proprietor

110, the AME 108 uses the age corrector 124 to correct age-based demographic group misattributions in the impression data.

[0077] A flowchart representative of example machine readable instructions for implementing the age corrector 124 of FIG. 3 is shown in FIG. 7. A flowchart representative of example machine readable instructions for implementing the age updater 310 of FIG. 3 is shown in FIG. 8. A flowchart representative of example machine readable instructions for implementing the demographic impression aggregator 126 of FIG. 1 is shown in FIG. 9. In this example, the machine readable instructions comprise a program for execution by a processor such as the processor 1012 shown in the example processor platform 1000 discussed below in connection with FIG. 10. The program may be embodied in software stored on a tangible computer readable storage medium such as a CD-ROM, a floppy disk, a hard drive, a digital versatile disk (DVD), a Blu-ray disk, or a memory associated with the processor 1012, but the entire program and/or parts thereof could alternatively be executed by a device other than the processor 1012 and/or embodied in firmware or dedicated hardware. Further, although the example programs are described with reference to the flowcharts illustrated in FIGS. 7, 8, and/or 9, many other methods of implementing the example age corrector 124, the example age updater 310, and/or the example demographic impression aggregator 126 may alternatively be used. For example, the order of execution of the blocks may be changed, and/or some of the blocks described may be changed, eliminated, or combined.

[0078] As mentioned above, the example processes of FIGS. 7, 8, and/or 9 may be implemented using coded instructions (e.g., computer and/or machine readable instructions) stored on a tangible computer readable storage medium such as a hard disk drive, a flash memory, a read-only memory (ROM), a compact disk (CD), a digital versatile disk (DVD), a cache, a random-access memory (RAM) and/or any other storage device or storage disk in which information is stored for any duration (e.g., for extended time periods, permanently, for brief instances, for temporarily buffering, and/or for caching of the information). As used herein, the term tangible computer

readable storage medium is expressly defined to include any type of computer readable storage device and/or storage disk and to exclude propagating signals and to exclude transmission media. As used herein, "tangible computer readable storage medium" and "tangible machine readable storage medium" are used interchangeably. Additionally or alternatively, the example processes of FIGS. 7, 8, and/or 9 may be implemented using coded instructions (e.g., computer and/or machine readable instructions) stored on a non-transitory computer and/or machine readable medium such as a hard disk drive, a flash memory, a read-only memory, a compact disk, a digital versatile disk, a cache, a random-access memory and/or any other storage device or storage disk in which information is stored for any duration (e.g., for extended time periods, permanently, for brief instances, for temporarily buffering, and/or for caching of the information). As used herein, the term non-transitory computer readable medium is expressly defined to include any type of computer readable storage device and/or storage disk and to exclude propagating signals and to exclude transmission media. As used herein, when the phrase "at least" is used as the transition term in a preamble of a claim, it is open-ended in the same manner as the term "comprising" is open ended.

[0079] FIG. 7 is a flow diagram of example machine readable instructions 700 that may be executed to implement the data corrector 124 of FIGS. 1, 2, and/or 5 to correct ages associated with demographic impressions associated with subscribers of a database proprietor 110. Initially, at block 702, the age predictor 304 (FIG. 3) obtains subscriber data (e.g., the subscriber data 136 of FIGS. 1, 2, and 3). The example subscriber data 136 includes an identifier (e.g., UID 130 of FIGS. 1, 2, and 3) and subscriber activity metrics (e.g., the subscriber activity metrics 138 of FIGS. 1, 2, and 3) for one or more subscribers of the database proprietor 110. In some examples, the subscriber data 136 is obtained in response to an age prediction model 306 (FIG. 3) being generated so that the age predictor 310 can make age predictions based on the subscriber data 136 soon after the age prediction model 306 is generated. In some examples, the age predictor 304 retrieves subscriber data 136 for a subset of the subscribers of the database proprietor 110. For example, the age

predictor 304 may retrieve subscriber data 136 of subscribers that have subscribed to the database proprietor 110 since the age prediction model 306 was last generated.

[0080] At block 704, the age predictor 304 obtains an age PDF (e.g., the age PDF 305 of FIG. 3) from the prediction model 306 using the subscriber activity metrics 138. In some examples, using the subscriber activity metrics 138 as inputs into the age prediction model 306, the intermediate nodes of the age prediction model 306 are traversed until the age prediction model 306 reaches a terminal node. In such an example, the terminal node identifies an age PDF (e.g., the age PDF 305 of FIGS. 3 and 4) and/or an age PDF ID (e.g., the age PDF ID 404 of FIGS. 4 and 5). In some examples, the age PDF 305 is identified by the age PDF ID. At block 706, the age predictor 304 stores the predicted age PDF 305 and/or the age PDF ID 404, the UID 130 included in the subscriber data 138, and the date the prediction was made (e.g., the prediction date 502 of FIG. 5) in the age cache 308. At block 708, the age predictor 304 determines whether there is another subscriber (e.g., more subscriber data 136) for which to obtain a predicted age PDF. If there is another subscriber, program control returns to block 704. Otherwise, if there is not another subscriber, program control advances to block 710.

[0081] At block 710, the age updater 310 waits until an age request (e.g., the age request 128 of FIGS. 1 and 2) has been received. For example, the age updater 310 may receive the age request 128 from the impression handler 120 (FIG. 1) or the age adjuster 206 (FIG. 2) to determine the age-corrected PDF 134 at a later time (e.g., days, weeks, months, etc.) after the age predictor 304 determines an initial predicted age PDF 305 represented in the age cache 308. If an age request 128 has been received, program control advances to block 712. At block 712, the age updater 310 retrieves a predicted age PDF 305 and/or an age PDF ID 404 and a prediction date 502 from the age cache 308 based on a UID 130 included in the age request 128. At block 714, the age updater 310 calculates an age-corrected PDF 134 (FIGS. 1, 2 and 3). An example process for implementing the operation of block 714 is

described below in connection with FIG. 8. At block 716, the age updater 310 determines if there is another age request 128. If there is another age request 128, program control returns to block 710. Otherwise, if there is not another age request 128, the example program 700 ends.

[0082] FIG. 8 is a flow diagram of example machine readable instructions 714 that may be executed to implement the age updater 310 of FIG. 3 to calculate an age-corrected PDF 134 (FIGS. 1, 2 and 3) based on a predicted age PDF with a corresponding prediction date 502 (FIG. 5) and an age request 128 (FIGS. 1, 2, and 3) that includes an impression date 302. The example process of FIG. 8 may be used to implement clock 714 of FIG. 7. Initially, at block 800, the age updater 310 selects a demographic bucket from the predicted age PDF for which to calculate an aging factor. For example, if the predicted age PDF (e.g., the predicted age PDFs 305a-305c of FIG. 4) has a 2-11 year-old demographic bucket, a 12-14 year-old demographic bucket, a 15-17 year-old demographic bucket, an 18-20 year-old demographic bucket, a 21-24 year-old demographic bucket, a 25-29 year-old demographic bucket, a 30-34 year-old demographic bucket, a 35-39 year-old demographic bucket, a 40-44 year-old demographic bucket, a 45-49 year-old demographic bucket, a 50-54 year-old demographic bucket, a 55-64 year-old demographic bucket, and a 65+ year-old demographic bucket, the age updater 310 may select the 2-11 year-old demographic bucket. At block 802, the age updater 310 calculates an aging factor for the selected demographic bucket. In some examples, the aging factors are calculated in accordance with Equation 2 above. For example, if the impression date 302 is September 25, 2014, the prediction date 502 is March 2, 2014, and the span of the selected demographic bucket is five years (e.g., an ages 2-6 year-old demographic bucket, an ages 7-11 year-old demographic bucket, etc.), the aging factor is 0.114 (208/1826) because 208 is the number of days between September 25, 2014, and March 2, 2014, and 1826 is the number of days in five years.

[0083] At block 804, the age updater 310 determines whether another aging factor is to be calculated for another demographic bucket. For example, another aging factor is to be calculated if one or more demographic buckets of

a predicted age PDF have not had corresponding aging factors calculated. If another aging factor is to be calculated for another demographic bucket, program control returns to block 800. Otherwise, if another aging factor is not to be calculated for another demographic bucket, program control advances to block 806. At block 806, the age updater 310 applies the aging factor(s) calculated at block 802 to the predicted age PDF. For example, the age updater 310 applies aging factors to the predicted age PDF to generate the age-corrected PDF 134 (FIGS. 1, 2, and 3). In some examples, the aging factors are applied to the predicted age PDF in accordance with Equation 3 and Equation 4 above. Example program 714 then ends.

[0084] FIG. 9 is a flow diagram of example machine readable instructions 900 that may be executed to implement the demographic impression aggregator 126 of FIG. 1 to calculate aggregated age-corrected PDFs to include with aggregate demographic impressions 106 (FIG. 1). The example process of FIG. 9 may be used when aggregate demographic impressions 106 are to be sent to the AME 108. Initially, at block 902, the demographic impression aggregator 126 selects characteristics (e.g., gender, marital status, media, device type, location, etc.) to include and/or exclude to generate aggregate demographic impressions 106. For example, the demographic impression aggregator 126 may select “males, ‘Ricky and Morty’, New England” (e.g., a gender, a media presentation, and a geographical region) as the characteristics to use to aggregate the demographic impressions. In some examples, the selected characteristics are based on agreements between the AME 108 (FIG. 1) and the database proprietor 110 (FIG. 1). In some examples, the selected characteristics are based on requests for aggregate demographic data 106 by the AME 108. At block 904, the demographic impression aggregator 126 retrieves the demographic impressions corresponding to the characteristics selected at block 902 from the impression database 116.

[0085] At block 906, the demographic impression aggregator 126 calculates an aggregate age-corrected PDF based on the age-corrected PDFs corresponding to the demographic impressions retrieved at block 904. In some

examples, to calculate an aggregate age-corrected PDF, the demographic impression aggregator 126 averages the probabilities for each demographic bucket. For example, if the probabilities associated with the 18-20 year-old demographic bucket for predicted age PDFs to be aggregated are 13%, 20%, 5% and 10%, the aggregate probability for the 18-20 year-old demographic bucket would be 12% $((13\%+20\%+5\%+10\%)/4)$. In some examples, the aggregate age-corrected PDF is calculated in accordance with Equation 1 above. At block 908, the demographic impression aggregator 126 generates the aggregate demographic impression data 106. In some examples, the demographic impression aggregator 126 generates the aggregate demographic impression data 106 by associating the aggregate age-corrected PDF calculated at block 906 with a count of the demographic impressions retrieved at block 904. Example program 900 then ends.

[0086] FIG. 10 is a block diagram of an example processor platform 1100 structured to execute the instructions of FIGS 7, 8, and/or 9 to implement the age corrector 124, the impression handler 120, the demographic impression aggregator 126, the impressions collectors 206, the age adjuster 206, the age predictor 304, the age prediction model 306, the age cache 308, and/or the age updater 310 of FIGS. 1, 2, and 3. By way of example, FIG. 10 shows the age predictor 304 and the age updater 310, but processor platform 1000 may include more of, fewer of, or different ones of the age corrector 124, the impression handler 120, the demographic impression aggregator 126, the impressions collectors 206, the age adjuster 206, the age predictor 304, the age prediction model 306, the age cache 308, and/or the age updater 310 of FIGS. 1, 2, and 3. The processor platform 1000 can be, for example, a server, a personal computer, a workstation, or any other type of computing device.

[0087] The processor platform 1000 of the illustrated example includes a processor 1012. The processor 1012 of the illustrated example is hardware. For example, the processor 1012 can be implemented by one or more integrated circuits, logic circuits, microprocessors or controllers from any desired family or manufacturer.

[0088] The processor 1012 of the illustrated example includes a local memory 1013 (e.g., a cache). The processor 1012 of the illustrated example is in communication with a main memory including a volatile memory 1014 and a non-volatile memory 1016 via a bus 1018. The volatile memory 1014 may be implemented by Synchronous Dynamic Random Access Memory (SDRAM), Dynamic Random Access Memory (DRAM), RAMBUS Dynamic Random Access Memory (RDRAM) and/or any other type of random access memory device. The non-volatile memory 1016 may be implemented by flash memory and/or any other desired type of memory device. Access to the main memory 1014, 1016 is controlled by a memory controller.

[0089] The processor platform 1000 of the illustrated example also includes an interface circuit 1020. The interface circuit 1020 may be implemented by any type of interface standard, such as an Ethernet interface, a universal serial bus (USB), and/or a PCI express interface.

[0090] In the illustrated example, one or more input devices 1022 are connected to the interface circuit 1020. The input device(s) 1022 permit(s) a user to enter data and commands into the processor 1012. The input device(s) can be implemented by, for example, an audio sensor, a microphone, a camera (still or video), a keyboard, a button, a mouse, a touchscreen, a track-pad, a trackball, isopoint and/or a voice recognition system.

[0091] One or more output devices 1024 are also connected to the interface circuit 1020 of the illustrated example. The output devices 1024 can be implemented, for example, by display devices (e.g., a light emitting diode (LED), an organic light emitting diode (OLED), a liquid crystal display, a cathode ray tube display (CRT), a touchscreen, a tactile output device, a printer and/or speakers). The interface circuit 1020 of the illustrated example, thus, typically includes a graphics driver card, a graphics driver chip or a graphics driver processor.

[0092] The interface circuit 1020 of the illustrated example also includes a communication device such as a transmitter, a receiver, a transceiver, a modem and/or network interface card to facilitate exchange of data with external machines (e.g., computing devices of any kind) via a

network 1026 (e.g., an Ethernet connection, a digital subscriber line (DSL), a telephone line, coaxial cable, a cellular telephone system, etc.).

[0093] The processor platform 1000 of the illustrated example also includes one or more mass storage devices 1028 for storing software and/or data. Examples of such mass storage devices 1028 include floppy disk drives, hard drive disks, compact disk drives, Blu-ray disk drives, RAID systems, and digital versatile disk (DVD) drives.

[0094] Coded instructions 1032 to implement the machine readable instructions of FIGS. 7, 8, and/or 9 may be stored in the mass storage device 1028, in the volatile memory 1014, in the non-volatile memory 1016, and/or on a removable tangible computer readable storage medium such as a CD or DVD.

[0095] FIG. 11 depicts an example system 1100 to collect user information (e.g., user information 1102a, 1102b) from distributed database proprietors 110a, 110b for associating with impressions of media presented at a client device 104. Examples disclosed herein to predict ages may be used in connection with demographic impressions collected using example techniques described below in connection with FIG. 11.

[0096] In the illustrated examples, user information 1102a, 1102b or user data includes one or more of demographic data, purchase data, and/or other data indicative of user activities, behaviors, and/or preferences related to information accessed via the Internet, purchases, media accessed on electronic devices, physical locations (e.g., retail or commercial establishments, restaurants, venues, etc.) visited by users, etc. In some examples, user information 1102a, 1102b is stored in the subscriber accounts database 122 of FIG. 1. Examples of FIG. 11 are described in connection with a mobile device, which may be a mobile phone, a mobile communication device, a tablet, a gaming device, a portable media presentation device, an in-vehicle or vehicle-integrated communication system, such as an automobile infotainment system with wireless communication capabilities, etc. However, examples disclosed herein may be implemented in connection with impressions corresponding to mobile and/or non-mobile devices. Example non-mobile devices include

internet appliances, smart televisions, internet terminals, computers, or any other device capable of presenting media received via network communications.

[0097] In the illustrated example of FIG. 11, to track media impressions on the client device 104, the AME 108 partners with or cooperates with an app publisher 1110 to download and install a data collector 1112 on the client device 104. The app publisher 1110 of the illustrated example may be a software app developer that develops and distributes apps to mobile devices and/or a distributor that receives apps from software app developers and distributes the apps to mobile devices. The data collector 1112 may be included in other software loaded onto the client device 104, such as the operating system 1114, an application (or app) 1116, a web browser 1117, and/or any other software. In some examples, the example client device 104 of FIG. 11 is a non-locally metered device. For example, the client device 104 of a non-panelist household does not support and/or has not been provided with specific metering software (e.g., dedicated metering software provided directly by the AME 108 and executing as a foreground or background process for the sole purpose of monitoring media accesses/exposure).

[0098] Any of the example software 1114-1117 may present media 1118 received from a media publisher 1120. The media 1118 may be an advertisement, video, audio, text, a graphic, a web page, news, educational media, entertainment media, or any other type of media. In the illustrated example, a media ID 1122 is provided in the media 1118 to enable identifying the media 1118 so that the AME 108 can credit the media 1118 with media impressions when the media 1118 is presented on the client device 104 or any other device that is monitored by the AME 108.

[0099] The data collector 1112 of the illustrated example includes instructions (e.g., Java, java script, or any other computer language or script) that, when executed by the client device 104, cause the client device 104 to collect the media ID 1122 of the media 1118 presented by the app program 1116 and/or the client device 104, and to collect one or more device/user identifier(s) 114 stored in the client device 104. The device/user identifier(s)

114 of the illustrated example include identifiers that can be used by corresponding ones of the partner database proprietors 110a, 110b to identify the user or users of the client device 104, and to locate user information 1102a, 1102b corresponding to the user(s). For example, the device/user identifier(s) 114 used in connection with examples of FIG. 11 may include hardware identifiers (e.g., an international mobile equipment identity (IMEI), a mobile equipment identifier (MEID), a media access control (MAC) address, etc.), an app store identifier (e.g., a Google Android ID, an Apple ID, an Amazon ID, etc.), an open source unique device identifier (OpenUDID), an open device identification number (ODIN), a login identifier (e.g., a username), an email address, user agent data (e.g., application type, operating system, software vendor, software revision, etc.), third-party service identifiers (e.g., advertising service identifiers, device usage analytics service identifiers, demographics collection service identifiers), web storage data, document object model (DOM) storage data, local shared objects (also referred to as “Flash cookies”), an automobile vehicle identification number (VIN), etc. In some examples, fewer or more device/user identifier(s) 114 may be used. In addition, although only two partner database proprietors 110a, 110b are shown in FIG. 11, the AME 108 may partner with any number of partner database proprietors to collect distributed user information (e.g., the user information 1102a, 1102b).

[0100] In some examples, the client device 104 may not allow access to identification information stored in the client device 104. For such instances, the disclosed examples enable the AME 108 to store an AME-provided identifier (e.g., an identifier managed and tracked by the AME 108) in the client device 104 to track media impressions on the client device 104. For example, the AME 108 may provide instructions in the data collector 1112 to set an AME-provided identifier in memory space accessible by and/or allocated to the app program 1116. The data collector 1112 uses the identifier as a device/user identifier 114. In such examples, the AME-provided identifier set by the data collector 1112 persists in the memory space even when the app program 1116 and the data collector 1112 are not running. In this manner, the same AME-provided identifier can remain associated with the client device

104 for extended durations and from app to app. In some examples in which the data collector 1112 sets an identifier in the client device 104, the AME 108 may recruit a user of the client device 104 as a panelist, and may store user information collected from the user during a panelist registration process and/or collected by monitoring user activities/behavior via the client device 104 and/or any other device used by the user and monitored by the AME 108. In this manner, the AME 108 can associate user information of the user (from panelist data stored by the AME 108) with media impressions attributed to the user on the client device 104.

[0101] In the illustrated example, the data collector 1112 sends the media ID 1122 and the one or more device/user identifier(s) 114 as collected data 1126 to the app publisher 1110. Alternatively, the data collector 1112 may be configured to send the collected data 1126 to another collection entity (other than the app publisher 1110) that has been contracted by the AME 108 or is partnered with the AME 108 to collect media ID's (e.g., the media ID 1122) and device/user identifiers (e.g., the device/user identifier(s) 114) from mobile devices (e.g., the client device 104). In the illustrated example, the app publisher 1110 (or a collection entity) sends the media ID 1122 and the device/user identifier(s) 114 as impression data 102 to a server 1132 at the AME 108. The impression data 102 of the illustrated example may include one media ID 1122 and one or more device/user identifier(s) 114 to report a single impression of the media 1118, or it may include numerous media ID's 1122 and device/user identifier(s) 114 based on numerous instances of collected data (e.g., the collected data 1126) received from the client device 104 and/or other mobile devices to report multiple impressions of media.

[0102] In the illustrated example, the server 1132 stores the impression data 102 in an AME media impressions store 1134 (e.g., a database or other data structure). Subsequently, the AME 108 sends the device/user identifier(s) 114 to corresponding partner database proprietors (e.g., the partner database proprietors 110a, 110b) to receive user information (e.g., the user information 1102a, 1102b) corresponding to the device/user identifier(s) 114 from the partner database proprietors 110a, 110b so that the AME 108 can associate the

user information with corresponding media impressions of media (e.g., the media 1118) presented at mobile devices (e.g., the client device 104).

[0103] In some examples, to protect the privacy of the user of the client device 104, the media identifier 1122 and/or the device/user identifier(s) 114 are encrypted before they are sent to the AME 108 and/or to the partner database proprietors 110a, 110b. In other examples, the media identifier 1122 and/or the device/user identifier(s) 114 are not encrypted.

[0104] After the AME 108 receives the device/user identifier(s) 114, the AME 108 sends device/user identifier logs 1136a, 1136b to corresponding partner database proprietors (e.g., the partner database proprietors 110a, 110b). In some examples, each of the device/user identifier logs 1136a, 1136b includes a single device/user identifier. In some examples, some or all of the device/user identifier logs 1136a, 1136b include numerous aggregate device/user identifiers received at the AME 108 over time from one or more mobile devices. After receiving the device/user identifier logs 1136a, 1136b, each of the partner database proprietors 110a, 110b looks up its users corresponding to the device/user identifiers 124 in the respective logs 1136a-b. In this manner, each of the partner database proprietors 110a-b collects user information 1102a, 1102b corresponding to users identified in the device/user identifier logs 1136a, 1136b for sending to the AME 108. For example, if the partner database proprietor 110a is a wireless service provider and the device/user identifier log 1136a includes IMEI numbers recognizable by the wireless service provider, the wireless service provider accesses its subscriber records to find users having IMEI numbers matching the IMEI numbers received in the device/user identifier log 1136a. When the users are identified, the wireless service provider copies the users' user information to the user information 1102a for delivery to the AME 108.

[0105] In some other examples, the example data collector 1112 sends the device/user identifier(s) 114 from the client device 104 to the app publisher 1110 in the collected data 1126, and it also sends the device/user identifier(s) 114 to the media publisher 1120. In such other examples, the data collector 1112 does not collect the media ID 1122 from the media 1118 at the

client device 104 as the data collector 1112 does in the example system 1100 of FIG. 11. Instead, the media publisher 1120 that publishes the media 1118 to the client device 104 retrieves the media ID 1122 from the media 1118 that it publishes. The media publisher 1120 then associates the media ID 1122 to the device/user identifier(s) 114 received from the data collector 1112 executing in the client device 104, and sends collected data 138 to the app publisher 110 that includes the media ID 1122 and the associated device/user identifier(s) 114 of the client device 104. For example, when the media publisher 1120 sends the media 1118 to the client device 104, it does so by identifying the client device 104 as a destination device for the media 1118 using one or more of the device/user identifier(s) 114 received from the client device 104. In this manner, the media publisher 1120 can associate the media ID 1122 of the media 1118 with the device/user identifier(s) 114 of the client device 104 indicating that the media 1118 was sent to the particular client device 104 for presentation (e.g., to generate an impression of the media 1118).

[0106] Alternatively, in some other examples in which the data collector 1112 is configured to send the device/user identifier(s) 114 to the media publisher 1120, and the data collector 1112 does not collect the media ID 1122 from the media 1118 at the client device 104, the media publisher 1120 sends impression data 102 to the AME 108. For example, the media publisher 1120 that publishes the media 1118 to the client device 104 also retrieves the media ID 1122 from the media 1118 that it publishes, and associates the media ID 1122 with the device/user identifier(s) 114 of the client device 104. The media publisher 1120 then sends the media impression data 102, including the media ID 1122 and the device/user identifier(s) 114, to the AME 108. For example, when the media publisher 1120 sends the media 1118 to the client device 104, it does so by identifying the client device 104 as a destination device for the media 1118 using one or more of the device/user identifier(s) 114. In this manner, the media publisher 1120 can associate the media ID 1122 of the media 1118 with the device/user identifier(s) 114 of the client device 104 indicating that the media 1118 was sent to the particular client device 104 for presentation (e.g., to generate an impression of the media

118). In the illustrated example, after the AME 108 receives the impression data 102 from the media publisher 1120, the AME 108 can then send the device/user identifier logs 1136a, 1136b to the partner database proprietors 110a, 110b to request the user information 1102a, 1102b as described above.

[0107] Although the media publisher 1120 is shown separate from the app publisher 1110 in FIG. 11, the app publisher 1110 may implement at least some of the operations of the media publisher 1120 to send the media 1118 to the client device 104 for presentation. For example, advertisement providers, media providers, or other information providers may send media (e.g., the media 1118) to the app publisher 1110 for publishing to the client device 104 via, for example, the app program 1116 when it is executing on the client device 104. In such examples, the app publisher 1110 implements the operations described above as being performed by the media publisher 1120.

[0108] Additionally or alternatively, in contrast with the examples described above in which the client device 104 sends identifiers to the AME 108 (e.g., via the application publisher 1110, the media publisher 1120, and/or another entity), in other examples the client device 104 (e.g., the data collector 1112 installed on the client device 104) sends the identifiers (e.g., the user/device identifier(s) 114) directly to the respective database proprietors 110a, 110b (e.g., not via the AME 108). In such examples, the example client device 104 sends the media identifier 1122 to the AME 108 (e.g., directly or through an intermediary such as via the application publisher 1110), but does not send the media identifier 1122 to the database proprietors 110a, 110b.

[0109] As mentioned above, the example partner database proprietors 110a, 110b provide the user information 1102a, 1102b to the example AME 108 for matching with the media identifier 1122 to form media impression information. As also mentioned above, the database proprietors 110a, 110b are not provided copies of the media identifier 1122. Instead, the client device 104 provides the database proprietors 110a, 110b with impression identifiers 1140. An impression identifier 1140 uniquely identifies an impression event relative to other impression events of the client device 104 so that an occurrence of an impression at the client device 104 can be distinguished from other

occurrences of impressions. However, the impression identifier 1140 does not itself identify the media associated with that impression event. In such examples, the impression data 102 from the client device 104 to the AME 108 also includes the impression identifier 1140 and the corresponding media identifier 1122. To match the user information 1102a, 1102b with the media identifier 1122, the example partner database proprietors 110a, 110b provide the user information 1102a, 1102b to the AME 108 in association with the impression identifier 1140 for the impression event that triggered the collection of the user information 1102a, 1102b. In this manner, the AME 108 can match the impression identifier 1140 received from the client device 104 via the impression data 102 to a corresponding impression identifier 1140 received from the partner database proprietors 110a, 110b via the user information 1102a, 1102b to associate the media identifier 1122 received from the client device 106 with demographic information in the user information 1102a, 1102b received from the database proprietors 110a, 110b.

[0110] The impression identifier 1140 of the illustrated example is structured to reduce or avoid duplication of audience member counts for audience size measures. For example, the example partner database proprietors 110a, 110b provide the user information 1102a, 1102b and the impression identifier 1140 to the AME 108 on a per-impression basis (e.g., each time a client device 104 sends a request including an encrypted identifier and an impression identifier 1140 to the partner database proprietor 110a, 110b) and/or on an aggregated basis. When aggregate impression data is provided in the user information 1102a, 1102b, the user information 1102a, 1102b includes indications of multiple impressions (e.g., multiple impression identifiers 1140) at mobile devices. In some examples, aggregate impression data includes unique audience values (e.g., a measure of the quantity of unique audience members exposed to particular media), total impression count, frequency of impressions, etc. In some examples, the individual logged impressions are not discernable from the aggregate impression data.

[0111] As such, it is not readily discernable from the user information 1102a, 1102b whether instances of individual user-level impressions logged at

the database proprietors 110a, 110b correspond to the same audience member such that unique audience sizes indicated in the aggregate impression data of the user-information 1102a, 1102b are inaccurate for being based on duplicate counting of audience members. However, the impression identifier 1140 provided to the AME 108 enables the AME 108 to distinguish unique impressions and avoid overcounting a number of unique users and/or devices accessing the media. For example, the relationship between the user information 1102a from the partner A database proprietor 110a and the user information 1102b from the partner B database proprietor 110b for the client device 104 is not readily apparent to the AME 108. By including an impression identifier 1140 (or any similar identifier), the example AME 108 can associate user information corresponding to the same user between the user information 1102a, 1102b based on matching impression identifiers 140 stored in both of the user information 1102a, 1102b. The example AME 108 can use such matching impression identifiers 1140 across the user information 1102a, 1102b to avoid overcounting mobile devices and/or users (e.g., by only counting unique users instead of counting the same user multiple times).

[0112] A same user may be counted multiple times if, for example, an impression causes the client device 104 to send multiple user/device identifiers to multiple different database proprietors 110a, 110b without an impression identifier (e.g., the impression identifier 1140). For example, a first one of the database proprietors 110a sends first user information 1102a to the AME 108, which signals that an impression occurred. In addition, a second one of the database proprietors 110b sends second user information 1102b to the AME 108, which signals (separately) that an impression occurred. In addition, separately, the client device 104 sends an indication of an impression to the AME 108. Without knowing that the user information 1102a, 1102b is from the same impression, the AME 108 has an indication from the client device 104 of a single impression and indications from the database proprietors 110a, 110b of multiple impressions.

[0113] To avoid overcounting impressions, the AME 108 can use the impression identifier 1140. For example, after looking up user information

1102a, 1102b, the example partner database proprietors 110a, 110b transmit the impression identifier 1140 to the AME 108 with corresponding user information 1102a, 1102b. The AME 108 matches the impression identifier 1140 obtained directly from the client device 104 to the impression identifier 1140 received from the database proprietors 110a, 110b with the user information 102a-b to thereby associate the user information 1102a, 1102b with the media identifier 1122 and to generate impression information. This is possible because the AME 108 received the media identifier 1122 in association with the impression identifier 1140 directly from the client device 104. Therefore, the AME 108 can map user data from two or more database proprietors 110a, 110b to the same media exposure event, thus avoiding double counting.

[0114] Each unique impression identifier 1140 in the illustrated example is associated with a specific impression of media on the client device 104. The partner database proprietors 110a, 110b receive the respective user/device identifiers 114 and generate the user information 1102a, 1102b independently (e.g., without regard to others of the partner database proprietors 104a-b) and without knowledge of the media identifier 1122 involved in the impression. Without an indication that a particular user demographic profile in the user information 1102a (received from the partner database proprietor 110a) is associated with (e.g., the result of) the same impression at the client device 104 as a particular user demographic profile in the user information 1102b (received from the partner database proprietor 110b independently of the user information 1102a received from the partner database proprietor 110a), and without reference to the impression identifier 1140, the AME 108 may not be able to associate the user information 1102a with the user information 1102b and/or cannot determine that the different pieces of user information 1102a, 1102b are associated with a same impression and could, therefore, count the user information 1102a and the user information 1102b as corresponding to two different users/devices and/or two different impressions.

[0115] The examples of FIG. 11 illustrate methods and apparatus for collecting impression data at an audience measurement entity (or other entity). The examples of FIG. 11 may be used to collect impression information for any type of media, including static media (e.g., advertising images), streaming media (e.g., streaming video and/or audio, including content, advertising, and/or other types of media), and/or other types of media. For static media (e.g., media that does not have a time component such as images, text, a webpage, etc.), the example AME 108 records an impression once for each occurrence of the media being presented, delivered, or otherwise provided to the client device 104. For streaming media (e.g., video, audio, etc.), the example AME 108 measures demographics for media occurring over a period of time. For example, the AME 108 (e.g., via the app publisher 1110 and/or the media publisher 1120) provides beacon instructions to a client application or client software (e.g., the OS 1114, the web browser 1117, the app 1116, etc.) executing on the client device 104 when media is loaded at client application/software 1114-1117. In some examples, the beacon instructions are embedded in the streaming media and delivered to the client device 106 via the streaming media. In some examples, the beacon instructions cause the client application/software 1114-1117 to transmit a request (e.g., a pingback message) to an impression monitoring server 1132 at regular and/or irregular intervals (e.g., every minute, every 30 seconds, every 2 minutes, etc.). The example impression monitoring server 1132 identifies the requests from the web browser 1117 and, in combination with one or more database proprietors, associates the impression information for the media with demographics of the user of the web browser 1117.

[0116] In some examples, a user loads (e.g., via the browser 1117) a web page from a web site publisher, in which the web page corresponds to a particular 60-minute video. As a part of or in addition to the example web page, the web site publisher causes the data collector 1112 to send a pingback message (e.g., a beacon request) to a beacon server 1142 by, for example, providing the browser 1117 with beacon instructions. For example, when the beacon instructions are executed by the example browser 1117, the beacon

instructions cause the data collector 1112 to send pingback messages (e.g., beacon requests, HTTP requests, pings) to the impression monitoring server 1132 at designated intervals (e.g., once every minute or any other suitable interval). The example beacon instructions (or a redirect message from, for example, the impression monitoring server 1132 or a database proprietor 104a-b) further cause the data collector 1112 to send pingback messages or beacon requests to one or more database proprietors 110a, 110b that collect and/or maintain demographic information about users. The database proprietor 110a, 110b transmits demographic information about the user associated with the data collector 1112 for combining or associating with the impression determined by the impression monitoring server 1132. If the user closes the web page containing the video before the end of the video, the beacon instructions are stopped, and the data collector 1112 stops sending the pingback messages to the impression monitoring server 1132. In some examples, the pingback messages include timestamps and/or other information indicative of the locations in the video to which the numerous pingback messages correspond. By determining a number and/or content of the pingback messages received at the impression monitoring server 1132 from the client device 104, the example impression monitoring server 1132 can determine that the user watched a particular length of the video (e.g., a portion of the video for which pingback messages were received at the impression monitoring server 1132).

[0117] The client device 104 of the illustrated example executes a client application/software 1114-1117 that is directed to a host website (e.g., www.acme.com) from which the media 1118 (e.g., audio, video, interactive media, streaming media, etc.) is obtained for presenting via the client device 104. In the illustrated example, the media 1118 (e.g., advertisements and/or content) is tagged with identifier information (e.g., a media ID 1122, a creative type ID, a placement ID, a publisher source URL, etc.) and a beacon instruction. The example beacon instruction causes the client application/software 1114-1117 to request further beacon instructions from a beacon server 1142 that will instruct the client application/software 1114-1117

on how and where to send beacon requests to report impressions of the media 1118. For example, the example client application/software 1114-1117 transmits a request including an identification of the media 1118 (e.g., the media identifier 1122) to the beacon server 1142. The beacon server 1142 then generates and returns beacon instructions 1144 to the example client device 104. Although the beacon server 1142 and the impression monitoring server 132 are shown separately, in some examples the beacon server 1142 and the impression monitoring server 1132 are combined. In the illustrated example, beacon instructions 1144 include URLs of one or more database proprietors (e.g., one or more of the partner database proprietors 110a, 110b) or any other server to which the client device 104 should send beacon requests (e.g., impression requests). In some examples, a pingback message or beacon request may be implemented as an HTTP request. However, whereas a transmitted HTTP request identifies a webpage or other resource to be downloaded, the pingback message or beacon request includes the audience measurement information (e.g., ad campaign identification, content identifier, and/or device/user identification information) as its payload. The server to which the pingback message or beacon request is directed is programmed to log the audience measurement data of the pingback message or beacon request as an impression (e.g., an ad and/or content impression depending on the nature of the media tagged with the beaconing instructions). In some examples, the beacon instructions received with the tagged media 1118 include the beacon instructions 1144. In such examples, the client application/software 1114-1117 does not need to request beacon instructions 1144 from a beacon server 1142 because the beacon instructions 1144 are already provided in the tagged media 1118.

[0118] When the beacon instructions 1144 are executed by the client device 104, the beacon instructions 1144 cause the client device 104 to send beacon requests (e.g., repeatedly at designated intervals) to a remote server (e.g., the impression monitoring server 1132, the media publisher 1120, the database proprietors 110a, 110b, or another server) specified in the beacon instructions 1144. In the illustrated example, the specified server is a server of

the AME 108, namely, at the impression monitoring server 1132. The beacon instructions 1144 may be implemented using JavaScript or any other types of instructions or script executable via a client application (e.g., a web browser) including, for example, Java, HTML, etc.

[0119] FIG. 12 depicts an example age prediction model 306 generated based on a classification tree (e.g., a decision tree) that uses subscriber activity metrics 138 to predict real ages (e.g., predicted age PDFs 305) of subscribers of a database proprietor 110 (FIG. 1). The example subscriber activity metrics 138 include, but are not limited to, user age reported to database proprietor, number of contacts (e.g., friends, connections, etc.), median age of contacts, privacy of birthdate (e.g., birthdate is or is not displayed to contacts), year of high school graduation, gender, whether a current address is associated with the subscriber's profile, whether a profile picture is associated with the subscriber's profile, whether a mobile phone number is associated with the subscriber's profile, subscriber activity that occurred within the last thirty days, subscriber activity that occurred within the last seven days, email address the subscriber used to register is in the .edu domain, percent of contacts that are female, degree of privacy settings (e.g., high privacy, medium privacy, low privacy, etc.), frequency of login (F_{LOGIN}), frequency of posting status updates (F_{POST}), percentage of logins done via a mobile device (LOGIN), whether a relationship status is associated with the subscriber's profile, number of days since the subscriber registered with the database proprietor, number of messages sent in the past seven days, webpages viewed in the past seven days, etc.

[0120] In the illustrated example of FIG. 12, the age prediction model 306 is a classification tree. In the illustrated example of FIG. 12, the classification tree implementing the age prediction model 306 may include many nodes and terminals. A detailed view of a portion of the example age prediction model 306 is shown in FIG. 12 in which the example age prediction model has intermediate (e.g., decision) nodes 1200a-1200c and terminal (e.g., prediction) nodes 1202a-1202d. The intermediate nodes 1200a-1200c represent test conditions based on the subscriber activity metrics 138 that

result in branching paths. For example, an intermediate node 1200a may represent a test condition based on the frequency with which a subscriber logs into the database proprietor 110. In such an example, the test condition may be whether the subscriber logs into the database proprietor either (i) less than five times per week (< 5 times/week), or (ii) greater than or equal to five times per week (≥ 5 times/week). The branching paths may either lead to another intermediate node 1200b-1200c or one of the terminal nodes 1202a-1202d. In the illustrated example, the example terminal nodes 1202a-1202d represent a classification (e.g., the predicted age PDF 305) based on the subscriber activity metrics 138 that is output from the age prediction model 306 (e.g., and used by the age predictor 304 of FIG. 3).

[0121] However, the predictive value of the age prediction model 306 model degrades over time as subscriber behaviors (e.g., as measured by subscriber activity metrics 138) on which the age prediction model 306 is based change over time. In some examples, small changes in subscriber-behavior can change the terminal node 1202a-1202d reached by the age prediction model 306. The manner in which outcomes are affected by such small behavior changes is observable using the example age prediction model 306 of FIG. 12. For example, in FIG. 12, the age prediction model 306 is used to predict the real age (e.g., a predicted age PDF 305 associated with terminal nodes 1202a-1202d) of a subscriber (e.g., “Subscriber A”) at day $D=0$ and at day $D=822$. In the illustrated example, the age prediction model 306 is generated on day $D=0$. On day $D=0$, the login frequency of the subscriber A is greater than or equal to five times per week. As a result, a decision at an intermediate node 1200a using login frequency (F_{LOGIN}) as a test condition, the branches down a path that leads to an intermediate node 1202c. In the illustrated example, a decision at an intermediate node 1200c using login type (LOGIN) as a test condition branches down a path leading to a terminal node 1202c because less than 50% of the logins by subscriber A are via a mobile device type when logging into the database proprietor 110. In the illustrated example, by day $D=822$ (e.g., 822 days after day $D=0$), the subscriber activity metrics 138 of the subscriber A change so that frequency of login (F_{LOGIN}) is

less than five times per week. As a result, if the prediction model 306 generated on day $D=0$ is used to predict the real age of the subscriber on day $D=822$, the resulting terminal node 1202b would be different compared to the results on day $D=0$.

[0122] From the foregoing, it will be appreciated that examples have been disclosed which allow association of accurate age-based demographic groups with impressions generated to exposure to media. Additionally, it will be appreciated that examples have been disclosed which enhance the operations of a computer to improve the accuracy of impression-based data so that computers and processing systems therein can be relied upon to produce audience analysis information with higher accuracies.

[0123] Although certain example methods, apparatus and articles of manufacture have been disclosed herein, the scope of coverage of this patent is not limited thereto. On the contrary, this patent covers all methods, apparatus and articles of manufacture fairly falling within the scope of the claims of this patent.

What Is Claimed Is:

1. A method comprising:

predicting a first age probability density function for an audience member at a first time, the first age probability density function indicative of a first probability that the audience member is in a first age range at the first time; and

in response to receiving media monitoring data associated with the audience member at a second time after the first time:

determining a second age probability density function by applying a first aging factor to the first age probability density function for the audience member, the second age probability density function indicative of a second probability that the audience member is in the first age range at the second time, and

associating the media monitoring data with the second age probability density function.

2. A method as defined in claim 1, further comprising determining the first aging factor by dividing a first number of days between the first time and the second time by a second number of total days in the first age range.

3. A method as defined in claim 1, wherein the first age probability density function is indicative of a third probability that the audience member is in a second age range at the first time, and further comprising in response to receiving the media monitoring data associated with the audience member at the second time after the first time:

determining the second age probability density function by applying a second aging factor to the third probability indicated by the first age probability density function for the audience member to create a fourth probability that the audience member is in the second age range at the second time.

4. The method of claim 3, wherein the fourth probability comprises a portion of the first probability and a portion of the third probability.

5. The method of claim 1, wherein the audience member is a subscriber to a database proprietor, and further comprising predicting the first age probability density function for the audience member further using an age prediction model based on subscriber characteristics of the audience member collected by the database proprietor.

6. The method of claim 5, wherein the age prediction model is a classification tree and the first age probability density function is a terminal node of the classification tree.

7. An apparatus, comprising:

an age predictor to predict a first age probability density function for an audience member at a first time, the first age probability density function indicative of a first probability that the audience member is in a first age range at the first time; and

an age updater to, in response to receiving media monitoring data associated with the audience member at a second time after the first time:

determine a second age probability density function by applying a first aging factor to the predicted first age probability density function for the audience member, the second age probability density function indicative of a second probability that the audience member is in the first age range at the second time, and

associate the media monitoring data with the second age probability density function.

8. An apparatus as defined in claim 7, wherein in the age updater is further to determine the first aging factor by dividing a first number of days between the first time and the second time by a second number of total days in the first age range.

9. An apparatus as defined in claim 7, wherein the first age probability density function is indicative of a third likelihood that the audience member is in a second age range at the first time, and wherein in response to receiving media monitoring data associated with the subscriber at the second time after the first time, the age updater is further to:

determine the second age probability density function by applying a second aging factor to the third probability of the first age probability density function for the audience member to create a fourth probability that the audience member is in the second age range at the second time.

10. An apparatus as defined in claim 9, wherein the fourth probability comprises a portion of the first probability and a portion of the third probability.

11. The apparatus of claim 7, wherein the audience member is a subscriber to a database proprietor, and the age predictor is to predict the first age probability density function for the audience member fusing an age prediction model based on subscriber characteristics of the audience member collected by the database proprietor.

12. The apparatus of claim 11, wherein the age prediction model is a classification tree and the first age probability density function is a terminal node of the classification tree.

13. A tangible computer readable storage medium comprising instructions which, when executed, cause a machine to at least:

predict a first age probability density function for an audience member at a first time, the first age probability density function indicative of a first probability that the audience member is in a first age range at the first time; and

in response to receiving media monitoring data associated with the subscriber at a second time after the first time:

determine a second age probability density function by applying a first aging factor to the first age probability density function for the audience member, the second age probability density function indicative of a second probability that the audience member is in the first age range at the second time, and

associate the media monitoring data with the second age probability density function.

14. A method as defined in claim 13, further comprising determining the first aging factor by dividing a first number of days between the first time and the second time by a second number of total days in the first age range.

15. A method as defined in claim 13, wherein the first age probability density function is indicative of a third probability that the audience member is in a second age range at the first time, and further comprising in response to receiving the media monitoring data associated with the audience member at the second time after the first time:

determining the second age probability density function by applying a second aging factor to the third probability indicated by first age probability density function for the subscriber to create a fourth probability that the subscriber is in the second age range at the second time.

16. The tangible computer readable storage medium of claim 15, wherein the fourth probability comprises a portion of the first probability and a portion of the third probability.

17. The tangible computer readable storage medium of claim 13, wherein the audience member is a subscriber to a database proprietor, wherein the instructions, when executed are further to cause the machine to predict the first age probability density function for the audience member using an age prediction model based on subscriber characteristics of the audience member collected by the database proprietor.

18. The tangible computer readable storage medium of claim 17, wherein the age prediction model is a classification tree and the first age probability density function is a terminal node of the classification tree.

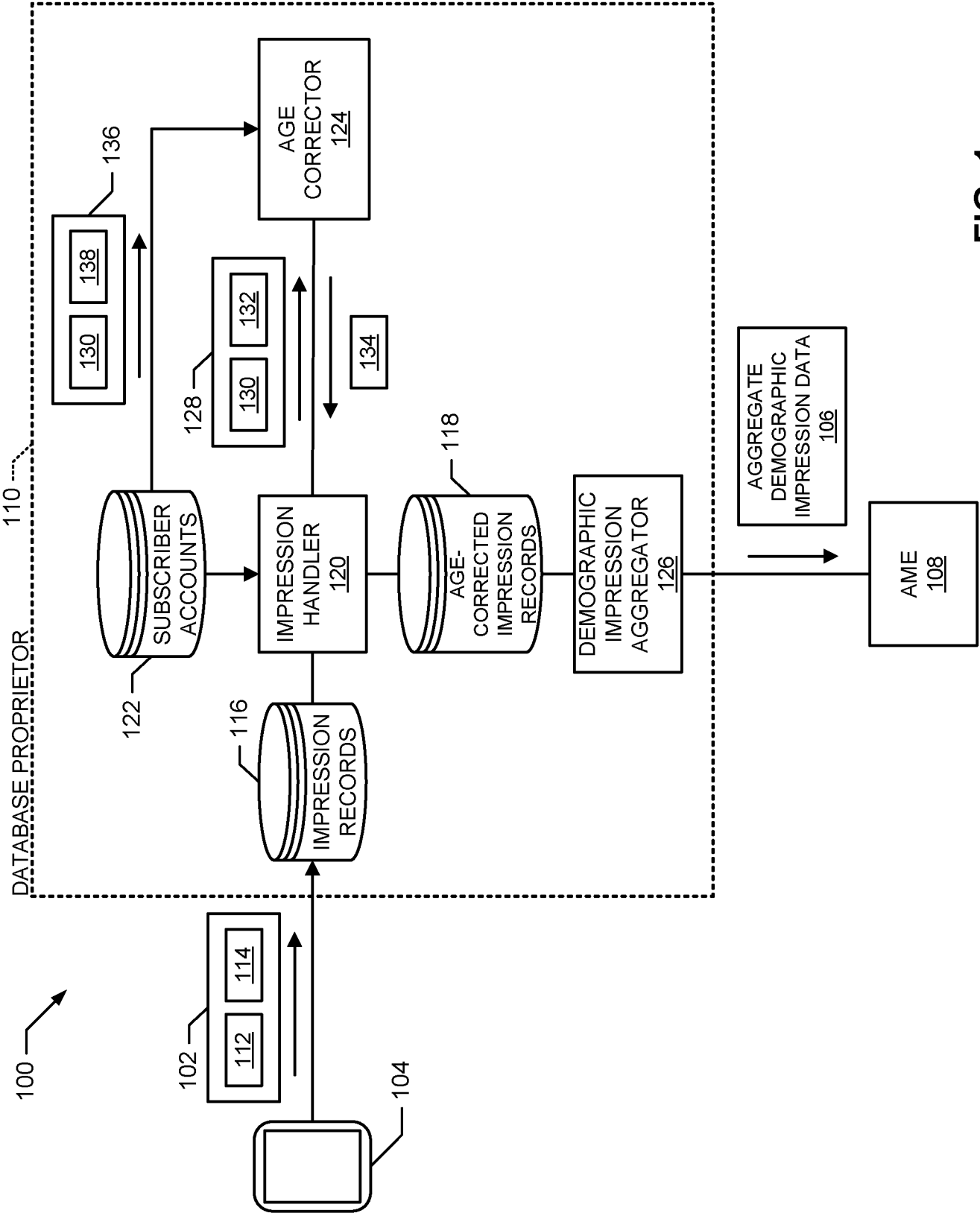


FIG. 1

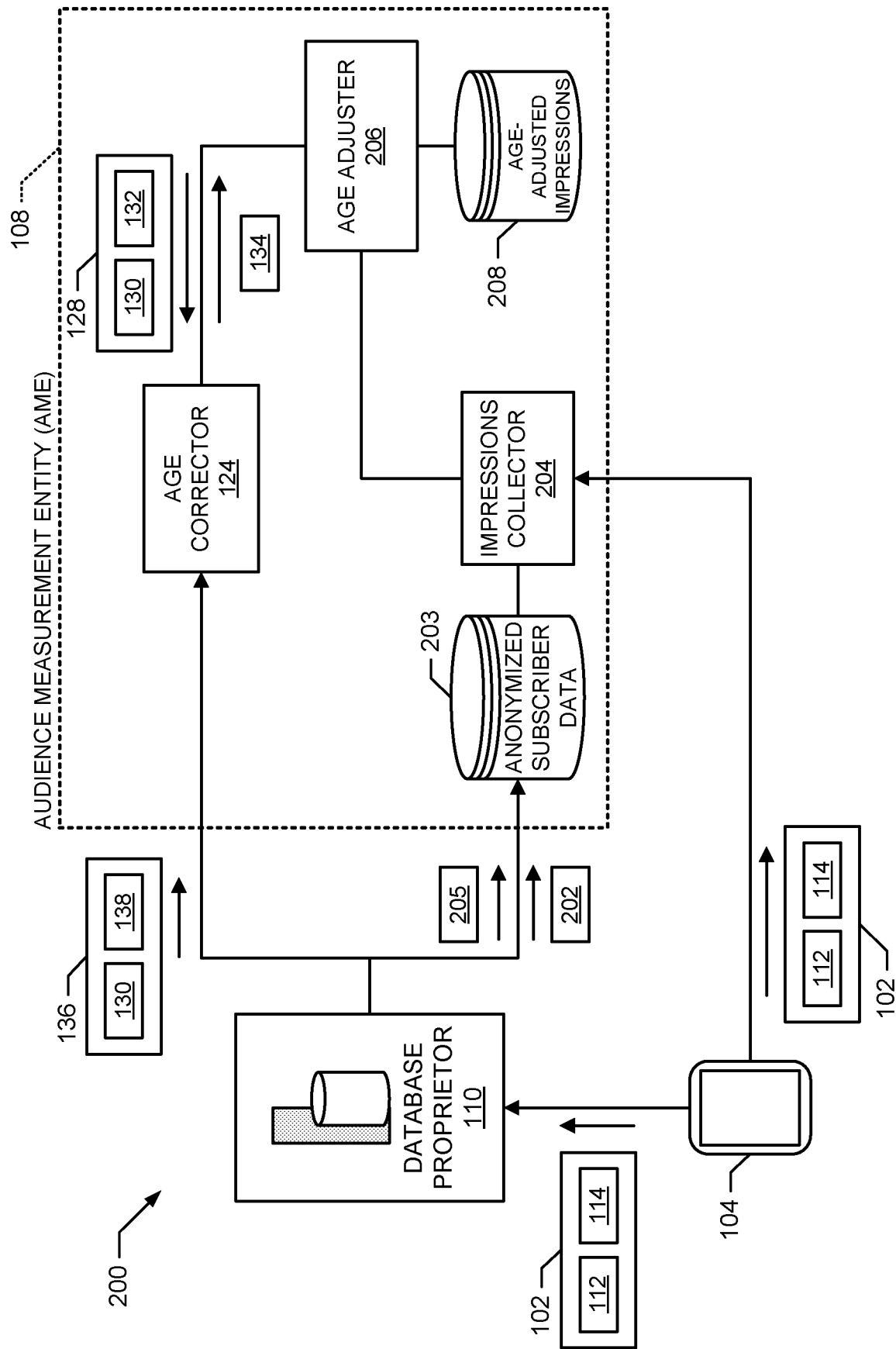


FIG. 2

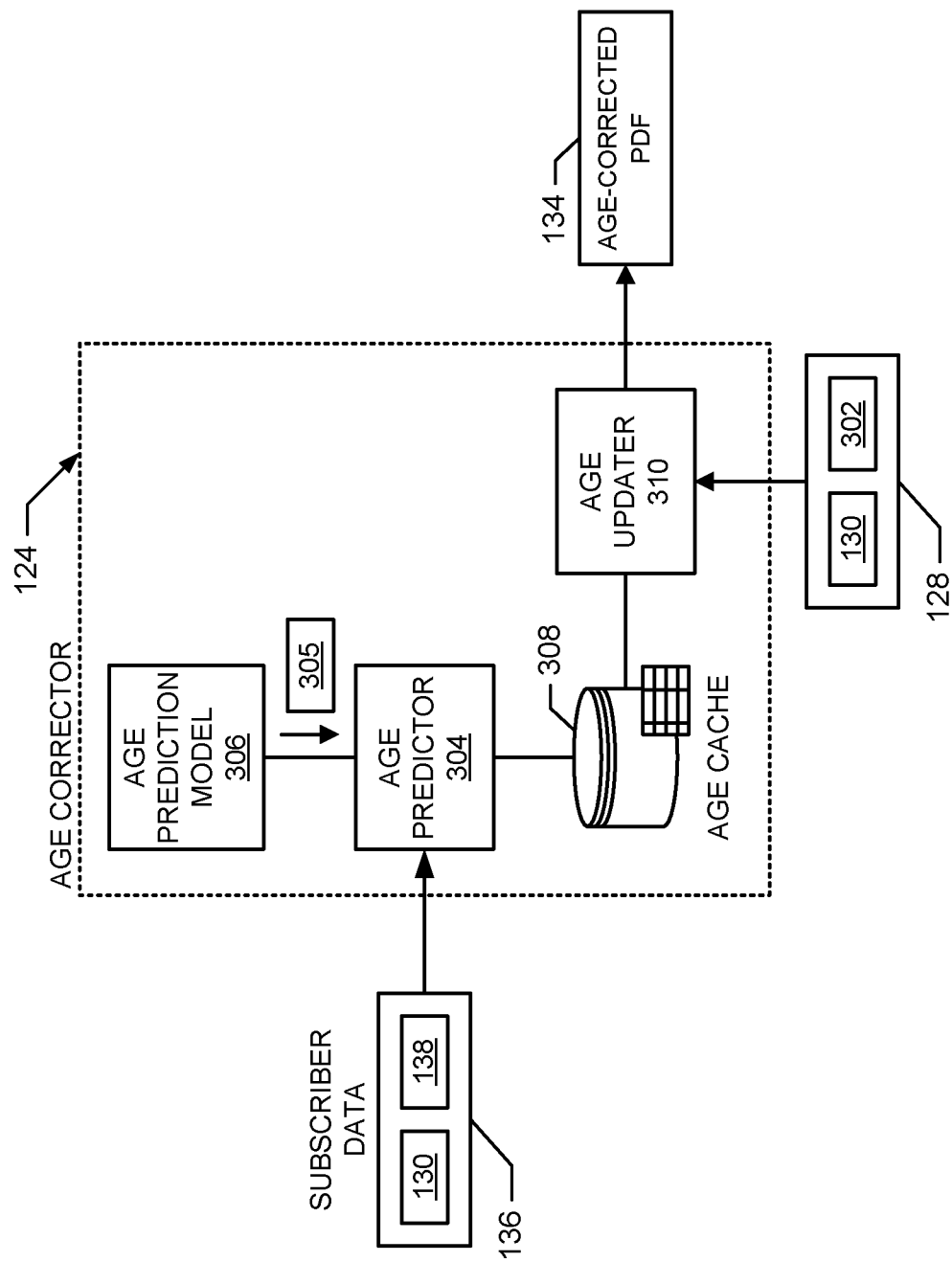
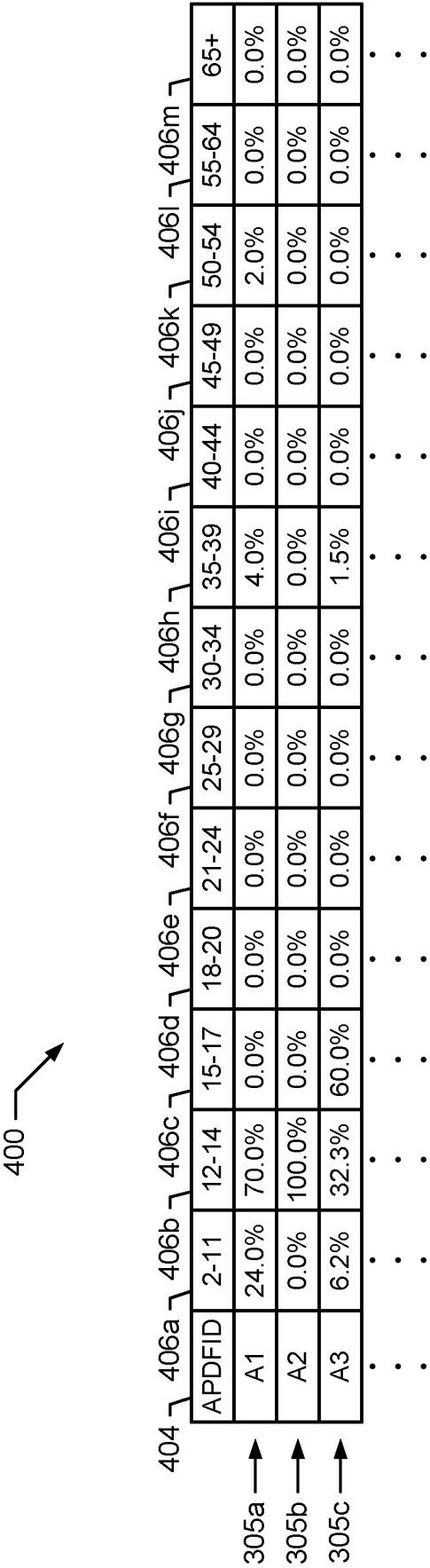


FIG. 3



AGE PROBABILITY DENSITY FUNCTIONS

FIG. 4

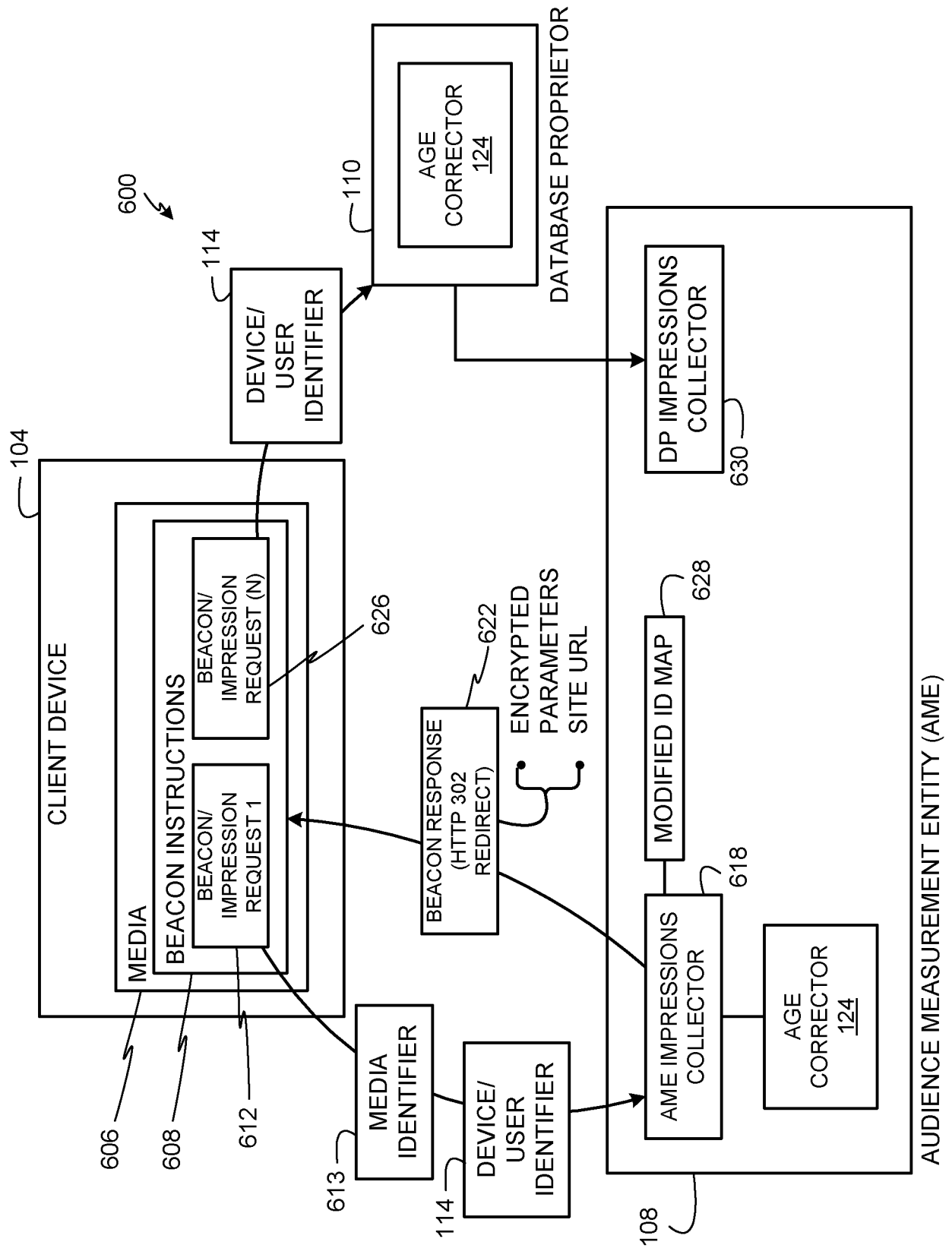
500

130 UID	404 AGE PDF ID	502 PREDICTION DATE
8PR6PYRGQC	A7	2015-10-23
ERX269AS22	B9	2016-11-24
ZBX9A8C5NA	D1	2018-02-09
W7WWUTDJFF	A3	2018-12-05
3RX2VH3K2E	C5	2019-07-02
⋮	⋮	⋮

PREDICTED AGE PROBABILITY DENSITY FUNCTIONS
STORED IN THE AGE CACHE

FIG. 5

FIG. 6



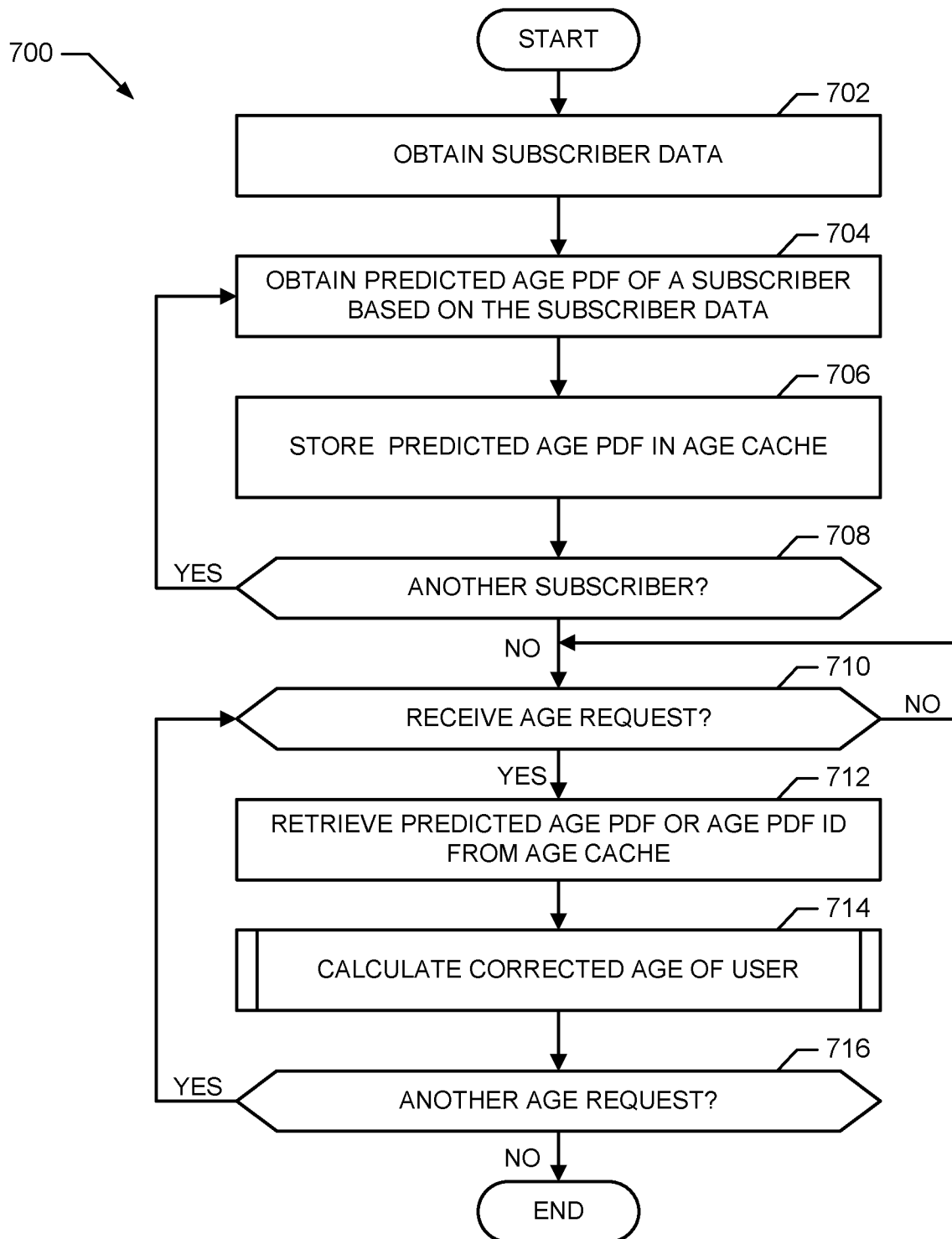
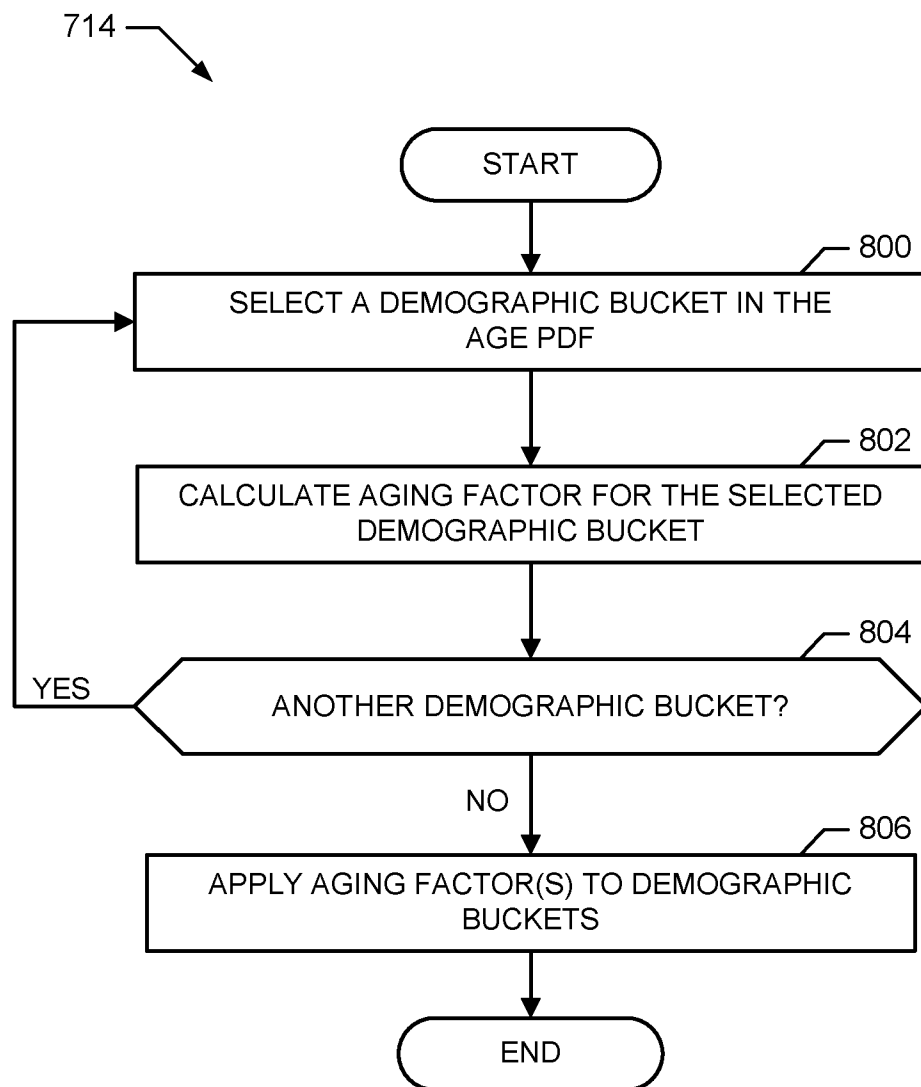
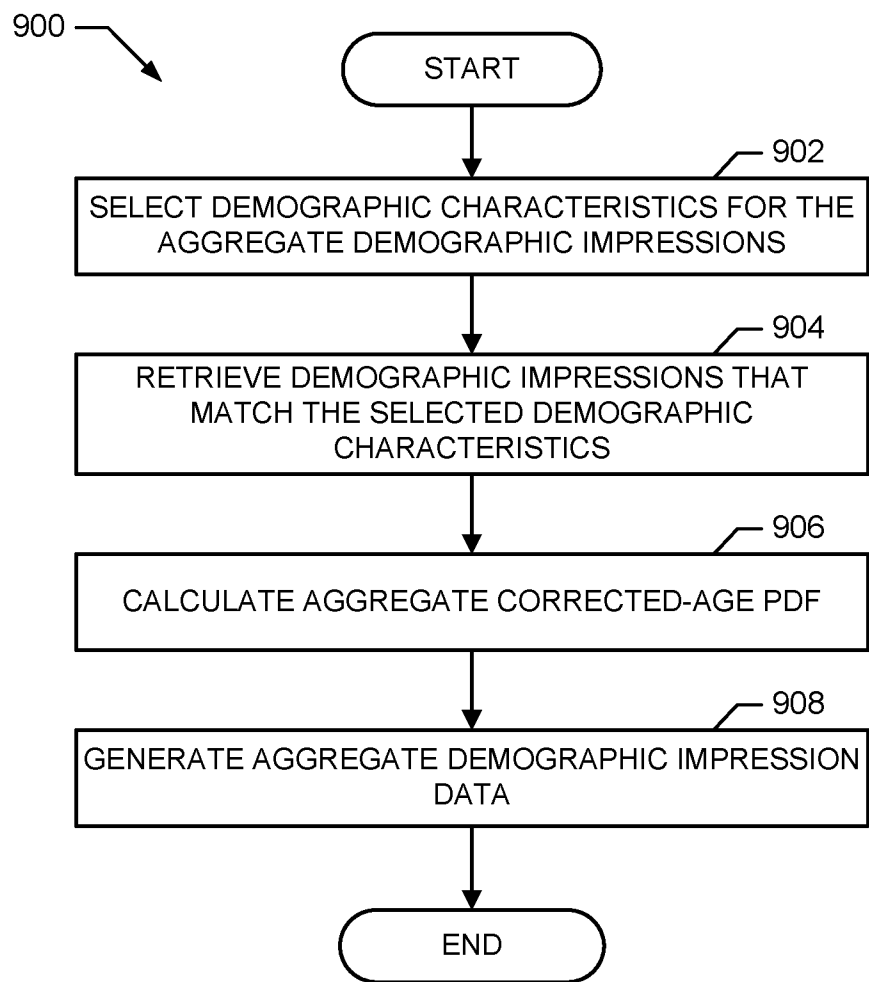


FIG. 7

**FIG. 8**

**FIG. 9**

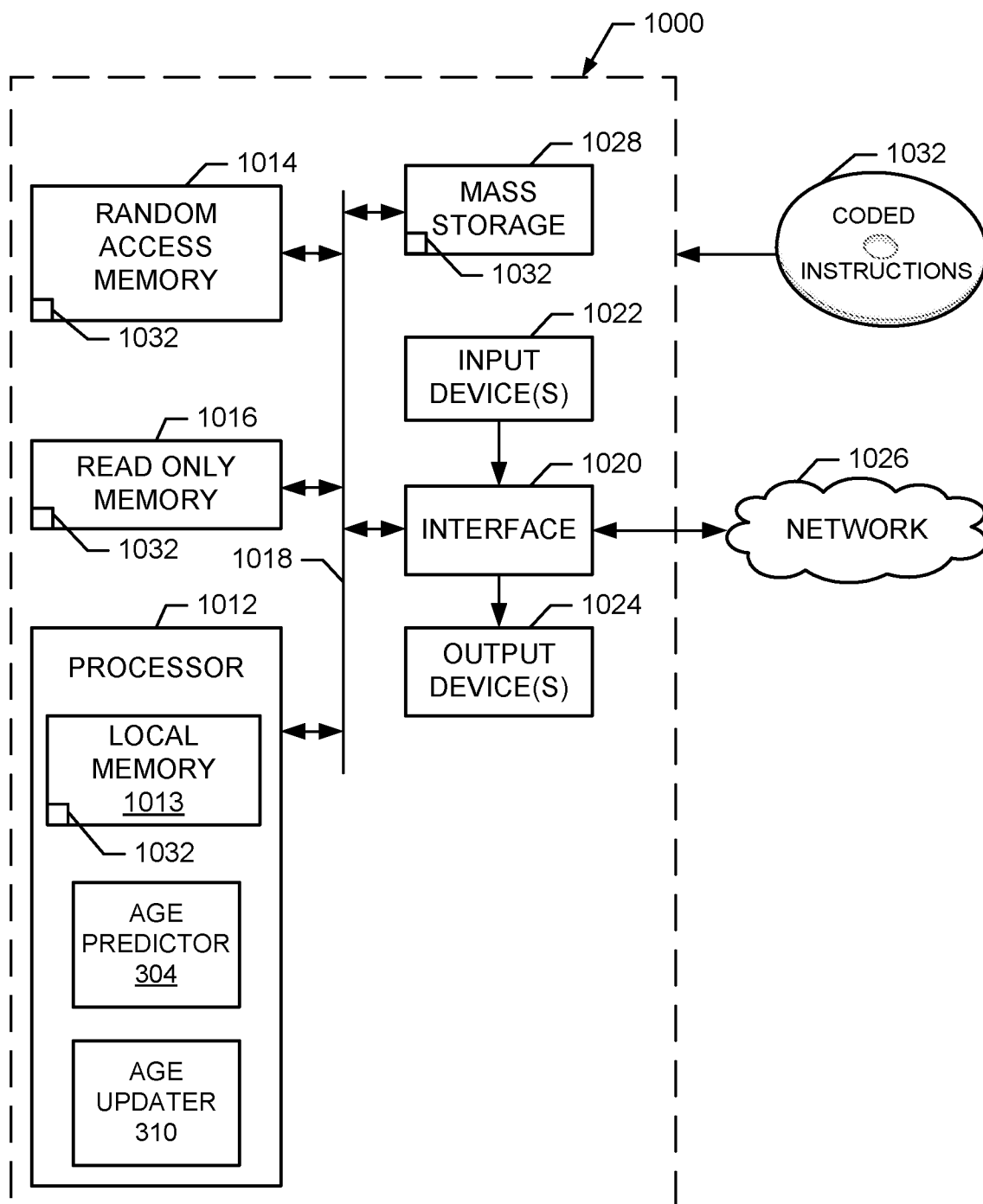


FIG. 10

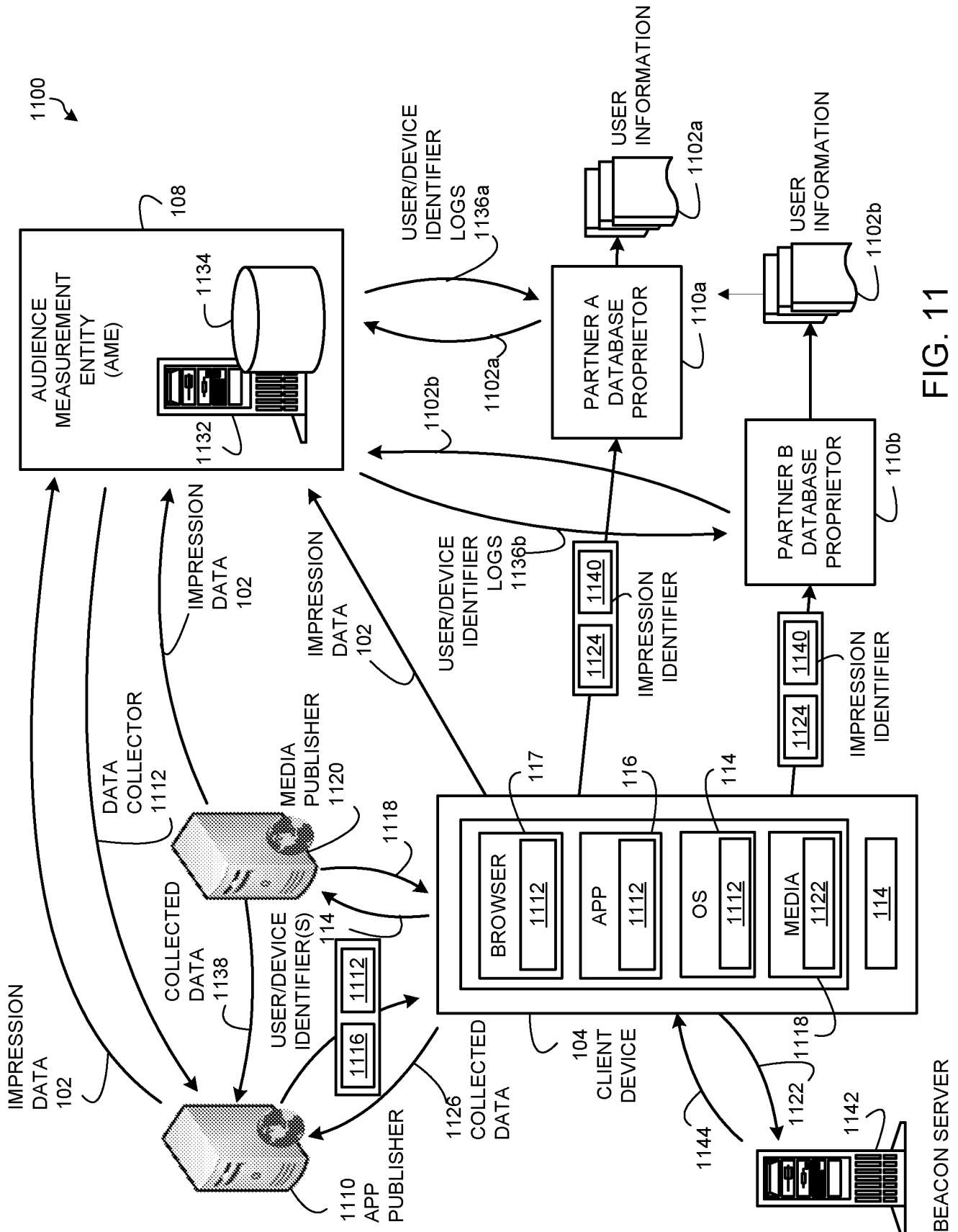


FIG. 11

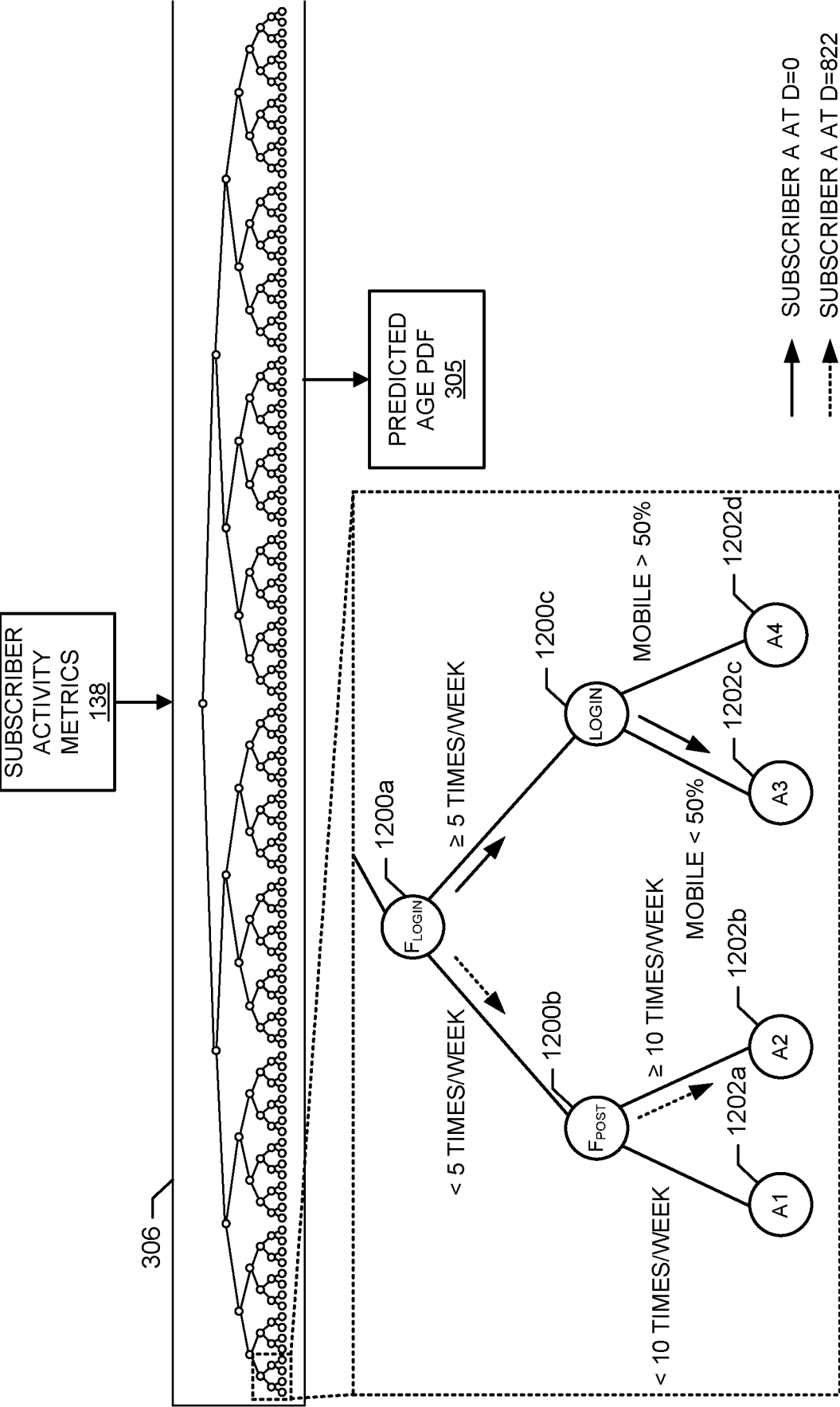


FIG. 12

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2015/067917**A. CLASSIFICATION OF SUBJECT MATTER****H04N 21/258(2011.01)i, H04N 21/45(2011.01)i, H04N 21/442(2011.01)i**

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04N 21/258; G06Q 30/02; G10L 15/00; H04N 21/442; G06F 17/30; G06Q 30/02; H04H 60/32; H04N 21/25; H04N 21/45

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS(KIPO internal) & Keywords: age range, probability, audience, time, aging factor, media impression

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2014-0324544 A1 (PAUL DONATO et al.) 30 October 2014 See paragraphs [0037], [0170]-[0176]; claim 1; and figures 13, 16.	1,3-7,9-13,15-18
A		2,8,14
Y	US 2013-0013308 A1 (YANG CAO et al.) 10 January 2013 See paragraphs [0019]-[0022], [0031], [0044]; claim 35; and figure 1.	1,3-7,9-13,15-18
A	US 2012-0072469 A1 (ALBERT R. PEREZ et al.) 22 March 2012 See paragraphs [0020]-[0028], [0034]-[0049]; and figures 1-3, 6-7.	1-18
A	US 8910195 B1 (GOOGLE INC.) 09 December 2014 See column 9, line 13 - column 13, line 16; claims 1-8; and figures 2A-2B.	1-18
A	KR 10-2011-0017895 A (ALMONDNET, INC.) 22 February 2011 See paragraphs [0044]-[0049]; claims 6-7; and figure 2.	1-18



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

23 May 2016 (23.05.2016)

Date of mailing of the international search report

23 May 2016 (23.05.2016)

Name and mailing address of the ISA/KR

International Application Division
Korean Intellectual Property Office
189 Cheongsu-ro, Seo-gu, Daejeon, 35208, Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

LEE, Jin Ick

Telephone No. +82-42-481-5770



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2015/067917

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2014-0324544 A1	30/10/2014	WO 2014-176343 A1	30/10/2014
US 2013-0013308 A1	10/01/2013	CN 102834842 A	19/12/2012
		EP 2550614 A1	30/01/2013
		US 9105053 B2	11/08/2015
		WO 2011-116514 A1	29/09/2011
US 2012-0072469 A1	22/03/2012	AU 2011-374955 B2	03/09/2015
		US 2015-0332297 A1	19/11/2015
		US 9092797 B2	28/07/2015
		WO 2013-025227 A1	21/02/2013
US 8910195 B1	09/12/2014	US 2015-0237395 A1	20/08/2015
		WO 2015-126437 A1	27/08/2015
KR 10-2011-0017895 A	22/02/2011	AU 2009-256278 A1	10/12/2009
		CA 02724627 A1	10/12/2009
		CN 102150435 A	10/08/2011
		CN 102150435 B	16/07/2014
		CN 104168492 A	26/11/2014
		EP 2304672 A2	06/04/2011
		JP 2011-526100 A	29/09/2011
		JP 5737518 B2	17/06/2015
		US 2009-0300675 A1	03/12/2009
		US 2015-0312614 A1	29/10/2015
		US 9083853 B2	14/07/2015
		WO 2009-149128 A2	10/12/2009
		WO 2009-149128 A3	01/04/2010