US009118641B1

(12) **United States Patent**
Paris, III

(10) **Patent No.: US 9,118,641 B1**
(45) **Date of Patent: Aug. 25, 2015**

(54) **DE-IDENTIFYING MEDICAL HISTORY INFORMATION FOR MEDICAL UNDERWRITING**

(75) Inventor: **Andrew L. Paris, III**, Victor, NY (US)

(73) Assignee: **Vigilytics LLC**, Victor, NY (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 242 days.

(21) Appl. No.: **12/827,745**

(22) Filed: **Jun. 30, 2010**

Related U.S. Application Data

(60) Provisional application No. 61/222,428, filed on Jul. 1, 2009.

(51) **Int. Cl.**

H04L 9/32	(2006.01)
H04L 29/06	(2006.01)
G06F 1/26	(2006.01)
G08B 13/00	(2006.01)

(52) **U.S. Cl.**
CPC ***H04L 63/08*** (2013.01); ***H04L 63/0428*** (2013.01)

(58) **Field of Classification Search**
CPC H04L 63/08; H04L 9/08
USPC 380/28–30, 255–283; 713/150–154, 713/160–167, 189–193; 705/3, 5
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2004/0143403 A1 7/2004 Brandon et al.
2005/0234739 A1* 10/2005 Schoenberg 705/2
2005/0261941 A1* 11/2005 Scarlat 705/3
2006/0004772 A1 1/2006 Hagan et al.
2006/0020611 A1 1/2006 Gilbert et al.
2006/0184524 A1* 8/2006 Pollanz 707/6
2006/0229919 A1* 10/2006 Pugh 705/3
2007/0043594 A1 2/2007 Lavergne
2007/0219824 A1 9/2007 Rawlings et al.
2008/0010254 A1 1/2008 Settimi
2008/0120296 A1 5/2008 Kariathungal et al.
2008/0133273 A1 6/2008 Marshall
2008/0147554 A1 6/2008 Stevens et al.
2009/0024416 A1* 1/2009 McLaughlin et al. 705/3
2009/0070146 A1 3/2009 Haider et al.
2009/0287502 A1 11/2009 Roberts et al.
2010/0162355 A1 6/2010 Zimmerman et al.
2011/0184748 A1 7/2011 Fierro et al.
2011/0225007 A1 9/2011 Theis
2011/0258000 A1* 10/2011 Green et al. 705/3

OTHER PUBLICATIONS

"Group," The American Heritage Dictionary, 2014 Houghton Mifflin Harcourt, Retrieved online: https://ahdictionary.com/word/search.html?q=group (accessed Oct. 23, 2014), 3 pages.
"Health Information Privacy: Business Associates," U.S. Department of Health and Human Services, Apr. 3, 2003, Retrieved online: http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/businessassociates.html, (accessed Oct. 9, 2014), 2 pages.

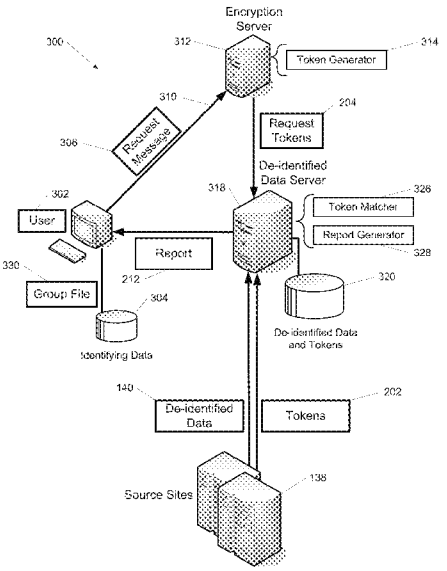
* cited by examiner

Primary Examiner — Evans Desrosiers
(74) *Attorney, Agent, or Firm* — Fish & Richardson P.C.

(57) **ABSTRACT**

A computer-implemented method includes producing information that characterizes a group of individuals from a set of private data representing characteristics of the individuals. The identity of the individuals is unattainable from the produced information. The method also includes providing the produced information to report the characteristics of the group.

19 Claims, 7 Drawing Sheets



Datavant Exhibit 1001

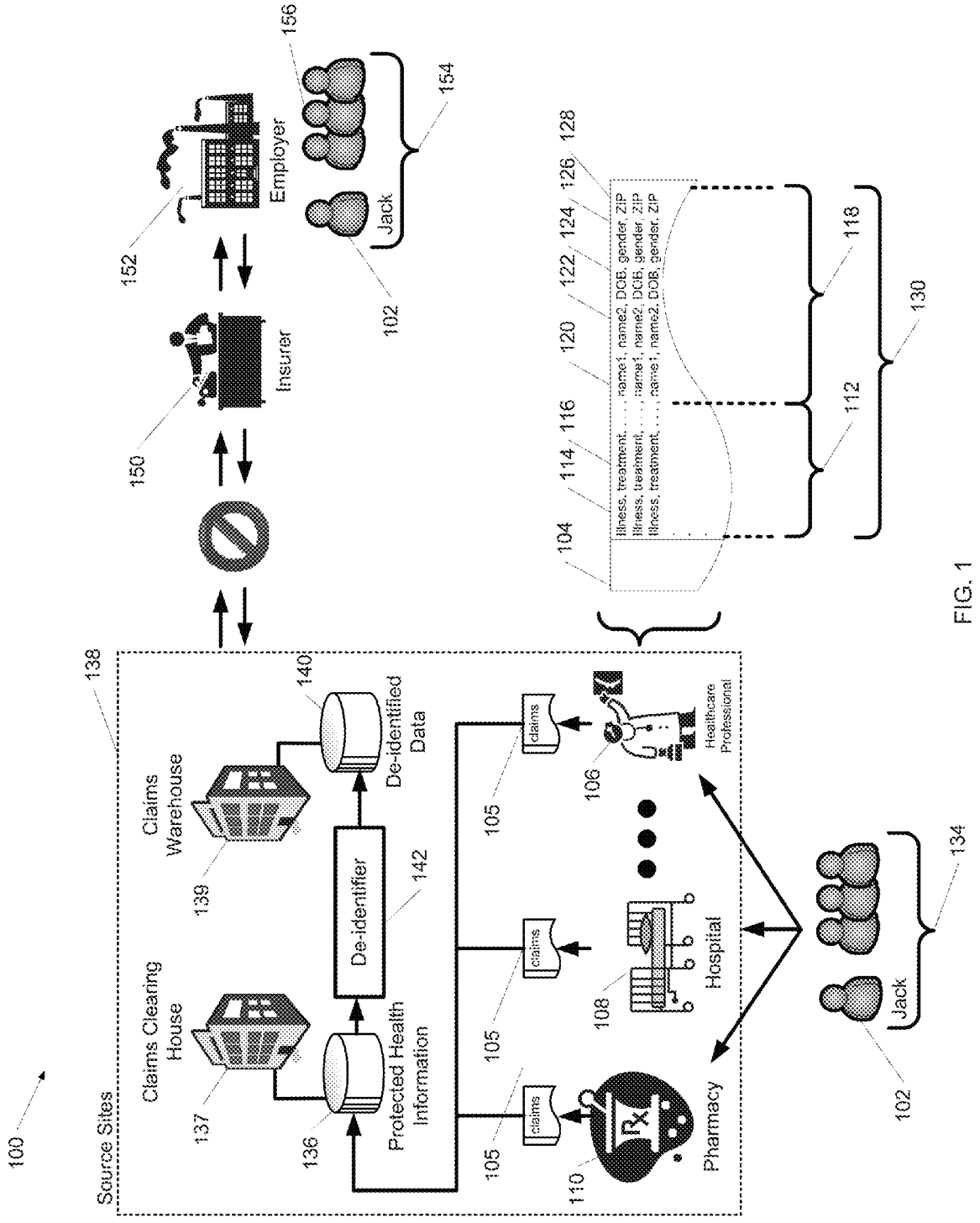

100
Source Sites
138
Claims Clearing House
137
Claims Warehouse
139
Protected Health Information
136
De-identifier
142
De-identified Data
140
claims
105
Pharmacy
110
Hospital
108
Healthcare Professional
106
Jack
102
134
Insurer
150
Employer
152
Jack
154
156
104
Illness, treatment, . . . , name1, name2, DOB, gender, ZIP
Illness, treatment, . . . , name1, name2, DOB, gender, ZIP
Illness, treatment, . . . , name1, name2, DOB, gender, ZIP
114
116
120
122
124
126
128
112
118
130

FIG. 1

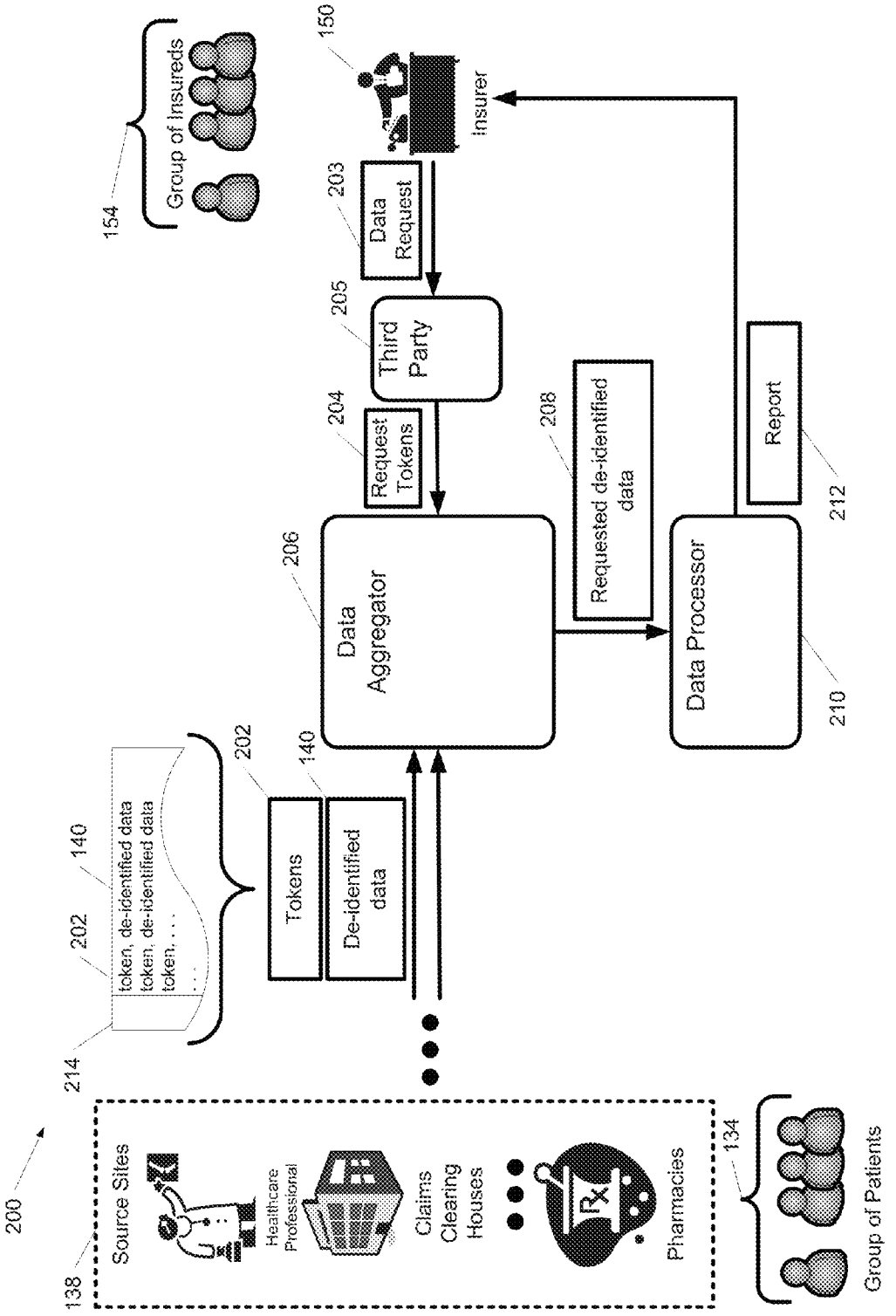

200
138
Source Sites
Healthcare Professional
Claims Clearing Houses
Pharmacies
134
Group of Patients
214
202
token, de-identified data
token, de-identified data
token, . . .
. . .
140
202
Tokens
De-identified data
206
Data Aggregator
204
Request Tokens
205
Third Party
203
Data Request
150
Insurer
154
Group of Insureds
208
Requested de-identified data
210
Data Processor
212
Report

FIG. 2

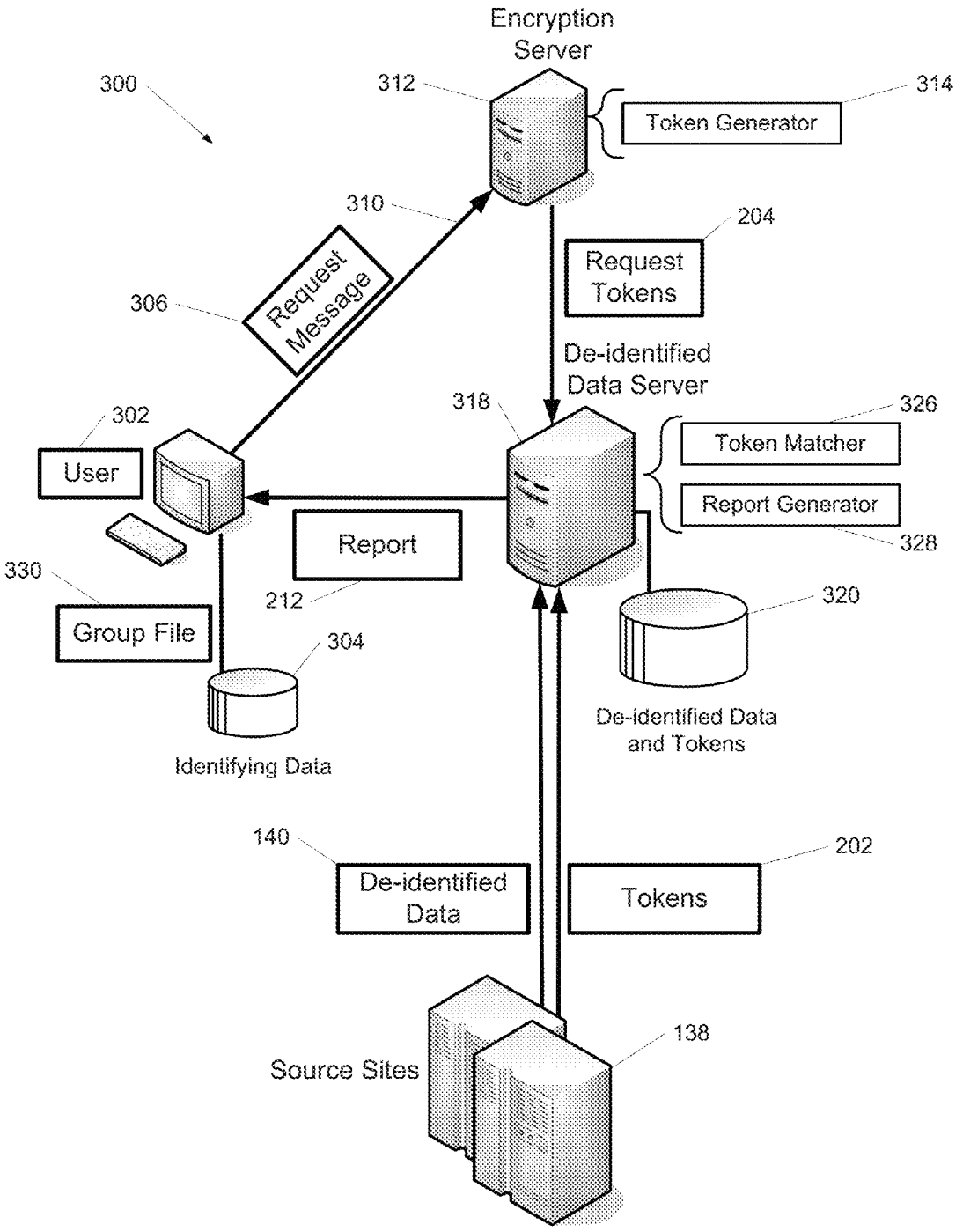

Encryption Server
300
312
314
Token Generator
310
204
Request Tokens
306
Request Message
De-identified Data Server
318
326
Token Matcher
302
User
Report Generator
328
Report
330
212
320
Group File
304
De-identified Data and Tokens
Identifying Data
140
202
De-identified Data
Tokens
Source Sites
138

FIG. 3

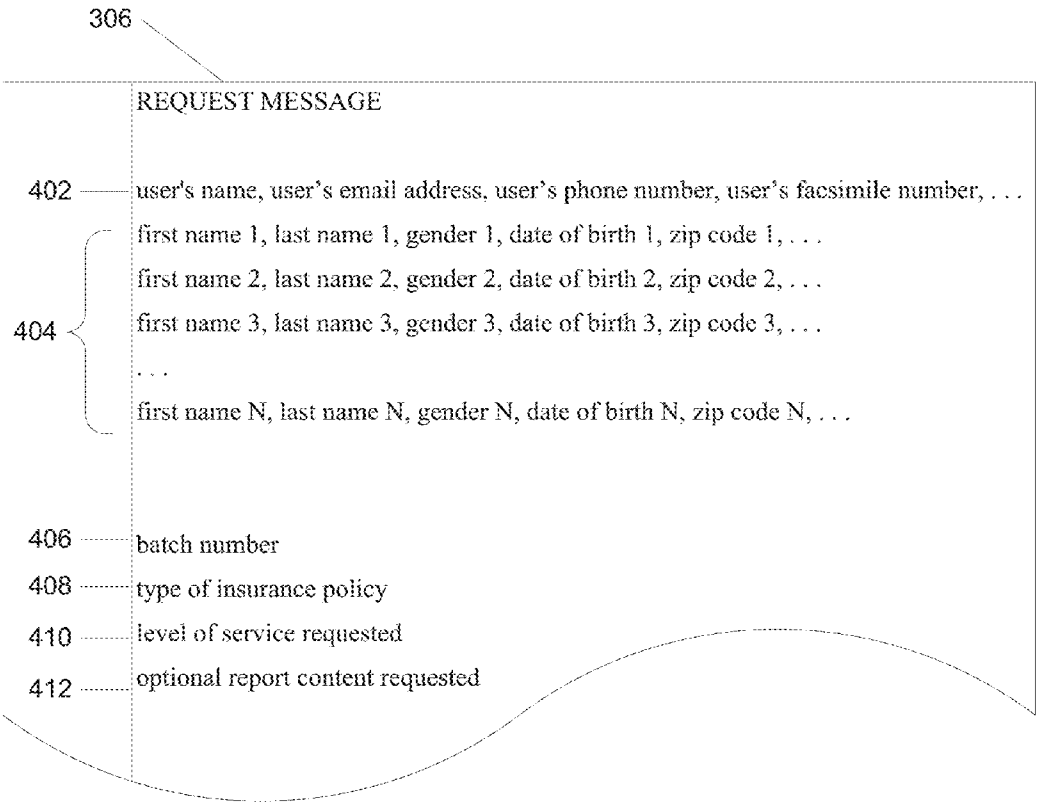

306
REQUEST MESSAGE
402
user's name, user's email address, user's phone number, user's facsimile number, . . .
404
first name 1, last name 1, gender 1, date of birth 1, zip code 1, . . .
first name 2, last name 2, gender 2, date of birth 2, zip code 2, . . .
first name 3, last name 3, gender 3, date of birth 3, zip code 3, . . .
. . .
first name N, last name N, gender N, date of birth N, zip code N, . . .
406
batch number
408
type of insurance policy
410
level of service requested
412
optional report content requested

FIG. 4

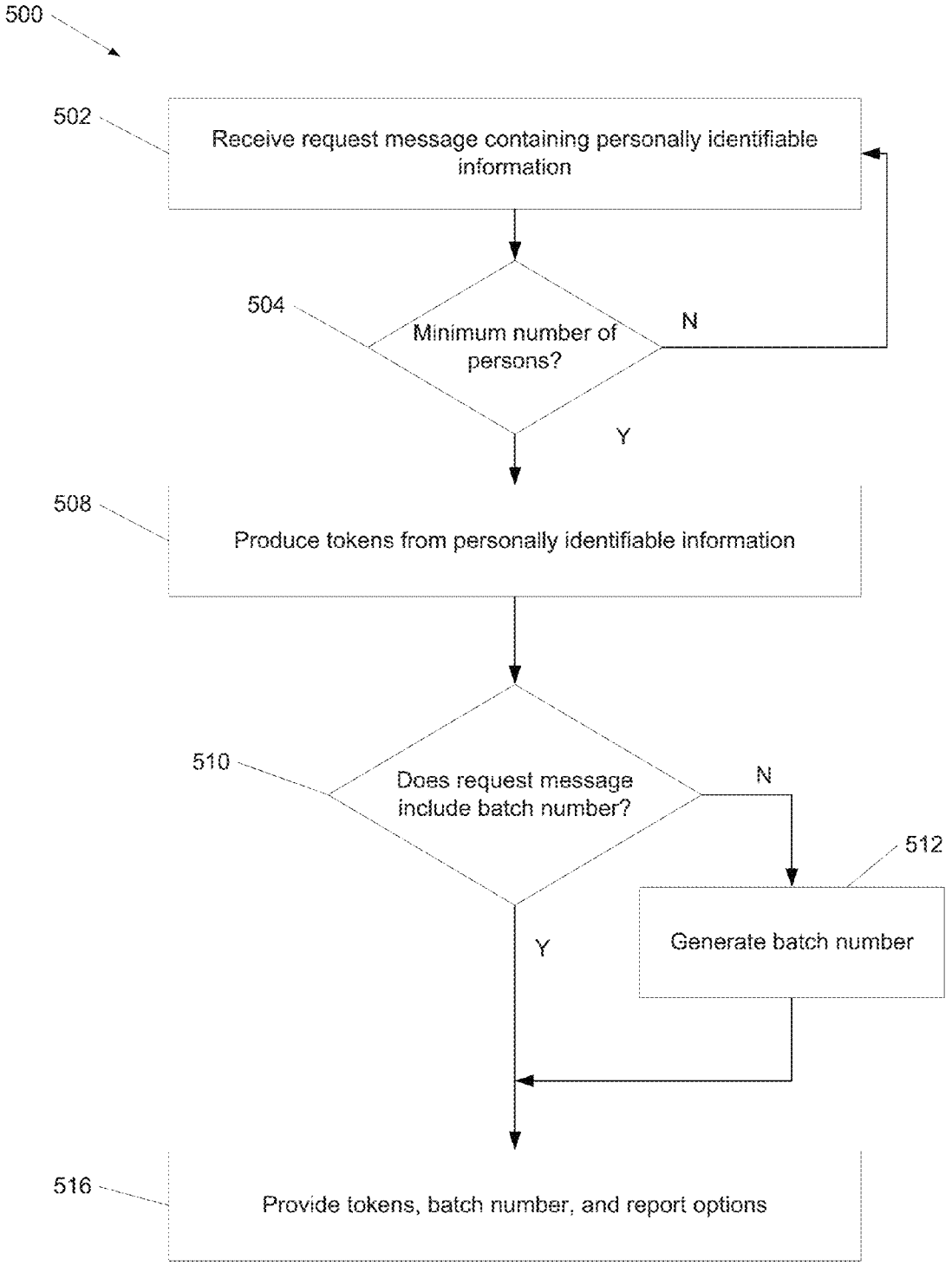

500
502
Receive request message containing personally identifiable information
504
Minimum number of persons?
N
Y
508
Produce tokens from personally identifiable information
510
Does request message include batch number?
N
512
Generate batch number
Y
516
Provide tokens, batch number, and report options

FIG. 5

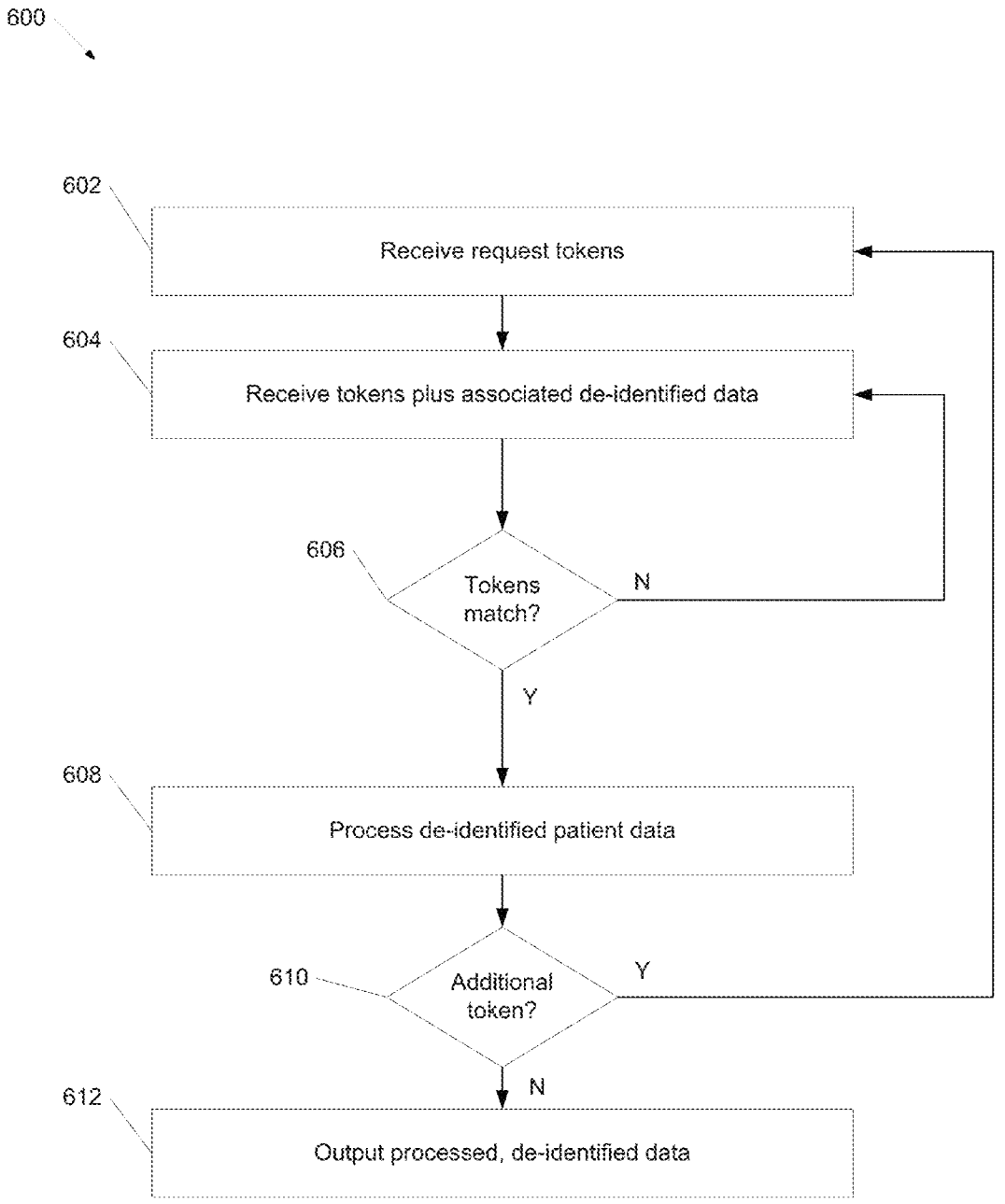

600
602
Receive request tokens
604
Receive tokens plus associated de-identified data
606
Tokens match?
N
Y
608
Process de-identified patient data
610
Additional token?
Y
N
612
Output processed, de-identified data

FIG. 6

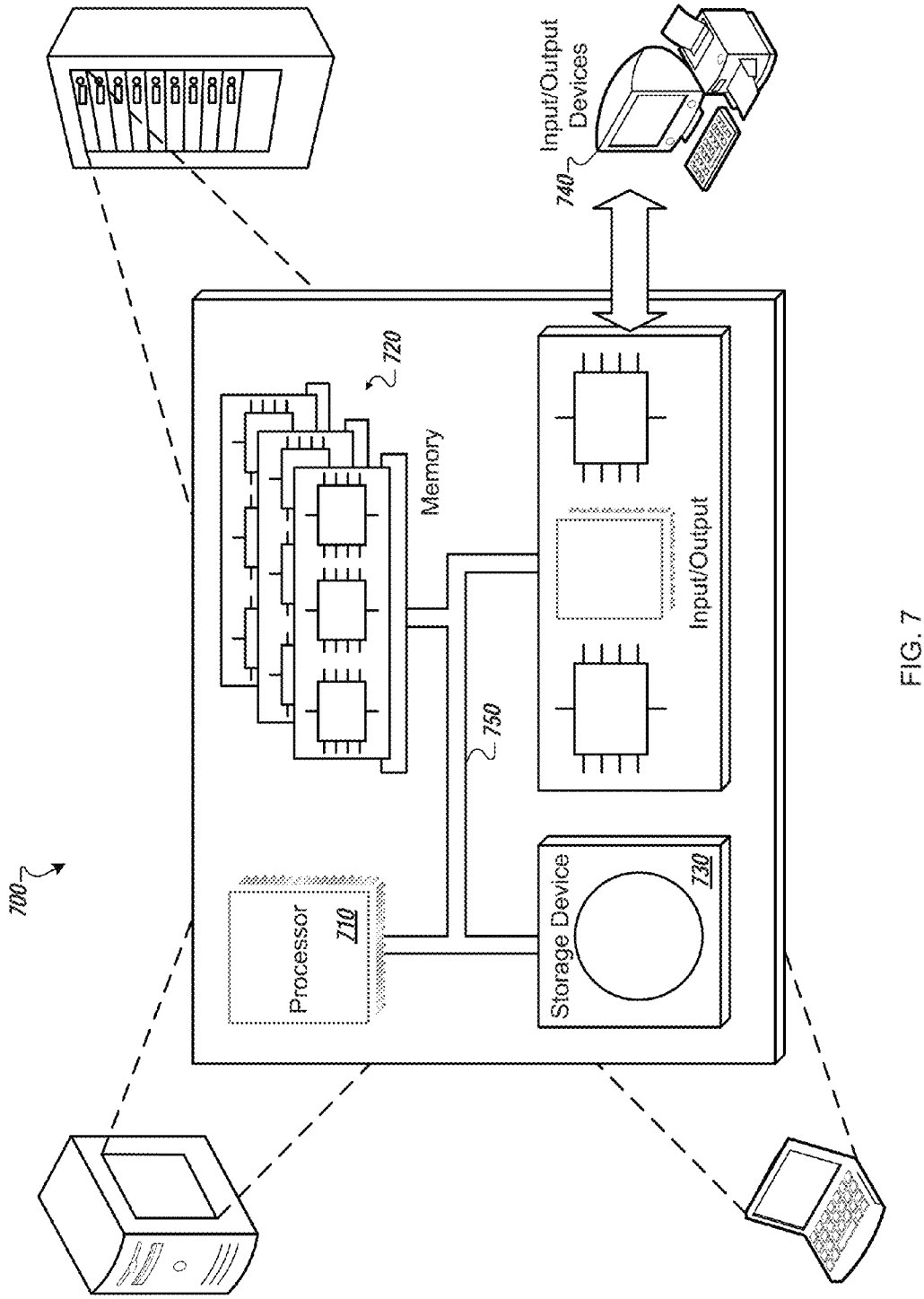

Input/Output Devices
740
720
Memory
Input/Output
750
Storage Device
730
Processor
710
700

FIG. 7

DE-IDENTIFYING MEDICAL HISTORY INFORMATION FOR MEDICAL UNDERWRITING

CLAIM OF PRIORITY

This application claims priority under 35 USC §119(e) to U.S. Patent Application Ser. No. 61/222,428, filed on Jul. 1, 2009, the entire contents of which are hereby incorporated by reference.

BACKGROUND

The present disclosure relates to processing and transmitting personal data, the dissemination of which is restricted by federal law.

Due to federal privacy laws, companies that offer health insurance to small and midsize groups (containing 10-1000 individuals) are often unable to obtain the health history data they need to estimate the risk of insuring such groups commonly referred to as "experience rating". This frequently leads to less than optimal pricing for the companies, the groups, or both. Healthcare providers such as pharmacies and hospitals generate private healthcare data about patients, including medical and prescription drug history, and administrative healthcare claims data. Data that associates patient identity with health information is known as protected health information (PHI). Healthcare providers can store protected health information in electronic databases for future use in patient care and insurance claims processing. Insurance companies have developed techniques to estimate their risk from insuring a group of people by processing the protected health information about the group. Federal privacy laws, however, prevent the insurance companies from obtaining protected health information without individual authorizations from the each person in the group.

SUMMARY

The systems and techniques described here relate to de-identifying medical history information.

In one aspect, a computer-implemented method includes producing information that characterizes a group of individuals from a set of private data representing characteristics of the individuals. The identity of the individuals is unattainable from the produced information. The method also includes providing the produced information to report the characteristics of the group.

Implementations may include any of all of the following features. Producing information that characterizes the group may include producing a request token for each individual included in the group. Producing such a request token for each individual may include encrypting respective data that identifies each individual. Producing information that characterizes the group may include comparing the request tokens to tokens associated with the information to be produced. The tokens associated with the information to be produced and the request tokens may be similarly encrypted. Producing information that characterizes the group may includes determining if the comparison provides at least a minimum number of matches. Producing information that characterizes the group may include requesting a predefined portion of the information. Additionally, producing information that characterizes the group may include determining if the group includes at least a minimum number of individuals. The private data may represent medical related information associated with the individuals of the group.

In another aspect a system includes an encryption server for producing a request token of each individual included in a group identified in a request for information that characterizes the group. The system may also include a data server for producing the information that characterizes the group from a set of private data representing characteristics of the individuals. The identity of the individuals is unattainable from the produced information. The data server is also configured to provide the produced information to report the characteristics of the group.

Implementations may include any of all of the following features. The data server may provide a request token for each individual included in the group to produce the information that characterizes the group. The request token for each individual may represent encrypted data that identifies the corresponding individual. The data server may be configured to compare the request tokens to tokens associated with the information to be produced. The tokens associated with the information to be produced and the request tokens may be similarly encrypted. The data server may be configured to determine if the comparison provides at least a minimum number of matches. The request may represent a predefined portion of information to use for producing the information that characterizes the group. The encryption server may be configured to determine if the group includes at least a minimum number of individuals. The private data may represent medical related information associated with the individuals of the group.

In another aspect, one or more computer readable media storing instructions that are executable by a processing device, and upon such execution cause the processing device to perform operations that include producing information that characterizes a group of individuals from a set of private data representing characteristics of the individuals. The identity of the individuals is unattainable from the produced information. The operations also include providing the produced information to report the characteristics of the group.

Implementations may include any of all of the following features. Producing information that characterizes the group may include producing a request token for each individual included in the group. Producing such a request token for each individual may include encrypting respective data that identifies each individual. Producing information that characterizes the group may include comparing the request tokens to tokens associated with the information to be produced. The tokens associated with the information to be produced and the request tokens may be similarly encrypted. Producing information that characterizes the group may includes determining if the comparison provides at least a minimum number of matches. Producing information that characterizes the group may include requesting a predefined portion of the information. Additionally, producing information that characterizes the group may include determining if the group includes at least a minimum number of individuals. The private data may represent medical related information associated with the individuals of the group.

The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

DESCRIPTION OF DRAWINGS

FIG. **1** illustrates exemplary circumstances in which protected health information is stored by healthcare providers, claims clearing houses, and other source sites, and requested by an insurer.

FIG. **2** illustrates an exemplary method and system that enables an insurer to obtain medical information about a group of people without violating the privacy of persons in the group.

FIG. **3** illustrates an exemplary system incorporating an encryption server and a de-identified data server to enable a user who is not permitted to obtain the private data of a group of people to instead obtain a report that characterizes the group as a whole.

FIG. **4** illustrates an exemplary message requesting de-identified data about a group of people.

FIG. **5** is a flowchart that represents exemplary operations of a token generator.

FIG. **6** is a flowchart that represents exemplary operations of a token matcher.

FIG. **7** represents a computer system and related components.

Like reference symbols in the various drawings indicate like elements.

DETAILED DESCRIPTION

Referring to FIG. **1**, when a doctor treats Jack **102** for high blood pressure, a surgeon removes Jack's gallstones, and a pharmacist fills Jack's prescription for insulin, Jack leaves a trail of electronic records **104** with healthcare professionals (e.g., healthcare profession **106**), in their offices (e.g., doctor's office, nurse's station, etc.), in healthcare facilities (e.g., a hospital **108**, a pharmacy **110**, a nursing home, etc.) and the like. The electronic records **104** contain medical data **112** about Jack, for example his illnesses **114** and treatments **116**. Each piece of medical data **112** is associated with personally identifiable information **118** that identifies Jack and distinguishes him from all other patients, such as his first name **120**, last name **122**, date of birth **124**, gender **126**, and zip code **128**. Together, the pieces of medical data **112** and the personally identifiable information **118** make up Jack's protected health information (PHI) **130**.

Each of Jack's healthcare providers may submit health insurance claims **105** containing PHI **130**, as well as additional PHI **130**, to a claims clearing house **137**. The claims clearing house **137** may store the PHI **130** of many patients **134**, including Jack, in a PHI database **136**. A de-identifier **142** can process the PHI **136** to generate irreversibly de-identified data **140** by removing all personally identifiable information **118** or otherwise transforming the PHI **130** so that it cannot be associated with a particular person. A claims warehouse **139** stores de-identified data **140** about many patients **134**. Claims clearing houses **137** and claims warehouses **139** are optimized for retrieving and providing PHI **136** and de-identified data **140** for use in further processing, but health care providers such as healthcare professional **106** and pharmacies **110** may also serve as source sites **138** for de-identified data in a distributed system.

By searching for Jack's personally identifiable information **118** in its electronic records **104**, the pharmacy **110** is able to look up the various drugs in Jack's PHI **130** and, for example, check for bad interactions among them. Jack does not object to this use of his PHI **130** because it improves the quality of his healthcare. On the other hand, Jack would object to the pharmacy **110** giving his PHI **130** to third parties without his permission because that would disclose personal, private information about him that Jack's potential employers, for example, might use to discriminate against him.

Federal laws, such as the Health Insurance Portability and Accountability Act (HIPAA), protect Jack by prohibiting the source sites **138** possessing Jack's PHI **130** from releasing it to third parties without Jack's permission. Under HIPAA, the source sites **138** can release only irreversibly de-identified data **140** without Jack's permission.

A health insurer **150**, for example, may be interested in the medical history information contained in Jack's PHI **130**. Jack's employer **152** may like to buy a group health insurance policy for the group **154** made up of Jack **102** and his coworkers **156**. If the group **154** is too small, the insurance company **150** will not have the knowledge to fully understand the potential risk from future medical claims for the group **154** and need to set the rate to cover this unknown risk. In that case, the insurer **150** would like to use the PHI **136** from the group **154** to assess the risk of claims and set the group health insurance premium appropriately. Unfortunately, the group **154** may also be too large for the insurance company **150** to practically obtain permission from each person in the group, without which HIPAA prohibits the healthcare providers **138** from releasing the PHI **136** to the insurer **150**. The de-identified data **140**, which the healthcare providers or other source sites **138** could release without permission, is not useful to the insurer **150** because the insurer has no way to know whether it corresponds to the people in the group **154**. With no way to obtain medical history data for the group, the insurer **150** cannot set the group insurance premium acceptably.

Referring to FIG. **2**, to provide the insurer **150** with the medical history information necessary to calculate the insurance premium for the group **154**, without revealing the protected health information of persons in the group **154**, the exemplary system **200** associates unique tokens **202** with the irreversibly de-identified data **140**. The tokens **202** correspond to persons treated by healthcare providers but do not reveal the identities of the treated individuals. The insurer **150** can send a data request **203** to a third party **205**, who in turn can generate request tokens **204**, one for each person in the group **154**. The third party **205** can send the request tokens **204** to a data aggregator **206**. The data aggregator **206** stores de-identified data **140** and associated tokens **202**. By searching for the request tokens **204** among the stored tokens **202**, the data aggregator **206** can process the de-identified data to generate requested de-identified data **208** that corresponds to the group **154**. A data processor **210** can process the requested de-identified data **208** to generate a report **212** containing metrics such as underwriting scores that are useful to the insurer **150** in assessing the overall risk of insuring the group **154**.

By using the tokens **202** and request tokens **204**, no parties other than the authorized source sites **138** can associate de-identified data **140** with the identity of any person in the group of patients **134** or insureds (e.g., the group **154**). The insurer **150** requesting the report **212** can never receive data associated with individuals. Further, the information in the report **212** may be processed into metrics that characterize a large group and cannot be used to infer information about individuals. De-identified data records may be provided to the data processor **210** (e.g., a third part data processor), but that third party may not have access to any personally identifiable information **118** about the group **154**. Nor may any party with access to the de-identified data **140**, other than the healthcare providers **138**, also have access to the de-identifier **142**. These features of the system **200** maintain the privacy of the protected health information **130**.

While we describe a system in which an insurer **150** needs to estimate the risk of insuring a group of potential customers, the system **200** can work for applications in which information characterizing a group needs to be generated from the private data of group members. In one arrangement, the system may implement Microsoft Windows-based computers in

connection with internet-based components. However, other implementations may use other types of components that support the processing of medical history information from healthcare databases.

Referring to FIG. **3**, in an exemplary system **300**, a user **302** specifies a group of people using identifying data for each group member. Referring briefly to FIG. **4**, an exemplary set of identifying data **404** is illustrated. Returning to FIG. **3**, the user may obtain the personally identifiable information from a database **304** by retrieving it in the form of a group file **330**. The user **302** generates a request message, which may be a computer file, which uniquely identifies each member of the group using the identifying data **304**. Typically, the group contains at least a minimum number of unique members. A request message **306** may be generated at a remote computer operated by the user **302**.

The request message **306** contains a set of identifying data **404** (see FIG. **4**) for each group member, as well as data that identifies a user **402** (see FIG. **4**), such as the user's name, email address, phone number and facsimile number, and a batch number that identifies the request. The group members' identifying data **404** (see FIG. **4**) may include first name, last name, date of birth, gender, and zip code as well as other identifiers, such as social security number, that may be added to ensure that members are uniquely identified. In some examples, fewer identifiers may be used. The request message **306** may also include optional indications that the user **302** requests a particular type of report **212** or requests additional processing to enhance the value of the report.

The request message **306** is sent over a communications network **310** (e.g., the Internet, a LAN, etc.) to an encryption server **312**. If the request message **306** contains a minimum number of group members, the encryption server **312** creates a unique token, or identifier, for each person in the message **306**. The encryption server creates each token by applying a token generator **314** that encrypts the personally identifiable information of each group member. The set of tokens corresponding to all the group members constitutes a batch of request tokens **204**. The minimum number of group members, for example ten, is chosen to make it effectively impossible to associate individual group members with individual tokens in the batch of request tokens **204**.

The encryption server **312** provides the request tokens **204** to the de-identified data server **318**. The de-identified data server **318** stores records of de-identified data and corresponding tokens **320** obtained from source sites **138** such as pharmacies, healthcare professionals and electronic claims clearing houses. Each token obtained from the source sites **138** may have been created using the same token generator **314** used by the encryption server **312**, or using any other means that generates the identical token for the same personally identifiable information.

For example, a pharmacy tracks the prescription histories of the patients being served. De-identified prescription histories and corresponding tokens are sent to the de-identified data server **318**. At the request of a user **302**, such as an insurance company, the encryption server **312** generates tokens identical to the patients' tokens using the same personally identifiable information. A unique token corresponding to the same personally identifiable information permits the pharmacy and the insurance company to refer to the same anonymous people without the insurance company ever associating protected health information with a particular person. To produce tokens and request tokens, one or more encryption techniques may be utilized, for example, hash functions and other methodologies may be implemented.

A token matcher **326**, executed by the de-identified data server **318**, performs a look-up in the de-identified database **320** to find all tokens in the database that match the tokens sent from the encryption server **312**. All available data for matched tokens, the requested de-identified data **208**, is retrieved for use in the report generator **328**. The de-identified data may include, for example, prescription history, medical claims, and hospital claims. The de-identified data server **318**, or the data processor **210** may use the report generator **328** to process the requested de-identified data **208** in a way that leaves it irreversibly de-identified. An example of such processing is an underwriting algorithm that transforms the data into an underwriting assessment. The report generator formats the processed data into an electronic or hardcopy report **212** that is returned to the user.

While the report **212** is described as generated on a computer system, it may also be generated in part or entirely outside the computer system. For example, the report **212** could be conveyed to the user via regular mail or other similar technique. In particular, the report may be generated and printed at the site of the de-identified data server **318** and subsequently communicated to the user **302** without using the computer network. The report generator **328** may also reside in the data processor **210** separate from the de-identified data server **318**.

In one arrangement, once the relevant information and options have been selected, the user **302** submits the request by clicking a submit button. The request message **306** may be encrypted prior to being transmitted over a computer network **310**. At the encryption server **312**, the request message **306** is unencrypted and stored. The encryption server **312** may send an optional confirmation message to the user **302**. The confirmation message may include the time and date that the message was received, and may indicate the service level and options selected by the user **302**.

Referring to FIG. **4**, the user **302** can input the user's identifying information **402**, including name, email address, phone number and/or facsimile number, in the first record, or set of fields, of the request message **306**. The user can then input the personally identifiable information **404** for each of the group members, including first name, last name, gender, date of birth, and zip code. Equivalently, the user can incorporate the group file **330** into the request message **306**. Information about the insurance policy **408** for which the group has applied, or other insurance purpose for which the report **212** is being requested may also be included in the request message **306**. The format of the request message **306** may be adjusted based on the user's **302** needs.

The request message **306** may also include information about the level of service **410** requested by the user **302**, for example the quantity, quality or type of information. A first level of service may request up to six months of medical history; a second level may request up to twelve months of history; and a third level may request a two year medical history. Alternatively, instead of providing the user **302** with a variety of service level options **410**, the system **300** may simply retrieve all of the medical history information available for the group.

The user **302** may also request additional, optional information **412**. For example, the user **302** may request information regarding the drug categories and drug indications associated with the drugs in the de-identified data. Drug indications include the medical conditions associated with each drug. Drug categories include the type of drug. This data can be passed to the data processor **210** to include in the report **212**. Alternatively, this data may be returned as part of every report **212**.

Referring to FIG. **5**, a flowchart **500** represents a particular arrangement of operations of a token generator (e.g., the token generator **314** shown in FIG. **3**). Operations include receiving **502** the request message (e.g., request message **306**) containing the personally identifiable information of each person in the group (e.g., such as the group **154** shown in FIG. **1**). Upon receiving the request message, operations may also include determining **504** whether the message includes the minimum number of persons. If the message does not include the minimum number of persons, the token generator returns to receive **502** a request message. A limit on the minimum number of persons ensures that a report (e.g., the report **212**) covers enough people such that it is difficult (if not impossible) to infer any association between particular persons and the information in the report. If the message does include the minimum number of people, operations of the token generator include producing **508** tokens from the identifying data of each person. Producing the tokens may include applying an encryption algorithm to the identifying data of each person. The tokens uniquely identify each person in the group **154**.

Operations also include determining **510** if the request message includes a batch number. If the message does not include a batch number, operations include generating **512** a batch number. Operations also include providing **516** the tokens, batch number and report options. The individual tokens are placed in a batch file and may be encrypted before being transmitted over the network.

Each batch file of request tokens **204** also specifies the information needed for the report **212**. The request tokens are transmitted to the de-identified data server **318**, unencrypted and processed using rules for searching, matching, and retrieving healthcare data.

Referring to FIG. **6**, a flowchart **600** represents a particular arrangement of operations of a token matcher (e.g., the token matcher **326** shown in FIG. **3**). Operations include receiving **602** request tokens (e.g., such as the request tokens **204** shown in FIG. **2**). Operations may also include receiving **604** a token (e.g., such as the token **202**) and associated de-identified data (e.g., such as the de-identified data **140**), typically from source sites (e.g., such as the source site **138** shown in FIG. **1**). Upon receiving a request token, de-identified data and a token, operations include determining **606** whether the request token matches the token. If no match is found, operations include returning to receive **604** more tokens and de-identified data. If a match is found, operations include processing **608** the de-identified data. Processing may include storage for later retrieval. In some arrangements, operations may also include determining whether a minimum number of matches has been detected. If such a minimum number of matches has not occurred in comparing the request tokens and the tokens, action may be taken (e.g., pause or restart processing) until the predefined number of matches has been detected (e.g., so as not to increase the probability of one or more individuals being identified by a process of elimination).

The de-identified data (e.g., such as the de-identified data **140**) may include a list of drugs prescribed over the requested period for the members of the group (e.g., such as the group **154**). The list of drugs prescribed may include the drug name, form, strength, days supplied, and date dispensed. As part of processing **608** the de-identified data, operations may include determining the drug category and drug indications for each drug prescribed. Operations may also include accessing a database relating the drug category and indications to each possible drug. The database may be maintained within the de-identified data server **318** database, or may be accessed on a remote server maintained by a third party.

Upon processing the de-identified data, operations also include determining **610** if additional tokens remain in the batch of request tokens (e.g., such as the request tokens **204**). If additional tokens remain, operations include receiving **602** more request tokens. If there are no additional request tokens, operations may include outputting **612** the processed, de-identified data.

Operations may also include providing the number of tokens submitted, the number matched, and the overall match rate. The collected data covers the interval of historical data according to the level of service requested (e.g., as represented by the level of service **410** in FIG. **4**). De-identified diagnosis and procedures data may also be included from administrative medical claims from a doctor's office (e.g., office **106**) or a healthcare facility (e.g., the hospital **108**).

In addition to accessing and incorporating drug indication information for each drug prescribed to persons in the group (e.g., such as group **154**), operations may include further processing of the requested de-identified data. For instance, operations may include determining the probability that a particular drug indicates a particular condition. In this example, in addition to providing the possible indications, the requested de-identified data would include the likelihood that anonymous individuals associated with the request tokens (e.g., such as request tokens **204**) have each of the conditions indicated by the prescribed drugs. Operations may also include using expert rule systems to provide health status information based on the prescription drug history information. Alternatively, operations may include using diagnosis codes from medical claims data to assess health status.

The requested de-identified data **208** may be sent for further processing to a third party data processor (e.g., data processor **210**) who may apply proprietary algorithms, modify the data format, or generate additional reports, provided that no re-identifiable information is transmitted to the user **302**. Third parties may not have access to the request message **306** and the group file **330** so that no association may be inferred between the de-identified data **140** and particular persons in the group **154**.

The report **212** provides the insurer **150** with information for making an immediate, informed decision about the insurance related risks. In particular, the insurer **150** may accept, reject or adjust the group's insurance rating depending on the information in the report **212**. Actuarial tables and formulas may be used to determine which of the insurance actions are taken. The report **212** may be used alone to make decisions about the insurability of the group **154**, or may simply indicate that additional investigation is needed.

FIG. **7** is a schematic diagram of a generic computer system **700**. The system **700** can be used for the operations described in association with any of the computer-implemented methods described previously, according to one implementation. The system **700** includes a processor **710**, a memory **720**, a storage device **730**, and an input/output device **740**. Each of the components **710**, **720**, **730**, and **740** are interconnected using a system bus **750**. The processor **710** is capable of processing instructions for execution within the system **700**. In one implementation, the processor **710** is a single-threaded processor. In another implementation, the processor **710** is a multi-threaded processor. The processor **710** is capable of processing instructions stored in the memory **720** or on the storage device **730** to display graphical information for a user interface on the input/output device **740**.

The memory **720** stores information within the system **700**. In some implementations, the memory **720** is a computer-

readable medium. The memory **720** is a volatile memory unit in some implementations and is a non-volatile memory unit in other implementations.

The storage device **730** is capable of providing mass storage for the system **700**. In one implementation, the storage device **730** is a computer-readable medium. In various different implementations, the storage device **730** may be a floppy disk device, a hard disk device, an optical disk device, or a tape device.

The input/output device **740** provides input/output operations for the system **700**. In one implementation, the input/output device **740** includes a keyboard and/or pointing device. In another implementation, the input/output device **740** includes a display unit for displaying graphical user interfaces.

The features described can be implemented in digital electronic circuitry, or in computer hardware, firmware, software, or in combinations of them. The apparatus can be implemented in a computer program product tangibly embodied in an information carrier, e.g., in a machine-readable storage device, for execution by a programmable processor; and method steps can be performed by a programmable processor executing a program of instructions to perform functions of the described implementations by operating on input data and generating output. The described features can be implemented advantageously in one or more computer programs that are executable on a programmable system including at least one programmable processor coupled to receive data and instructions from, and to transmit data and instructions to, a data storage system, at least one input device, and at least one output device. A computer program is a set of instructions that can be used, directly or indirectly, in a computer to perform a certain activity or bring about a certain result. A computer program can be written in any form of programming language, including compiled or interpreted languages, and it can be deployed in any form, including as a stand-alone program or as a module, component, subroutine, or other unit suitable for use in a computing environment.

Suitable processors for the execution of a program of instructions include, by way of example, both general and special purpose microprocessors, and the sole processor or one of multiple processors of any kind of computer. Generally, a processor will receive instructions and data from a read-only memory or a random access memory or both. The essential elements of a computer are a processor for executing instructions and one or more memories for storing instructions and data. Generally, a computer will also include, or be operatively coupled to communicate with, one or more mass storage devices for storing data files; such devices include magnetic disks, such as internal hard disks and removable disks; magneto-optical disks; and optical disks. Storage devices suitable for tangibly embodying computer program instructions and data include all forms of non-volatile memory, including by way of example semiconductor memory devices, such as EPROM, EEPROM, and flash memory devices; magnetic disks such as internal hard disks and removable disks; magneto-optical disks; and CD-ROM and DVD-ROM disks. The processor and the memory can be supplemented by, or incorporated in, ASICs (application-specific integrated circuits).

To provide for interaction with a user, the features can be implemented on a computer having a display device such as a CRT (cathode ray tube) or LCD (liquid crystal display) monitor for displaying information to the user and a keyboard and a pointing device such as a mouse or a trackball by which the user can provide input to the computer.

The features can be implemented in a computer system that includes a back-end component, such as a data server, or that includes a middleware component, such as an application server or an Internet server, or that includes a front-end component, such as a client computer having a graphical user interface or an Internet browser, or any combination of them. The components of the system can be connected by any form or medium of digital data communication such as a communication network. Examples of communication networks include, e.g., a LAN, a WAN, and the computers and networks forming the Internet.

The computer system can include clients and servers. A client and server are generally remote from each other and typically interact through a network, such as the described one. The relationship of client and server arises by virtue of computer programs running on the respective computers and having a client-server relationship to each other.

A number of implementations have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the following claims.

What is claimed is:

1. A computer-implemented method comprising:

receiving a request for healthcare information that characterizes a group composed of multiple individuals, the healthcare information being assembled from healthcare data that comprises de-identified information;

producing, by a processor of an encryption server, tokens, wherein the identity of the individuals in the group is unattainable from the tokens, each token being associated with an individual in the group, each token being associated with a corresponding token in a set of de-identified healthcare data, and each token in the set of de-identified healthcare data being associated with healthcare data for a corresponding individual;

when the group includes at least a minimum number of individuals, producing the requested healthcare information that characterizes the group by identifying the tokens in the set of de-identified healthcare data and the associated de-identified healthcare data that correspond to the produced tokens, wherein the produced tokens and the tokens in the set of de-identified data are similarly encrypted, and the produced healthcare information is absent personally identifiable information and has been produced without the individuals in the group having authorized production of such healthcare information that characterizes the group; and

providing the produced healthcare information to report the characteristics of the group composed of multiple individuals absent authorization at any time from the individuals in the group of individuals to provide the requested healthcare information.

2. The computer-implemented method of claim **1**, wherein the healthcare information provided absent authorization from the group of individuals at any time is in conformity with the Health Insurance Portability and Accountability Act (HIPAA).

3. The computer-implemented method of claim **1**, wherein receiving a request for healthcare information that characterizes a group of individuals from healthcare data includes receiving personally identifiable information associated with members of the group of individuals.

4. The computer-implemented method of claim **1**, in which producing the requested healthcare information includes encrypting respective data that identifies each individual.

5. The computer-implemented method of claim 1, in which producing the requested healthcare information includes identifying at least a minimum number of corresponding tokens.

6. The computer-implemented method of claim 1, in which the healthcare data represents medical related information associated with the individuals of the group.

7. A system comprising:

an encryption server, having a processor and a memory, for producing a token for each individual included in a group composed of multiple individuals identified in a request for healthcare information that characterizes the group, wherein the identity of the individuals in the group is unattainable from the tokens, each token being associated with a corresponding token in a set of de-identified healthcare data, and each token in the set of de-identified healthcare data being associated with healthcare data for a corresponding individual;

a data server for producing the requested healthcare information that characterizes the group by identifying the tokens in the set of de-identified healthcare data and the associated de-identified healthcare data that correspond to the produced tokens, wherein the produced tokens and the tokens in the set of de-identified data are similarly encrypted, and the produced healthcare information is absent personally identifiable information and has been produced without the individuals in the group having authorized production of such healthcare information that characterizes the group; and

providing the produced healthcare information to report the characteristics of the group composed of multiple individuals absent authorization at any time from the individuals in the group of individuals to provide the requested healthcare information.

8. The system of claim 7, wherein the healthcare information provided absent authorization from the group of individuals at any time is in conformity with the Health Insurance Portability and Accountability Act (HIPAA).

9. The system of claim 7, wherein the request for healthcare information that characterizes a group of individuals from healthcare data includes personally identifiable information associated with members of the group of individuals.

10. The system of claim 7, in which producing the requested healthcare information includes encrypting respective data that identifies each individual.

11. The system of claim 7, in which producing the requested healthcare information includes identifying at least a minimum number of corresponding tokens.

12. The system of claim 7, in which producing the requested healthcare information includes determining if the group includes at least a minimum number of individuals.

13. The system of claim 7, in which the healthcare data represents medical related information associated with the individuals of the group.

14. One or more computer readable storage devices storing instructions that are executable by a processing device, and upon such execution cause the processing device to perform operations comprising:

receiving a request for healthcare information that characterizes a group composed of multiple individuals, the healthcare information being assembled from healthcare data that comprises de-identified information;

producing, by a processor of an encryption server, tokens, wherein the identity of the individuals in the group is unattainable from the tokens, each token being associated with an individual in the group, each token being associated with a corresponding token in a set of de-identified healthcare data, and each token in the set of de-identified healthcare data being associated with healthcare data for a corresponding individual;

when the group includes at least a minimum number of individuals, producing the requested healthcare information that characterizes the group by identifying the tokens in the set of de-identified healthcare data and the associated de-identified healthcare data that correspond to the produced tokens, wherein the produced tokens and the tokens in the set of de-identified data are similarly encrypted, and the produced healthcare information is absent personally identifiable information and has been produced without the individuals in the group having authorized production of such healthcare information that characterizes the group; and

providing the produced healthcare information to report the characteristics of the group composed of multiple individuals absent authorization at any time from the individuals in the group of individuals to provide the requested healthcare information.

15. The computer readable storage devices of claim 14, wherein the healthcare information provided absent authorization from the group of individuals at any time is in conformity with the Health Insurance Portability and Accountability Act (HIPAA).

16. The computer readable storage devices of claim 14, wherein the request for healthcare information that characterizes a group of individuals from healthcare data includes personally identifiable information associated with members of the group of individuals.

17. The computer readable storage devices of claim 14, in which producing the requested healthcare information includes encrypting respective data that identifies each individual.

18. The computer readable storage devices of claim 14, in which producing the requested healthcare information includes identifying at least a minimum number of corresponding tokens.

19. The computer readable storage devices of claim 14, in which the healthcare data represents medical related information associated with the individuals of the group.

* * * * *