

(43) International Publication Date
20 December 2012 (20.12.2012)(10) International Publication Number
WO 2012/173826 A2

(51) International Patent Classification:

G06Q 30/06 (2012.01)Redmond, WA 98052-6399 (US). **ROBINSON, David**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, WA 98052-6399 (US).

(21) International Application Number:

PCT/US2012/040951

(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, 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, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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.

(22) International Filing Date:

5 June 2012 (05.06.2012)

(25) Filing Language:

English

(26) Publication Language:

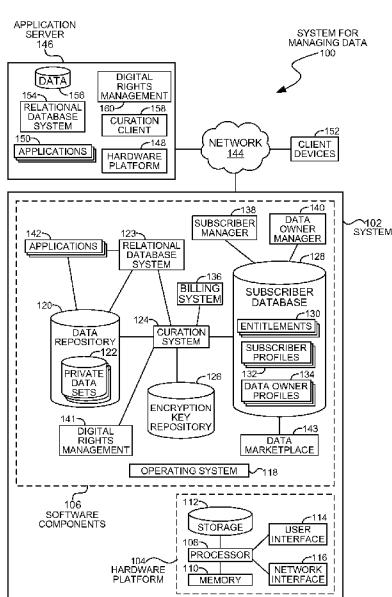
English

(30) Priority Data:

13/159,658 14 June 2011 (14.06.2011) US

(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, 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, ML, MR, NE, SN, TD, TG).(71) Applicant (for all designated States except US): **MICROSOFT CORPORATION** [US/US]; One Microsoft Way, Redmond, WA 98052-6399 (US).(72) Inventors: **SIVARAMAKRISHNAN, Hari**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, WA 98052-6399 (US). **DSOUZA, Roy**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, WA 98052-6399 (US). **NOVIK, Lev**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, WA 98052-6399 (US). **BICE, Nino**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way,*[Continued on next page]*

(54) Title: DATA CUSTODIAN AND CURATION SYSTEM

(57) **Abstract:** A data custodian and curation system may store data from a data supplier in encrypted form and may allow users to consume the data when the consumers obtain access to the data through an agreement. The curation system may manage access to the data, even allowing a consumer to use the data as if it were their own, but may track each usage of the data to implement a payment scheme or various usage restrictions. The curation system may encrypt downstream uses of the data and operate as a digital rights management system for the data. The custodian and curation system may operate as a cloud service that may contain encrypted data from many data providers but where the data providers may individually control access to their data in a managed system at any granular level.**FIG. 1**



Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

- *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

DATA CUSTODIAN AND CURATION SYSTEM

Background

[0001] Data are a valuable commodity, but access to valuable data may be difficult to manage. When data are sold to a consumer, the data may be used by the consumer over and over, or even combined with other data and resold without the control or knowledge of a data owner. This may be because the data comes into the possession of the consumer and may not be traceable or controllable after possession is transferred.

[0002] In such cases, the data owner may only be able to sell or monetize the data in a single transaction, but not be able to monetize the data for other use scenarios.

10

Summary

[0003] A data custodian and curation system may store data from a data supplier in encrypted form and may allow users to consume the data when the consumers obtain access to the data through an agreement. The curation system may manage access to the data, even allowing a consumer to use the data as if it were their own, but may track each usage of the data to implement a payment scheme or various usage restrictions. The curation system may encrypt downstream uses of the data and operate as a digital rights management system for the data. The custodian and curation system may operate as a cloud service that may contain encrypted data from many data providers but where the data providers may individually control access to their data in a managed system at any granular level.

[0004] This Summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This Summary is not intended to identify key features or essential features of the claimed subject matter, nor is it intended to be used to limit the scope of the claimed subject matter.

25

Brief Description of the Drawings

[0005] In the drawings,

[0006] FIGURE 1 is a diagram of an embodiment showing a network environment with a data management system.

[0007] FIGURE 2 is a flowchart of an embodiment showing a method for making data available by a data owner.

[0008] FIGURE 3 is a flowchart of an embodiment showing a method for setting up data access.

[0009] FIGURE 4 is a flowchart of an embodiment showing a method for permitting access with a curation system.

Detailed Description

[0010] A data management system may permit users to purchase and use data while 5 keeping the data within a controlled system. The data may be encrypted and managed so that each access to the data may be monitored and controlled. In many embodiments, the source data may be maintained and tracked within the data repository so that access by other parties may be monitored and controlled.

[0011] The data management system may be considered an ecosystem in which data may 10 be shared and consumed while maintaining the data in a secured manner. Each portion of data may be combined with other data sets to deliver data intensive applications. Each data source may be maintained in its original location and state, and each access to the data source may be tracked for billing and access control.

[0012] The data management system may have a relational database management system 15 that various applications may use. Within the relational database management system, application databases may be created that access tables, columns, rows, elements, or other database components that may be owned by data owners. Each of the database components may be separately encrypted and a curation system may provide keys to decrypt the data when requested.

[0013] The ecosystem may enable data to be shared at different granularities and for 20 different conditions, each of which may be managed by an entitlement or agreement to allow access. For example, an entitlement may permit access to a specific table of data on a per-access price. Another entitlement may permit access to a column of data for a predetermined period of time, regardless of the number of accesses.

[0014] Data provenance may be provided to a consumer when an application uses data 25 provided from a data owner. The encryption keys used to lock the source data may be provided to the consumer as evidence that the source data did indeed come from the source. Such provenance may be maintained even when multiple parties process the data within the ecosystem.

[0015] Throughout this specification, like reference numbers signify the same elements 30 throughout the description of the figures.

[0016] When elements are referred to as being “connected” or “coupled,” the elements can be directly connected or coupled together or one or more intervening elements may also be

present. In contrast, when elements are referred to as being “directly connected” or “directly coupled,” there are no intervening elements present.

[0017] The subject matter may be embodied as devices, systems, methods, and/or computer program products. Accordingly, some or all of the subject matter may be embodied in hardware and/or in software (including firmware, resident software, micro-code, state machines, gate arrays, etc.) Furthermore, the subject matter may take the form of a computer program product on a computer-usable or computer-readable storage medium having computer-usable or computer-readable program code embodied in the medium for use by or in connection with an instruction execution system. In the context of this document, a computer-usable or computer-readable medium may be any medium that can contain, store, communicate, propagate, or transport the program for use by or in connection with the instruction execution system, apparatus, or device.

[0018] The computer-usable or computer-readable medium may be, for example but not limited to, an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, device, or propagation medium. By way of example, and not limitation, computer readable media may comprise computer storage media and communication media.

[0019] Computer storage media includes volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage of information such as computer readable instructions, data structures, program modules or other data. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to store the desired information and which can be accessed by an instruction execution system. Note that the computer-usable or computer-readable medium could be paper or another suitable medium upon which the program is printed, as the program can be electronically captured, via, for instance, optical scanning of the paper or other medium, then compiled, interpreted, or otherwise processed in a suitable manner, if necessary, and then stored in a computer memory.

[0020] Communication media typically embodies computer readable instructions, data structures, program modules or other data in a modulated data signal such as a carrier wave or other transport mechanism and includes any information delivery media. The term “modulated data signal” means a signal that has one or more of its characteristics set

or changed in such a manner as to encode information in the signal. By way of example, and not limitation, communication media includes wired media such as a wired network or direct-wired connection, and wireless media such as acoustic, RF, infrared and other wireless media. Combinations of the any of the above should also be included within the
5 scope of computer readable media.

[0021] When the subject matter is embodied in the general context of computer-executable instructions, the embodiment may comprise program modules, executed by one or more systems, computers, or other devices. Generally, program modules include routines, programs, objects, components, data structures, etc. that perform particular tasks or
10 implement particular abstract data types. Typically, the functionality of the program modules may be combined or distributed as desired in various embodiments.

[0022] Figure 1 is a diagram of an embodiment 100, showing a system 102 that may provide a data ecosystem in which data may be shared while maintaining control of the data by the data owner. Embodiment 100 is a simplified example of a data ecosystem that
15 manages data.

[0023] The diagram of Figure 1 illustrates functional components of a system. In some cases, the component may be a hardware component, a software component, or a combination of hardware and software. Some of the components may be application level software, while other components may be operating system level components. In some
20 cases, the connection of one component to another may be a close connection where two or more components are operating on a single hardware platform. In other cases, the connections may be made over network connections spanning long distances. Each embodiment may use different hardware, software, and interconnection architectures to achieve the described functions.

25 [0024] Embodiment 100 illustrates one version of a system for managing data. Data may be sold by-the-use, while maintaining control and provenance of the data. An encryption system may encrypt each salable unit of data and access to the data may be monitored and controlled by each request for a key to decrypt the data.

30 [0025] Various data owners may make their data available to other users through an entitlement, which may entitle a user to access the data according to various predefined rules. The entitlement may be considered an access contract that defines the metes and bounds of the data the user may access.

[0026] The data may be maintained in a single location. A relational database system may permit users to create applications that include controlled data in their databases, but where each access to the controlled data may be monitored and logged.

[0027] Throughout this specification and claims, the term “private data” may be any data owned by a data owner for which access may be restricted and managed. In many embodiments, a data owner may make their data available for purchase or for access under restricted conditions.

[0028] In many embodiments, a data user may be a business that executes an application that consumes data. The application may have a database against which queries may be executed, and the database may contain one or more tables, rows, columns, or cells of data that are private data. When the application developer creates their database, the application developer may create links within the application database to private data.

[0029] The private data may be separately encrypted for each unit of data that may be made available. In some embodiments, the unit of data may be a group of tables, a single table, a set of rows or columns in a table, specific cells within a table, or some other unit of data in the database.

[0030] When an element of the private data is accessed, the relational database management system may determine that the data are encrypted, and may request an encryption key. An encryption key repository may contain a key that may be used to decrypt the private data and make the data available.

[0031] A curation system may manage the access to private data. The curation system may identify an access, ensure that the access is permitted by an existing entitlement, then may retrieve a key for decrypting the data. In cases where the access is billed on a per-access basis, the curation system may use a billing system to create a billable event.

[0032] The data encryption may be performed in a format preserving manner in some cases. Format preserving data encryption may result in an encrypted version of a data element where the encrypted version may have the same data type as the original version.

[0033] In some embodiments, some data elements may be encrypted in a non-format preserving manner. Such embodiments may be useful when the data have data types that are not amenable to format preserving encryption or where the data are to be protected using higher levels of security.

[0034] In many embodiments, the system may be able to track derived data which may contain or be derived from private data. When derived data are present, the derived data may be encrypted and managed in the same manner as the original private data. Such

embodiments may preserve the private data even though the private data may be merged or transformed with other data.

[0035] As data are used by one or more applications, the provenance of the data may be verified with the data management system. When each of the transformations of data are preserved using the data management system, a consumer may request a verified provenance of the data. The consumer may receive a public encryption key from the data owner that may be used to verify that the data are indeed from the data owner. The public encryption key may be used to decrypt some or all of the private data, or a token or other identifier associated with the private data. In some instances, such a token may be known as an authentication token.

[0036] Data provenance may be useful in applications where high value data are used. An application that uses data from a trusted and costly source may wish to advertise that their data came from a trusted source, and may be able to verify the source to the consumer.

[0037] Many embodiments may include analytics that may be used by a data owner. The analytics may include who used the data, which data elements were accessed, how and when the data were used, and other information. The analytics data may be useful for monitoring, marketing, and monetizing the data to notice trends, highly used data, or other information.

[0038] The system of embodiment 100 is illustrated as being contained in a single system 102. The system 102 may have a hardware platform 104 and software components 106.

[0039] The system 102 may represent a server or other powerful, dedicated computer system that may support multiple user sessions. In some embodiments, however, the system 102 may be any type of computing device, such as a personal computer, game console, cellular telephone, netbook computer, or other computing device.

[0040] The hardware platform 104 may include a processor 108, random access memory 110, and nonvolatile storage 112. The processor 108 may be a single microprocessor, multi-core processor, or a group of processors. The random access memory 110 may store executable code as well as data that may be immediately accessible to the processor 108, while the nonvolatile storage 112 may store executable code and data in a persistent state.

[0041] The hardware platform 104 may include user interface devices 114. The user interface devices 114 may include keyboards, monitors, pointing devices, and other user interface components.

[0042] The hardware platform 104 may also include a network interface 116. The network interface 116 may include hardwired and wireless interfaces through which the system 102 may communicate with other devices.

[0043] Many embodiments may implement the various software components using a hardware platform that is a cloud fabric. A cloud hardware fabric may execute software on multiple devices using various virtualization techniques. The cloud fabric may include hardware and software components that may operate multiple instances of an application or process in parallel. Such embodiments may have scalable throughput by implementing multiple parallel processes.

10 [0044] The software components 106 may include an operating system 118 on which various applications may execute. In some cloud based embodiments, the notion of an operating system 118 may or may not be exposed to an application.

15 [0045] A data repository 120 may contain many private data sets 122. The private data sets 122 may include a separate data set for each unit of data that may be separately tracked and managed. Each private data set may be encrypted with a different encryption key, and the encryption keys may be stored and managed in an encryption key repository 126.

20 [0046] A curation system 124 may manage the private data sets 122 and provide encryption keys from the encryption key repository 126 to decrypt the various private data sets 122, among other functions. The curation system 124 may operate with a relational database system 123, where the relational database system 123 may identify an encrypted private data set 122 and request access to the curation system 124.

25 [0047] The relational database management system 123 may be accessed by various applications 142 that may use and manipulate one or more of the various private data sets 122.

30 [0048] The curation system 124 may access a subscriber database 128 to determine if any entitlements 130 exist from which access may be granted, then retrieve an encryption key. The curation system 124 may also interact with a billing system 136 to track access to a private data set 122 when an entitlement is in place that causes a billing event to occur with an access.

[0049] In some embodiments, the curation system 124 may operate with twice encrypted keys. In one such embodiment, a data user or data owner may encrypt an encryption key for a private data set with a second encryption key. In such embodiments, the curation system 124 may provide an encryption key that may be encrypted with a second

encryption key known by the data owner or data user. The data owner or data user may then decrypt the encryption key to gain access to the key used to decrypt the data.

[0050] Such embodiments may be useful in protecting access to the private data sets 122 from the operator of the system 102. The double encryption may use a private key known 5 by a data owner or data user to ensure that the operator of the system 102 does not have full access to all the encryption keys in the encryption key repository 126.

[0051] The subscriber database 128 may include subscriber profiles 132 and data owner profiles 134. The subscriber profiles 132 may include login information, billing information, and other data that may be used to create an entitlement 130. Similarly, the 10 data owners may have data owner profiles 134 that may be used to manage the data being made available.

[0052] A subscriber manager 138 may be an interface through which a subscriber may establish and maintain a subscriber profile 132. The subscriber manager 138 may also assist a data user in finding data that may be available for purchase and use. Some 15 embodiments may have a data marketplace 143 which may be a web based or other shopping mechanism whereby a data user may identify the data to be consumed, then the data marketplace 143 may establish an entitlement 130 that gives access to the data user.

[0053] The data owner manager 140 may be an interface through which a data owner may identify data to share, configure the various private data sets 122, and establish terms and 20 conditions for various entitlements. The data marketplace 143 may present the private data sets to prospective data users, as well as the terms and conditions for entitlements.

[0054] In some embodiments, a digital rights management system 141 may permit or deny access to certain private data sets. In some cases, the private data sets 122 may be copied to a second location and used with another relational database management system. In 25 such cases, the digital rights management system 141 may apply protection to the data, and a corresponding digital rights management system may permit or deny access to those data according to the applicable entitlements.

[0055] The system 102 may be connected to a network 144 to which various client devices 152 and other application servers 146 may be attached. A client device 152 may be any 30 type of device that may access one of the applications that may use the private data sets 122. In some cases, the client devices 152 may also be used by a data owner or data user to access the data owner manager 140 or subscription manager 138, respectively.

[0056] The client devices 152 may be any network connected device. In some cases, the client devices 152 may be personal computers, server computers, mobile telephones, game consoles, laptop computers, tablet computers, or any other network enabled device.

[0057] Some embodiments may use an application server 146 to operate in conjunction

5 with the system 102. The application server 146 may have a hardware platform 148, which may be similar to the hardware platform 104 or may be a cloud based fabric.

[0058] The application server 146 may have applications 150 that may access the relational database system 123 of the system 102 or may access a local relational database system 154.

10 [0059] In some embodiments, the application server 146 may request private data sets 122 from the system 102 and may have the requested private data sets 122 transmitted to be stored in local data 156. In such embodiments, the private data sets may remain encrypted during transmission and storage. In order to access the private data sets stored in the local data 156, a curation client 158 may operate in conjunction with the curation system 124 to 15 retrieve an encryption key to decrypt the local data 156 for use. In other embodiments, the private data sets may be transmitted using digital rights management. In such embodiments, a digital rights management system 160 may permit or deny access to the private data sets.

[0060] Figure 2 is a flowchart illustration of an embodiment 200 showing a method for

20 making data available. Embodiment 200 is a simplified example of a method that may be performed by a data owner to identify and prepare private data for use.

[0061] Other embodiments may use different sequencing, additional or fewer steps, and

different nomenclature or terminology to accomplish similar functions. In some

embodiments, various operations or set of operations may be performed in parallel with

25 other operations, either in a synchronous or asynchronous manner. The steps selected here were chosen to illustrate some principles of operations in a simplified form.

[0062] Embodiment 200 is a simplified example of the steps performed when a data owner

may make private data available within a data management system. The operations of

embodiment 200 may be performed using a data owner manager interface, such as the data

30 owner manager 140 of embodiment 100.

[0063] In block 202, an account may be created. The account may include various data

about the data owner, including any access authorization that the data owner may have for

the data that may be shared.

[0064] In block 204, the data owner may create their database. In some embodiments, the database may be preexisting and the data owner may copy the database into a data management system for dissemination.

[0065] The data owner may identify private data sets in block 206. The private data sets 5 may be any unit of data that the data owner may wish to make available through a data management system. In some cases, the private data sets may be one or more tables, columns, rows, individual elements, or some other unit of a database.

[0066] For each private data set in block 208, a separate encryption key may be created in block 210 and used to encrypt the private data set in block 212.

10 [0067] After the private data set is encrypted, the private data set may be made available in a marketplace in block 214. The terms and conditions of any entitlements may be defined in block 216 and also included in the marketplace. The process may return to block 208 to process any additional private data sets.

15 [0068] Figure 3 is a flowchart illustration of an embodiment 300 showing a method for setting up access to a private data set. Embodiment 300 is a simplified example of a method that may be performed by a data user to identify and prepare private data for use.

20 [0069] Other embodiments may use different sequencing, additional or fewer steps, and different nomenclature or terminology to accomplish similar functions. In some embodiments, various operations or set of operations may be performed in parallel with other operations, either in a synchronous or asynchronous manner. The steps selected here were chosen to illustrate some principles of operations in a simplified form.

[0070] Embodiment 300 is a simplified example of the steps performed when a data user may begin using private data. The operations of embodiment 300 may be performed using a subscription interface, such as the subscription manager 138 of embodiment 100.

25 [0071] A data user may establish an account in block 302. The account may include payment information that may be debited to meet the terms and conditions of any entitlement that may be established to private data sets.

30 [0072] The data user may browse available private data sets in block 304 and may establish an agreement in block 306 with the data owner to obtain access. An entitlement may be created in block 308 and stored with a data management system, which may begin permitting the data user to access the data in block 310.

[0073] Figure 4 is a flowchart illustration of an embodiment 400 showing a method for interacting with and operating a curation system. The operations of a data user 402 may be illustrated in the left hand column. The operations of a curation system 404 may be

illustrated in the center column, and the operations of a data repository 406 may be illustrated in the right hand column. The operations illustrated for a data user 402 illustrate operations performed by a device or software under the control of the data user 402.

5 [0074] Other embodiments may use different sequencing, additional or fewer steps, and different nomenclature or terminology to accomplish similar functions. In some embodiments, various operations or set of operations may be performed in parallel with other operations, either in a synchronous or asynchronous manner. The steps selected here were chosen to illustrate some principles of operations in a simplified form.

10 [0075] Embodiment 400 is a simplified example of some steps that may be performed when a data user uses private data. The curation system 404 and data repository 406 may interact to decrypt private data and permit access when the conditions of an entitlement are satisfied.

15 [0076] Embodiment 400 further illustrates a simplified interaction that may be performed in order to provide provenance for private data.

[0077] In block 408, a data user 402 may request access to a data repository 406, which may receive the request in block 410. If the request is not for encrypted data in block 412, the data may be transmitted in block 414 and received in block 416.

20 [0078] The operations of blocks 412 through 416 represent a case where the data stored in the data repository is freely available to the data user. Such a situation may occur when the data being accessed are owned by the data user or when the data are publically available.

[0079] If the data are encrypted in block 412, the data repository 406 may request access in block 418, which may be received by the curation system 404 in block 420.

25 [0080] The curation system 404 may look up any entitlement to determine whether or not the data user 402 may be granted access in block 422. If the data user 402 is granted access, the curation system 404 may retrieve an encryption key in block 424 and transmit the key in block 426 to the data repository 406. The data repository 406 may receive the key in block 428, decrypt the data in block 430, and permit access in block 432. The data user 402 may receive permission in block 434 and access the data in block 436.

30 [0081] In other embodiments, the decryption may be performed by the data user 402. In such an embodiment, the encryption key may be transmitted to the data user 402 and used by the data user 402 to decrypt the data. In some embodiments, the encryption key may be

stored by the curation system 404 as an encrypted object, with the data user 402 having a second encryption key that can be used to decrypt the key held by the curation system 404.

[0082] After the curation system handles the encryption key look up in blocks 422 and 424, the curation system 404 may create a billable event in block 438 and store the billable event in a billing system in block 440.

[0083] In many embodiments, an entitlement may be created that charges a fee based on an access event. Some embodiments may, for example, charge a fee for each use. Other embodiments may charge a fixed fee for a certain number of accesses. Still other embodiments may permit unlimited accesses for a certain period of time after the first access. Various entitlements may be structured with different conditions.

[0084] In block 442, the data user 402 may request provenance for certain data. The request may be received by the data repository 406 in block 444. In response, the data repository 406 may transmit a public key in block 446 to the data user 402, which may receive the public key in block 448 and use the public key to verify against the encryption of the data in block 450.

[0085] The foregoing description of the subject matter has been presented for purposes of illustration and description. It is not intended to be exhaustive or to limit the subject matter to the precise form disclosed, and other modifications and variations may be possible in light of the above teachings. The embodiment was chosen and described in order to best explain the principles of the invention and its practical application to thereby enable others skilled in the art to best utilize the invention in various embodiments and various modifications as are suited to the particular use contemplated. It is intended that the appended claims be construed to include other alternative embodiments except insofar as limited by the prior art.

Claims

1. A system comprising:

a data repository comprising a plurality of private data sets, each of said private data sets comprising a set of data groups, each of said data groups being encrypted using a separate encryption key;

5 said data repository comprising said separate encryption keys for each of said private data sets;

10 a data curation system comprising a database system having a subscriber database comprising subscriber data and a first data group having a first encryption key and being owned by a first data owner, said subscriber having access to said first encryption key through a first entitlement, said data curation system having access to said first encryption key to decrypt said first data group.

2. The system of claim 1, said data curation system managing access to said first data group according to said first entitlement.

15 3. The system of claim 2, said first entitlement defining a limited number of accesses to said first data group, said data curation system limiting access to said first data group to said limited number.

4. The system of claim 1, said data curation system creating a first table comprising said subscriber data and said first data group, said first data group being stored in an encrypted form.

5. The system of claim 4, said first data group comprising a first portion having a first format in an unencrypted form, said first format being encrypted using format preserving encryption such that said encrypted form has said first format.

25 6. The system of claim 5, said first table further comprising said subscriber data in said unencrypted form.

7. The system of claim 6, said database comprising a second table comprising second data derived from said first portion, at least a portion of said second table being encrypted using said first encryption key.

30 8. The system of claim 7, said portion of said second table being derived from said first portion.

9. The system of claim 8, said second table having a data provenance.

10. The system of claim 1 further comprising:

a billing system that identifies an access to said data repository using said first entitlement and generates a billing event based on said access.

11. The system of claim 10, said billing event being detected when said first encryption key is requested.

12. A method comprising:

5 creating a plurality of datasets, each of said datasets being owned by a data owner and stored in a data store;

encrypting each of said plurality of datasets with a separate encryption key;

storing said encryption keys in a data curation system;

establishing a first entitlement between a data user and said data owner, said first entitlement allowing said data user to access a subset of said plurality of datasets;

10 receiving a request for a first dataset from said data user, said request comprising a reference to said first entitlement;

evaluating said first entitlement to determine that said data user has access to said first dataset; and

decrypting said first dataset using said first encryption key.

15 13. The method of claim 12 further comprising:

returning a first encryption key to said data user to decrypt said first dataset.

14. The method of claim 12 further comprising:

detecting that said data user requested said first dataset and logging said request in a billing system.

20 15. The method of claim 12 further comprising:

requesting a provenance from a consumer for said first dataset, said consumer being a user of said first dataset as provided by said data user; and

returning an authentication token for said first dataset.

1/4

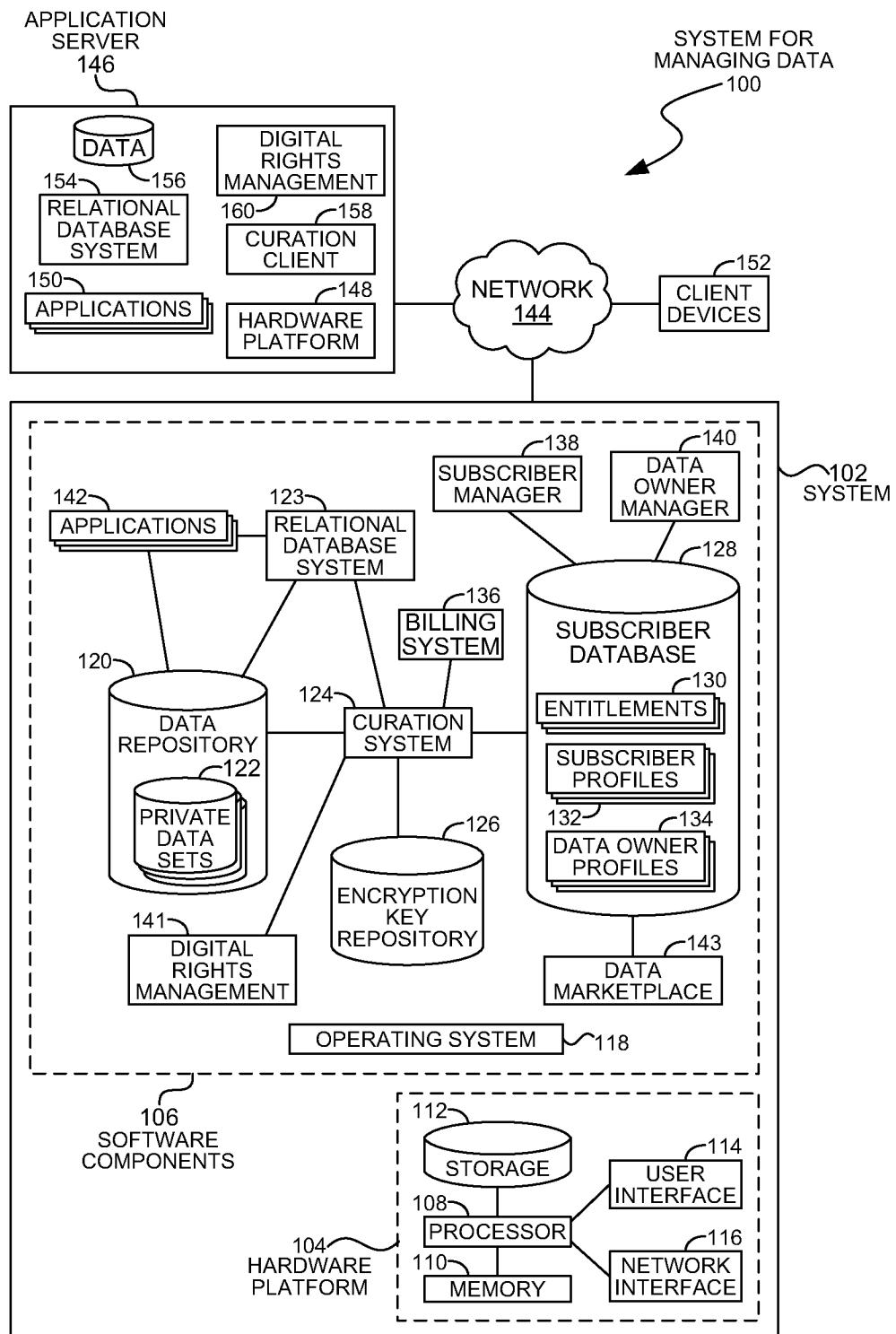
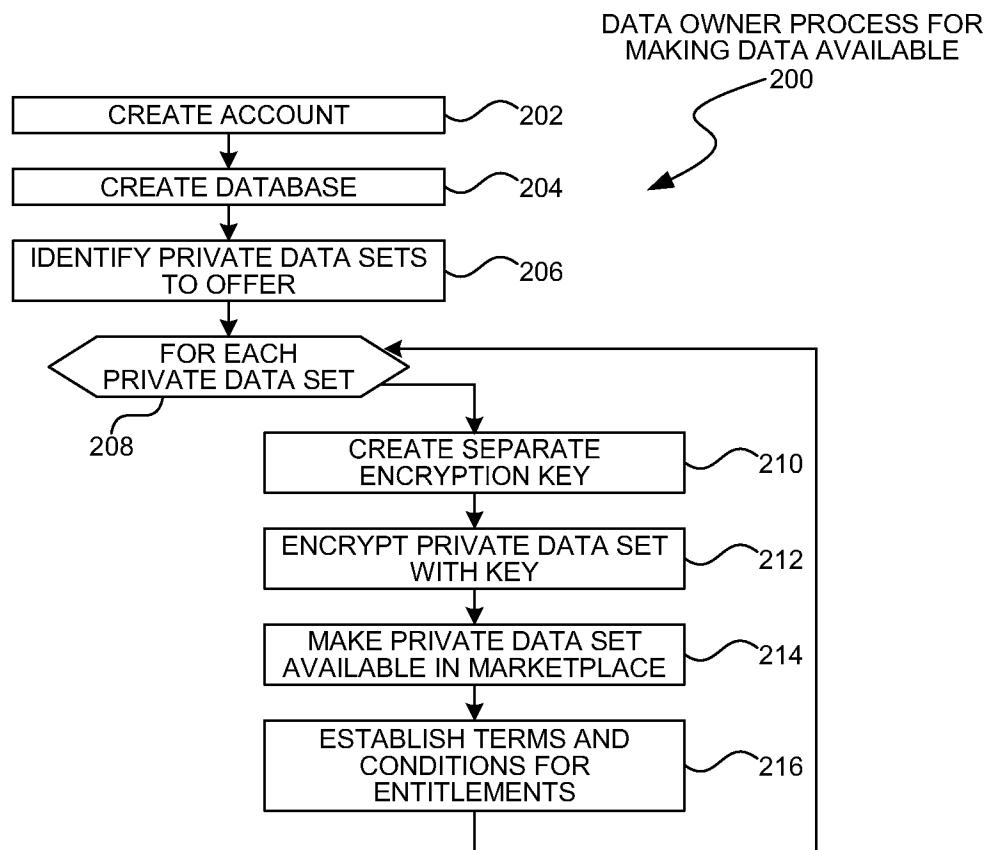
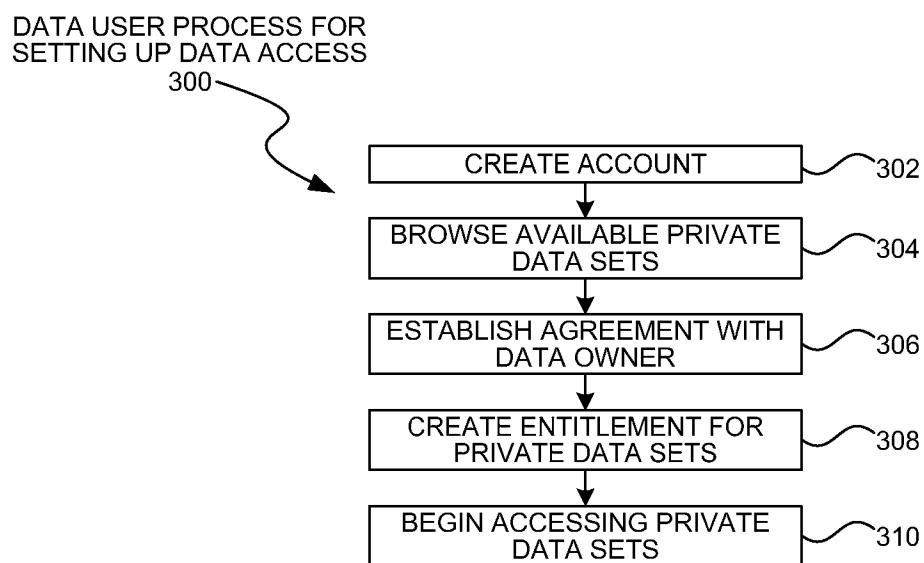


FIG. 1

**FIG. 2**

3/4

**FIG. 3**

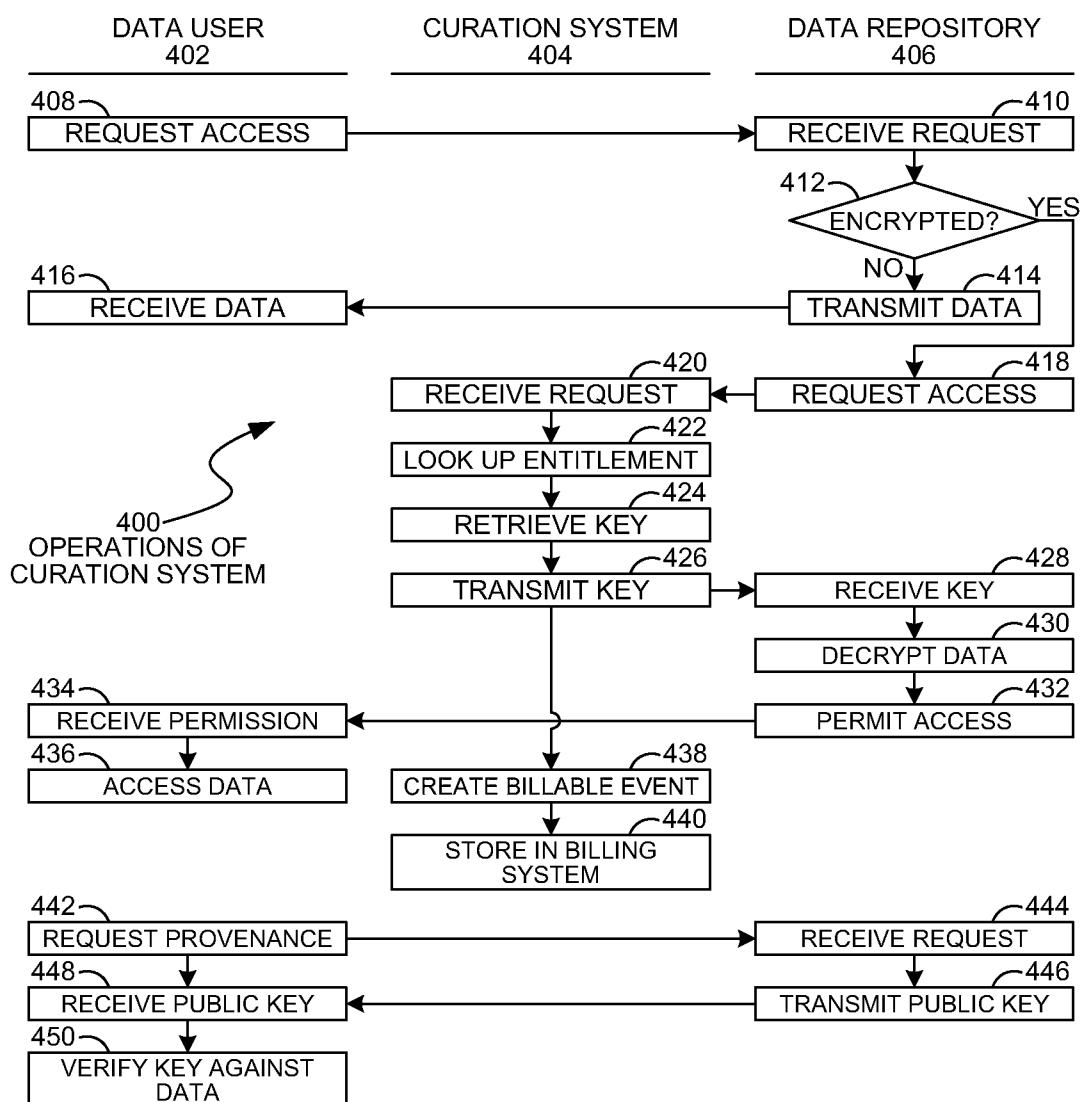


FIG. 4