



- (51) International Patent Classification:  
*G06Q 10/00* (2012.01)
- (21) International Application Number:  
PCT/US2016/056458
- (22) International Filing Date:  
11 October 2016 (11.10.2016)
- (25) Filing Language:  
English
- (26) Publication Language:  
English
- (30) Priority Data:  
62/240,497 12 October 2015 (12.10.2015) US  
14/884,703 15 October 2015 (15.10.2015) US  
14/939,986 12 November 2015 (12.11.2015) US
- (71) Applicant: **POKITDOK, INC.** [US/US]; 28 E. 3rd Avenue, Suite 300, San Mateo, CA 94401 (US).
- (72) Inventors: **DUNLEVY, Tim**; 28 E. 3rd Avenue, Suite 300, San Mateo, CA 94401 (US). **TANNER, Theodore, C.**; 28 E. 3rd Avenue, Suite 300, San Mateo, CA 94401 (US). **GOSNELL, Denise**; 28 E. 3rd Avenue, Suite 300, San Mateo, CA 94401 (US). **WHITEMIRE, Thomas, Dixon**; 28 E. 3rd Avenue, Suite 300, San Mateo, CA 94401 (US).

- (74) Agent: **LOHSE, Timothy, W.**; DLA Piper LLP (US), 2000 University Avenue, East Palo Alto, CA 94303 (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, DJ, 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, KW, 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).

[Continued on next page]

(54) Title: SYSTEMS AND METHOD FOR DYNAMIC AUTONOMOUS TRANSACTIONAL IDENTITY MANAGEMENT

(57) Abstract: A dynamic and autonomous identity management system and method are disclosed that provide a means of consolidating patient identity data from disparate data sources into a single system, making patient data easily accessible in a uniform and transparent manner.

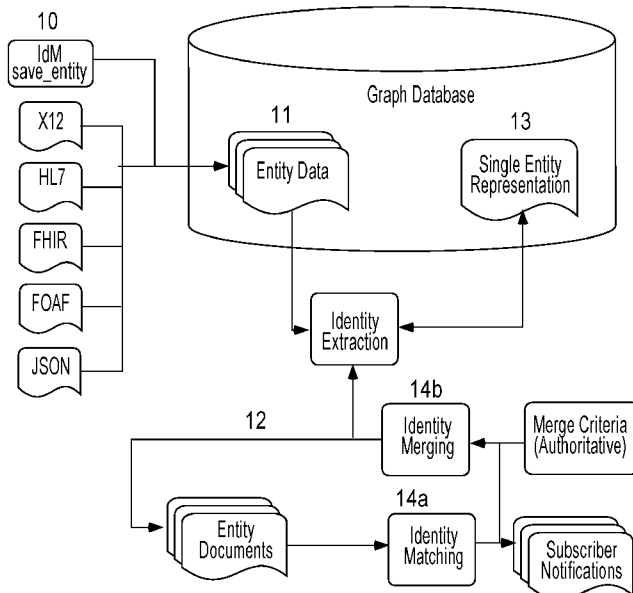


FIGURE 7

WO 2017/066202 A1

**Published:**

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

SYSTEM AND METHOD FOR DYNAMIC AUTONOMOUS TRANSACTIONAL IDENTITY  
MANAGEMENT

5 Field

The disclosure relates to health care identity management.

Background

The United States' meaningful use policy is intended to improve the efficiency of a health care provider's practice and improve patient outcomes through the use of an Electronic Health  
10 Record (EHR) system. Ideally, an EHR system records patient data, tracks clinical processes, and provides a means of sharing data with other providers involved in a patient's care. EHR systems are certified to ensure that they meet the guidelines set forth in the American Recovery and Reinvestment Act of 2009, aka The Recovery Act. While The Recovery Act defines minimum  
15 guidelines for meaningful use acceptance, it does not provide standards on electronic data interchange (EDI) between systems. The current standard widely used for EHR EDI transmissions is HL7, which is governed by Health Level Seven International.

While EHR systems generally support HL7 as a means of generating EDI transactions, implementations vary greatly between EHR systems. As a result sharing data between systems and consolidating records throughout multiple systems is a complex effort, as EHR systems  
20 support differing and sometimes customized versions of the HL7 standard.

Brief Description of the Drawings

Figure 1 illustrates an example of an embodiment of a healthcare identity management system;

Figures 2-3 illustrate an example of a data schema implemented for this identity  
25 management procedure;

Figure 4 illustrates an example of a data structure of the identity management system;

Figure 5 illustrates an example of an internal identity matching pipeline utilized in this system;

Figure 6 illustrates an example of an identity merging architecture of the system;

Figure 7 illustrates an example of another embodiment of a healthcare identity management system; and

Figure 8 illustrates an example of an implementation of the identity management system.

## 5 Detailed Description of One or More Embodiments

The disclosure is particularly applicable to an identity management system for healthcare as described below and it is in this context that the disclosure will be described. It will be appreciated, however, that the system and method has greater utility, such as to managing identities for other system that those below or be implemented in other manners than those  
10 described below. In the embodiment described below, the identity management system may be a PokitDok Identity Management system. As another example, the identity management system may be a standalone component/system or may be part of a healthcare system such as the PokitDok healthcare system in one embodiment.

The dynamic and autonomous identity management system described below provides a  
15 means of consolidating patient identity data from disparate data sources into a single system, making patient data easily accessible in a uniform and transparent manner.

Figure 1 illustrates an architecture and solution to identity management. Herein, the system transforms a streaming transactional data architecture into an identity management solution. Element 1.10 illustrates the save\_entity transaction as an aggregate transaction  
20 supporting multiple data interchange standards including, but not limited to: ASCII ANSI X12, HL7, FHIR, FOAF, and JSON. These formats are streamed as inputs and persisted separately within the graph database specified as element 1.11.

Algorithm 1 illustrates the save\_entity implementation used to persist entity data within the graph.

### 25 Algorithm 1

```
def save_entity(self, data_sets):  
    for data_set in data_sets:  
        data_adapter = self.load_adapter(data_set)  
        data_adapter.submit(data_set)
```

The implementation accepts one or more data sets as input, as specified by the `data_sets` parameter. The system matches a data adapter to each data set received and then persists the data set within the system. Data adapters encapsulate the operations required to work with each unique data set format such as JSON, ASCII ANSI X12, HL7, FHIR, FOAF, and JSON. Finally  
5 the data adapter submits each data set to the system in a non-blocking fashion as seen in Figure 1, element 1.1.

This solution extracts, persists, resolves and updates a unified identity for the entity observed throughout our stream of transactions. Element 1.13 may persist a conceptual property model of each entity in the graph with appropriate data provenance identifiers. The processes at  
10 elements 1.14, 1.15, 1.16, 1.17, and 1.18 create the dynamic autonomous pipeline of this identity management system. The data at 1.13 is lassoed and duplicates are identified via an autonomous identity matcher at 1.14. The matching criteria at 1.14 can be both learned thresholds (dynamic) and subscriber specific (static defined rules) regulations. This built in flexibility is necessary for protecting patient identity, the most critical piece of this system. With a set of identified  
15 duplicates, the data is checked for known subscriber regulations at 1.16, such as those outlined in the Health Insurance Portability and Accountability Act Privacy Rule or stricter payor specific regulations. At element 1.15, the system merges the identities and create identity profiles which adhere to the entity's regulations which are controlled at element 1.17. As transitions are continuously streamed into the architecture, the merged identities are stored in the graph database  
20 at 1.18 and are updated dynamically according to the optimization functions at 1.14.a and 1.14.b.

### Service Views

FIGURE 4 demonstrates the interaction of the Identity Management Solution with an external health care enterprise system such as an Enterprise Master Person Index (EMPI), Electronic Medical Record (EMR), Practice Management (PM), or Electronic Health Record  
25 (EHR) system. For a patient in the Identity Management database (element 1.13 in Figure 1), the system checks to see if the system has created the mapping for this identity to the respective external system. If the patient is not mapped in the system, the system queries the external interface with the patient's strong identifiers. If this patient does not exist, the system creates the patient in the external system, establish the patient mappings and use these transactions to update  
30 our internal transactional stream as described in FIGURE 1.

The identity record data structure may be:

```

{
  "first_name": "John",
  "middle_name": "P",
5   "last_name": "Doe"
  "birth_date": "1980-02-17",
  "gender": "M",
  "external_ids": {"external_system_a": "WQ123456789", "external_system_b": "P-
1432"}
10  }

```

The identity record data structure contains demographic data and is used as an input to query\_patient and save\_patient\_mapping functions throughout the system. The external\_ids field is a map structure, used to maintain an external system identity ids. A map entry is indexed on the external system id and resolves to the external system id value.

15 Algorithm 2 query\_patient() implementation:

```

query_patient(self, pd_identity_record, external_system_id):
  pd_identity_record = self.identity_service.find(pd_identity_record)

  if not pd_identity_record:
20   pd_identity_record = self.identity_service.create(pd_identity_record)

  system_adapter = self.load_system_adapter(external_system_id)
  external_patient = system_adapter.find(pd_identity_record)

25   if not external_patient:
     external_patient = system_adapter.create(pd_identity_record)

  return external_patient

```

30 The query\_patient function has two arguments the pd\_identity\_record (Figure 4) and the external\_system\_id. The external\_system\_id is a system defined unique identifier used to identify

each external systems that are supported within the Identity Management Solution. The external\_system\_id is used to load the appropriate system\_adapter for the external system. Each system adapter is used to encapsulate search operations for the external system.

Algorithm 3 save\_patient\_mapping() implementation:

```

5 save_patient_mapping(self, pd_identity_record, external_patient):
    if external_patient.system_id not in pd_identity_record.external_ids:
        pd_identity_record.external_ids[external_patient.system_id] = external_patient.id

```

The save\_patient\_mapping function updates an Identity Management Solution identity\_record (Figure 4) with the external system to external system patient id mapping.

10 Database Schema

The identity extraction process (element 1.12) relates multiple entity documents (Figure 1, element 1.11) to a consolidated document (Figure 1, element 1.18) using the matching (Figure 1, element 1.14) and optional merging (Figure 1, element 1.15) processes. Together, FIGURE 2 and FIGURE 3 provide a specific example of the schema implemented for this procedure.

15 Specifically, FIGURE 2 shows the immutable transactional property graph model for an eligibility request for an office visit. FIGURE 3 depicts the three stages of the identity persistence for a consumer. FIGURE 3.a corresponds to the root of this transaction from FIGURE 2; this is an immutable data structure in the graph database. Next, at FIGURE 3.b, the system extracts a bag of words consumer model from every transaction, which streams through the PokitDok EDI  
20 architecture. The bag-of-words model at FIGURE 3.b extracts and persists human readable key value pairs from the EDI transaction as one vertex in the graph architecture. Lastly, the system creates a merged property graph model at FIGURE 3.c that serves as the main identity for the consumer of interest. In its entirety, FIGURE 3 shows the persistence of the entity extraction from the transaction through the resulting merged property graph model for the duplicate  
25 consumers in the example. Persistence stages 3.a, 3.b, and 3.c in FIGURE 3 correlate to the architecture stages 1.11, 1.13, and 1.18, respectively, from FIGURE 1.

FIGURE 5 outlines the internal identity matching pipeline utilized in this system. The pipeline dynamically queries the models available at 1.13 and 1.18 in FIGURE 1 and identifies sets of duplicate entities. Each set of duplicate entities receives a unique match key to be  
30 exploited in the identity merging pipeline. FIGURE 6 details the identity merging architecture.

For each unique match key observed at 1.13 and 1.18 in FIGURE 1, the identity merging pipeline either creates or updates the merged representation for that identity, as shown in Figure 5.

FIGURE 6. To provide external control to regulate identities in the system, the identity merging pipeline creates a profiled representation for each entity which requires additional management. The regulations can create profiled representations within the identity management system via learned autonomous threshold or manual data extractions.

#### Formal definition of Entity Evolution

There are entire fields of theoretical research on *matching*, *entity resolution*, *fuzzy grouping*, and *object consolidation*. All of these approaches are referring to the same underlying generic problem: a dataset  $D$  contains  $m$  entries:

$D = \{d_1, d_2, \dots, d_m\}$ . The entries can be different types with varying relationships to other entities. The objective of a solution in entity resolution is to define a set  $R = \{r_1, r_2, \dots, r_n\}$  such that  $|R| = n \leq m = |D|$  and every element in  $D$  correctly maps to an element in  $R$ . That is,  $R$  is the set of instantiated attributes of the objects in  $D$ . Extensive reviews of previous methods and solutions in this space can be found in, but not limited to, the following: 1) Chen, Zhaoqi, Dmitri V. Kalashnikov, and Sharad Mehrotra. "Adaptive graphical approach to entity resolution." *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*. ACM, 2007; 2) Christen, Peter. *Data matching: concepts and techniques for record linkage, entity resolution, and duplicate detection*. Springer Science & Business Media, 2012; and 3) Cohen, William, Pradeep Ravikumar, and Stephen Fienberg. "A comparison of string metrics for matching names and records." *Kdd workshop on data cleaning and object consolidation*. Vol. 3. 2003. The system combines domain specific features with proven techniques in similarity scoring and novel approaches in graph isomorphisms to adapt feature thresholds  $\delta_i$  to dynamically create  $R$  from a continuous stream of healthcare transactions.

Recall the architecture outlined at 1.13 and 1.18 from FIGURE 1. Given the definition above, FIGURE 1.13 will now be referred to as set  $R$  whereas FIGURE 1.18 corresponds to set  $D$ . The system utilizes both the features and relationships from our architecture to learn, train, and update the matching algorithm at FIGURE 1 element 1.14. The system may combine similarity techniques from the bag of words model demonstrated at FIGURE 3.b, inferred

relationships at 3.c, feature similarity scoring, and domain knowledge to create the backbone of the identity matching algorithm. The system may also apply optimization techniques, such as linear programming, to dynamically create an evolved representation of an entity's identity by minimizing the difference in similarity across the identity's evolution.

5        Algorithm 4 calculate similarity\_score() implementation:  
       calculate\_similarity\_score(self, pd\_identity\_record\_a, pd\_identity\_record\_b):  
           similarity\_score = new ScoreIdentities()  
           similarity\_score.score\_domain\_features(pd\_identity\_record\_a, pd\_identity\_record\_b)  
                   similarity\_score.score\_semantic\_features(pd\_identity\_record\_a,  
10                pd\_identity\_record\_b)  
                   similarity\_score.score\_relationship\_inferences(pd\_identity\_record\_a,  
                   pd\_identity\_record\_b)  
           similarity\_score.score\_features(pd\_identity\_record\_a, pd\_identity\_record\_b)  
           similarity\_score.calculate\_priority\_score()  
15        return similarity\_score.total\_score

For example, Equation 1 below details the calculation of a graph connectivity score by calculating the total connection strength of shortest paths between two identity feature graphs:

Equation 1:

$$\text{graphConnectivityScore}(a,b) = \sum_{p \in \mathcal{P}_L(a,b)} w(p)$$

20        where the set  $L$  denotes the set of short paths between entity relationship graphs for identity  $a$  and identity  $b$ ,  $p \in L$  is a path in  $L$ , and  $w(p)$  denotes the total weight (or strength) of this path. To account for outliers in this distribution, the system normalizes the overall relationship inference score for two entities. In addition to applying optimization techniques to an identity's dynamically evolving *graphConnectivityScore*, the system utilizes a myriad of  
25        similarity algorithms such as, but not limited to, the following: cosine similarity, jaccard similarity, hamming distance, simple matching coefficients, graph isomorphisms, overlap coefficient, maximal matchings, Tversky index, Levenshtein distance, object frequency,

Hellinger distance, skew divergence, confusion probability, Kullback-Leibler divergence metric, ... etc.

Merging will be handled from highest to lowest priority merges, given that the similarity score for each merge exceeds the matching threshold  $\delta_i$ . After computing a combined similarity score with semantic similarity, relationship inference, feature similarity, and/or domain knowledge, the system merges a set of identities with the highest score, provided the overall score exceeds our matching threshold. In the case that the merge percolates a scoring change to another set of nodes, the scores will be updated and re-evaluated for merging.

The similarity scores for any two entities represent a co-reference distribution in which the system can positionally order the similarity of any two entities in relation to the global distribution. This ordering drives the supervised learning algorithm for identity management evolution. The system aims to merge identities using the similarity ordering such that the system minimizes the similarity difference function for the new mapping  $\phi(n): D \rightarrow R$  where:

$$\forall r \in R: \phi_r(n) - \phi_r(n-1) \geq \epsilon, \epsilon \geq 0$$

and

$$\phi_r(n) = \text{simScore}(r, \text{merge}(r, c))$$

where  $a, b \in r, c \in r$  and

$$\text{simScore}(a, b) > \delta_i \text{ and}$$

$$\text{simScore}(a, c) > \delta_j$$

That is to say, as the system evolves the mapping and state space observed in the entity evolution set  $R$ , the system seeks to augment the set such that we minimize the similarity difference from one step to another. The application of this objective function is shown in FIGURE 1 at 1.14a and 1.14b in Figure 7 in which those processes are coupled differently to the graph database and identity extraction as shown.

For example, consider the following two identities extracted from EDI transactions:

a = {  
 "first\_name": "John",  
 "middle\_name": "P",  
 "last\_name": "Doe"

```

    "birth_date": "1980-02-17",
    "gender": "M",
    "external_ids": {"external_system_a": "WQ123456789", "external_system_b": "P-
1432"}

```

5

}

b = {

```

    "first_name": "John",

```

```

    "last_name": "Doe"

```

```

    "birth_date": "1980-02-17",

```

10

```

    "gender": "M",

```

```

    "address": "1 Main Street Los Angeles, CA 55555",

```

```

    "external_ids": {"external_system_a": "WQ123456789", "external_system_b": "P-
1432"}

```

}

15

These two identities would surpass the domain knowledge similarity scoring function due to matching strong identifiers from an external system. If the domain similarity scoring function has a normalized range from  $[0, 1]$ , then for these two documents:

$$\text{domainSimScore}(a,b) = 1 > \delta_{\text{domain}}$$

The system would initialize  $D = \{a, b\}$  and

20  $R = \{a'\}$ ,  $\phi(n): D \rightarrow R = a, b \rightarrow a'$  and:

a' = {

```

    "first_name": "John",

```

```

    "middle_name": "P",

```

```

    "last_name": "Doe"

```

25

```

    "birth_date": "1980-02-17",

```

```

    "gender": "M",

```

```

    "address": "1 Main Street Los Angeles, CA 55555",

```

```

    "external_ids": {"external_system_a": "WQ123456789", "external_system_b": "P-
1432"}

```

30

}

Next, consider the following two identities which we would observe via separate EDI transactions:

```

c = {
  "first_name": "Jane",
  "middle_name": "P",
  "last_name": "Doe"
  "birth_date": "1980-02-17",
  "gender": "F",
  "address": "1 Main Street Los Angeles, CA 55555",
  "external_ids": {"external_system_a": "AB123456789", "external_system_b": "P-
ABCD"}
}

```

```

d = {
  "first_name": "J",
  "middle_name": "P",
  "last_name": "Doe"
  "birth_date": "1980-02-17",
  "gender": "M",
  "address": "1 Main St. LA, CA 55555",
  "external_ids": {"external_system_a": "WQ123456789", "external_system_b": "P-
1432"}
}

```

The system would initialize  $D = \{a, b, c, d\}$ . The system seeks to define a mapping  $\phi(n+1): D \rightarrow R$  such that every element in  $D$  maps to an element in  $R$ . First, let us consider

$$\text{semanticSimScore}(a', c) > \delta_{\text{semantic}}$$

due to the semantic similarities between the fields of the two documents. However,

$$\text{domainSimScore}(a', c) < \delta_{\text{domain}}$$

because the external system identity keys do not match. As such, the system will not augment  $a'$  to include identity  $c$  and the system will make an isolate identity  $c' \in R$  and define

$$\phi(n+1): c \rightarrow R$$

Lastly, the system must consider identity  $d$ . The system calculates

$$\text{semanticSimScore}(a', d) > \delta_{\text{semantic}}$$

$$\text{domainSimScore}(a', d) > \delta_{\text{domain}}$$

The system ensures an objective function is satisfied by observing that:

$$5 \quad \phi_{a'}(n+1) = \text{simScore}(a', \text{merge}(a', d)) \text{ and}$$

$$\phi_{a'}(n+1) - \phi_{a'}(n) \geq \epsilon, \epsilon \geq 0$$

Figure 8 illustrates an example of an implementation of the identity management system 800 that has one or more computing devices 802, such as 802A, 802B, ..., 802N as shown in Figure 8) which allow a user (patient, healthcare provider or any other entity with sufficient access to the system) to couple to and interact over a communications path 804 with an identity management backend system 806 in the manners described above. Each computing device 802 may be a processor based device with a display, memory, storage and connectivity capabilities. For example, each computing device may be a smartphone device, such an Apple iPhone or Android operating system based device, a personal computer, a laptop computer, a tablet computer, a terminal and the like. Each computing device 802 may have an application to facilitate connection to and communication with the identity management backend 806 such as a browser application, mobile application or any other application. Each computing device allows the user to issue a request to the identity management backend 806 and receive a response back.

The communications path 804 may be a wired or wireless network for a combination therefore that permit each computing device to connect to and interact with the identity management backend 806. For example, the communications path 804 may be one or more of Ethernet, the Internet, a wireless data network, a cellular digital data network, a computer network and the like. The communications path 804 may use various communications and data transfer protocols for its operation, such as HTTP, REST, HTTPS, TCP/IP and the like.

The identity management backend 806 may be one or more specially designed computing resources including one or more processors, memory, storage, a communications circuit and the like. For example, the identity management backend 806 may be implemented using one or more server computers, a database server, an application server, a blade server and/or cloud computing components. Each component of the identity management backend 806 may be implemented in

hardware or software. In a software implementation, each of the components is a plurality of lines of computer code executed by a processor of a computer system or of the computing resources to implement the functions of the system as described above. In the hardware implementation, each of the components may be a hardware device such as a microcontroller, a programmable logic device, a field programmable gate array and the like.

The identity management backend 806 may further comprise a user interface component 806A that manages the connections and communications with each computing device. In a client server type implementation, the user interface component may be a web server. The identity management backend 806 may further comprise an identity management component 806B that performs the identity management operations and processes described above. The identity management component 806B may further comprise an identity processing component 806B1 and a graph database component 806B2 as described above. The identity processing component 806B1 provides an interface to the graph database and may perform the identity extraction 12, identity matching 14 and identity merger 15 shown in Figure 1.

The foregoing description, for purpose of explanation, has been described with reference to specific embodiments. However, the illustrative discussions above are not intended to be exhaustive or to limit the disclosure to the precise forms disclosed. Many modifications and variations are possible in view of the above teachings. The embodiments were chosen and described in order to best explain the principles of the disclosure and its practical applications, to thereby enable others skilled in the art to best utilize the disclosure and various embodiments with various modifications as are suited to the particular use contemplated.

The system and method disclosed herein may be implemented via one or more components, systems, servers, appliances, other subcomponents, or distributed between such elements. When implemented as a system, such systems may include an/or involve, inter alia, components such as software modules, general-purpose CPU, RAM, etc. found in general-purpose computers. In implementations where the innovations reside on a server, such a server may include or involve components such as CPU, RAM, etc., such as those found in general-purpose computers.

Additionally, the system and method herein may be achieved via implementations with disparate or entirely different software, hardware and/or firmware components, beyond that set

forth above. With regard to such other components (e.g., software, processing components, etc.) and/or computer-readable media associated with or embodying the present inventions, for example, aspects of the innovations herein may be implemented consistent with numerous general purpose or special purpose computing systems or configurations. Various exemplary  
5 computing systems, environments, and/or configurations that may be suitable for use with the innovations herein may include, but are not limited to: software or other components within or embodied on personal computers, servers or server computing devices such as routing/connectivity components, hand-held or laptop devices, multiprocessor systems, microprocessor-based systems, set top boxes, consumer electronic devices, network PCs, other  
10 existing computer platforms, distributed computing environments that include one or more of the above systems or devices, etc.

In some instances, aspects of the system and method may be achieved via or performed by logic and/or logic instructions including program modules, executed in association with such components or circuitry, for example. In general, program modules may include routines,  
15 programs, objects, components, data structures, etc. that perform particular tasks or implement particular instructions herein. The inventions may also be practiced in the context of distributed software, computer, or circuit settings where circuitry is connected via communication buses, circuitry or links. In distributed settings, control/instructions may occur from both local and remote computer storage media including memory storage devices.

The software, circuitry and components herein may also include and/or utilize one or more type of computer readable media. Computer readable media can be any available media that is resident on, associable with, or can be accessed by such circuits and/or computing components. By way of example, and not limitation, computer readable media may comprise computer storage media and communication media. Computer storage media includes volatile and nonvolatile,  
25 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 tape, magnetic disk storage or other magnetic storage devices, or any other  
30 medium which can be used to store the desired information and can accessed by computing

component. Communication media may comprise computer readable instructions, data structures, program modules and/or other components. Further, communication media may include wired media such as a wired network or direct-wired connection, however no media of any such type herein includes transitory media. Combinations of the any of the above are also  
5 included within the scope of computer readable media.

In the present description, the terms component, module, device, etc. may refer to any type of logical or functional software elements, circuits, blocks and/or processes that may be implemented in a variety of ways. For example, the functions of various circuits and/or blocks can be combined with one another into any other number of modules. Each module may even be  
10 implemented as a software program stored on a tangible memory (e.g., random access memory, read only memory, CD-ROM memory, hard disk drive, etc.) to be read by a central processing unit to implement the functions of the innovations herein. Or, the modules can comprise programming instructions transmitted to a general purpose computer or to processing/graphics hardware via a transmission carrier wave. Also, the modules can be implemented as hardware  
15 logic circuitry implementing the functions encompassed by the innovations herein. Finally, the modules can be implemented using special purpose instructions (SIMD instructions), field programmable logic arrays or any mix thereof which provides the desired level performance and cost.

As disclosed herein, features consistent with the disclosure may be implemented via  
20 computer-hardware, software and/or firmware. For example, the systems and methods disclosed herein may be embodied in various forms including, for example, a data processor, such as a computer that also includes a database, digital electronic circuitry, firmware, software, or in combinations of them. Further, while some of the disclosed implementations describe specific hardware components, systems and methods consistent with the innovations herein may be  
25 implemented with any combination of hardware, software and/or firmware. Moreover, the above-noted features and other aspects and principles of the innovations herein may be implemented in various environments. Such environments and related applications may be specially constructed for performing the various routines, processes and/or operations according to the invention or they may include a general-purpose computer or computing platform selectively activated or  
30 reconfigured by code to provide the necessary functionality. The processes disclosed herein are

not inherently related to any particular computer, network, architecture, environment, or other apparatus, and may be implemented by a suitable combination of hardware, software, and/or firmware. For example, various general-purpose machines may be used with programs written in accordance with teachings of the invention, or it may be more convenient to construct a  
5 specialized apparatus or system to perform the required methods and techniques.

Aspects of the method and system described herein, such as the logic, may also be implemented as functionality programmed into any of a variety of circuitry, including programmable logic devices ("PLDs"), such as field programmable gate arrays ("FPGAs"), programmable array logic ("PAL") devices, electrically programmable logic and memory devices  
10 and standard cell-based devices, as well as application specific integrated circuits. Some other possibilities for implementing aspects include: memory devices, microcontrollers with memory (such as EEPROM), embedded microprocessors, firmware, software, etc. Furthermore, aspects may be embodied in microprocessors having software-based circuit emulation, discrete logic (sequential and combinatorial), custom devices, fuzzy (neural) logic, quantum devices, and  
15 hybrids of any of the above device types. The underlying device technologies may be provided in a variety of component types, e.g., metal-oxide semiconductor field-effect transistor ("MOSFET") technologies like complementary metal-oxide semiconductor ("CMOS"), bipolar technologies like emitter-coupled logic ("ECL"), polymer technologies (e.g., silicon-conjugated polymer and metal-conjugated polymer-metal structures), mixed analog and digital, and so on.

It should also be noted that the various logic and/or functions disclosed herein may be enabled using any number of combinations of hardware, firmware, and/or as data and/or instructions embodied in various machine-readable or computer-readable media, in terms of their behavioral, register transfer, logic component, and/or other characteristics. Computer-readable media in which such formatted data and/or instructions may be embodied include, but are not  
25 limited to, non-volatile storage media in various forms (e.g., optical, magnetic or semiconductor storage media) though again does not include transitory media. Unless the context clearly requires otherwise, throughout the description, the words "comprise," "comprising," and the like are to be construed in an inclusive sense as opposed to an exclusive or exhaustive sense; that is to say, in a sense of "including, but not limited to." Words using the singular or plural number also  
30 include the plural or singular number respectively. Additionally, the words "herein," "hereunder,"

"above," "below," and words of similar import refer to this application as a whole and not to any particular portions of this application. When the word "or" is used in reference to a list of two or more items, that word covers all of the following interpretations of the word: any of the items in the list, all of the items in the list and any combination of the items in the list.

5           Although certain presently preferred implementations of the invention have been specifically described herein, it will be apparent to those skilled in the art to which the invention pertains that variations and modifications of the various implementations shown and described herein may be made without departing from the spirit and scope of the invention. Accordingly, it is intended that the invention be limited only to the extent required by the applicable rules of law.

10           While the foregoing has been with reference to a particular embodiment of the disclosure, it will be appreciated by those skilled in the art that changes in this embodiment may be made without departing from the principles and spirit of the disclosure, the scope of which is defined by the appended claims.

Claims:

1. An identity management system, comprising:  
a computer system  
a data store associated with the computer system, the data store containing a plurality of patient records, wherein each patient record contains information about a particular patient;  
5 an identity management component hosted on the computer system;  
the identity management component configured to perform an identity extraction on a piece of input data to determine if the identity of the patient is contained in the data store, an identity matching on a plurality of pieces of input data about the patient to determine if the plurality of pieces of data describe the patient contained in the data store and an identity merging  
10 for autonomous evolution of the pieces of data and merging, in the graph database, the plurality of pieces of data that describe the patient contained in the data store.
2. The system of claim 1, wherein the identity extraction extracts a bag of words consumer model from every transaction for each patient.
3. The system of claim 1, wherein the identity matching generates a unique match  
15 key for each duplicate data for the patient.
4. The system of claim 1, wherein the identity merging generate a similarity score for each two pieces of data about the patient.
5. The system of claim 4, wherein the identity merging merges the pieces of data for the patient when the similar score exceeds a matching threshold.
- 20 6. The system of claim 4, wherein the identity merging first merges the pieces of data for the patient with a highest score.
7. The system of claim 1, wherein the data store further comprises a graph database.
8. An identity management method, comprising:  
providing a computer system and a data store associated with the computer system, the  
25 data store containing a plurality of patient records, wherein each patient record contains information about a particular patient;  
performing, based on the data store, an identity extraction on a piece of input data to determine if the identity of the patient is contained in the data store;

determining, if the plurality of pieces of data describe the patient contained in the data store;

autonomous evolving of the pieces of data; and

merging, in the data store, the plurality of pieces of data that describe the patient  
5 contained in the data store.

9. The method of claim 8, wherein performing the identity extraction further comprises extracting a bag of words consumer model from every transaction for each patient.

10. The method of claim 8, wherein performing the identity matching further comprises generating a unique match key for each duplicate data for the patient.

10 11. The method of claim 8, wherein identity merging further comprises generating a similarity score for each two pieces of data about the patient.

12. The method of claim 11 wherein identity further comprising merging the pieces of data for the patient when the similar score exceeds a matching threshold.

15 13. The method of claim 11, wherein the identity merging further comprises first merging the pieces of data for the patient with a highest score.

14. The method of claim 8, wherein the data store further comprises a graph database.

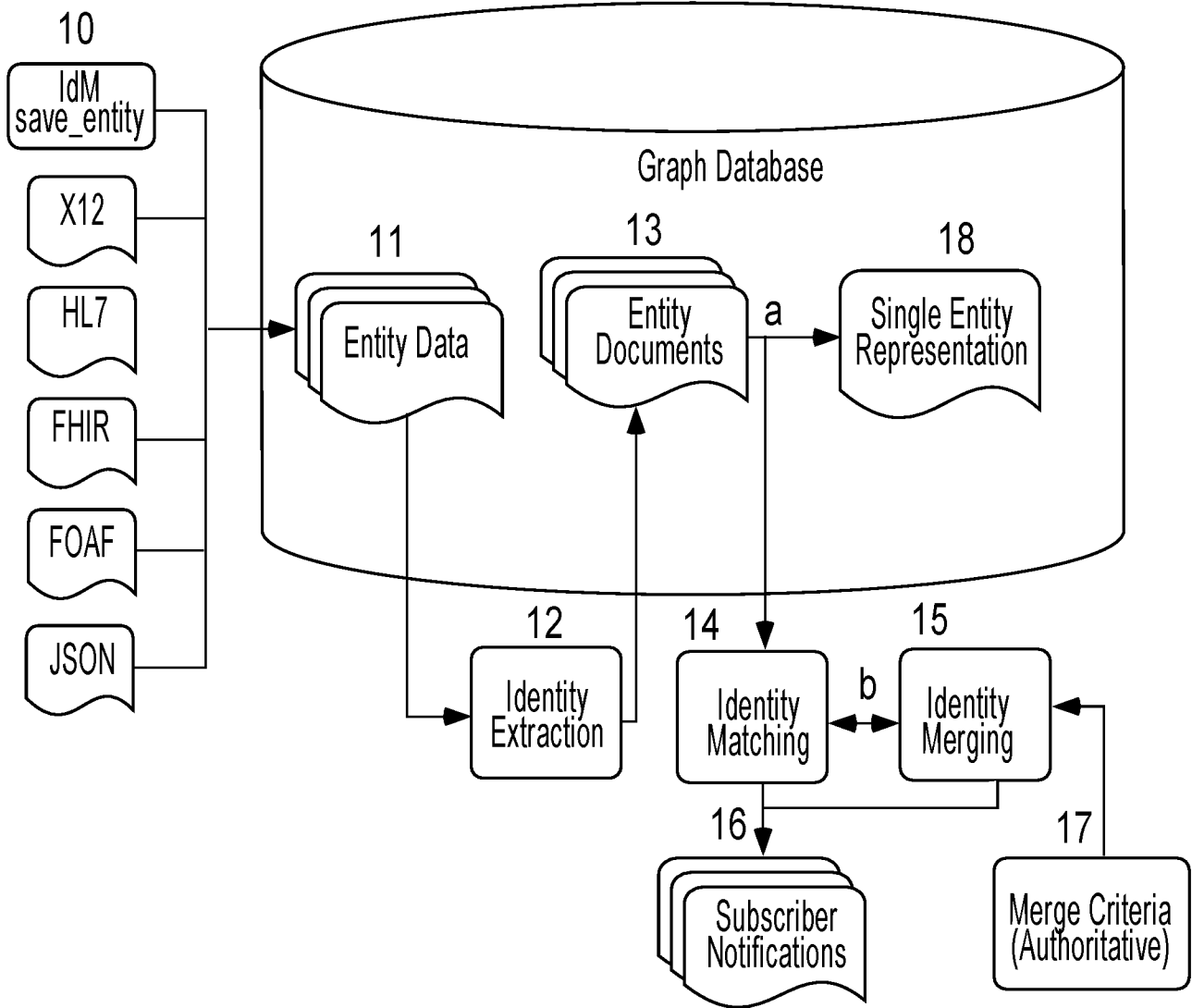


FIGURE 1



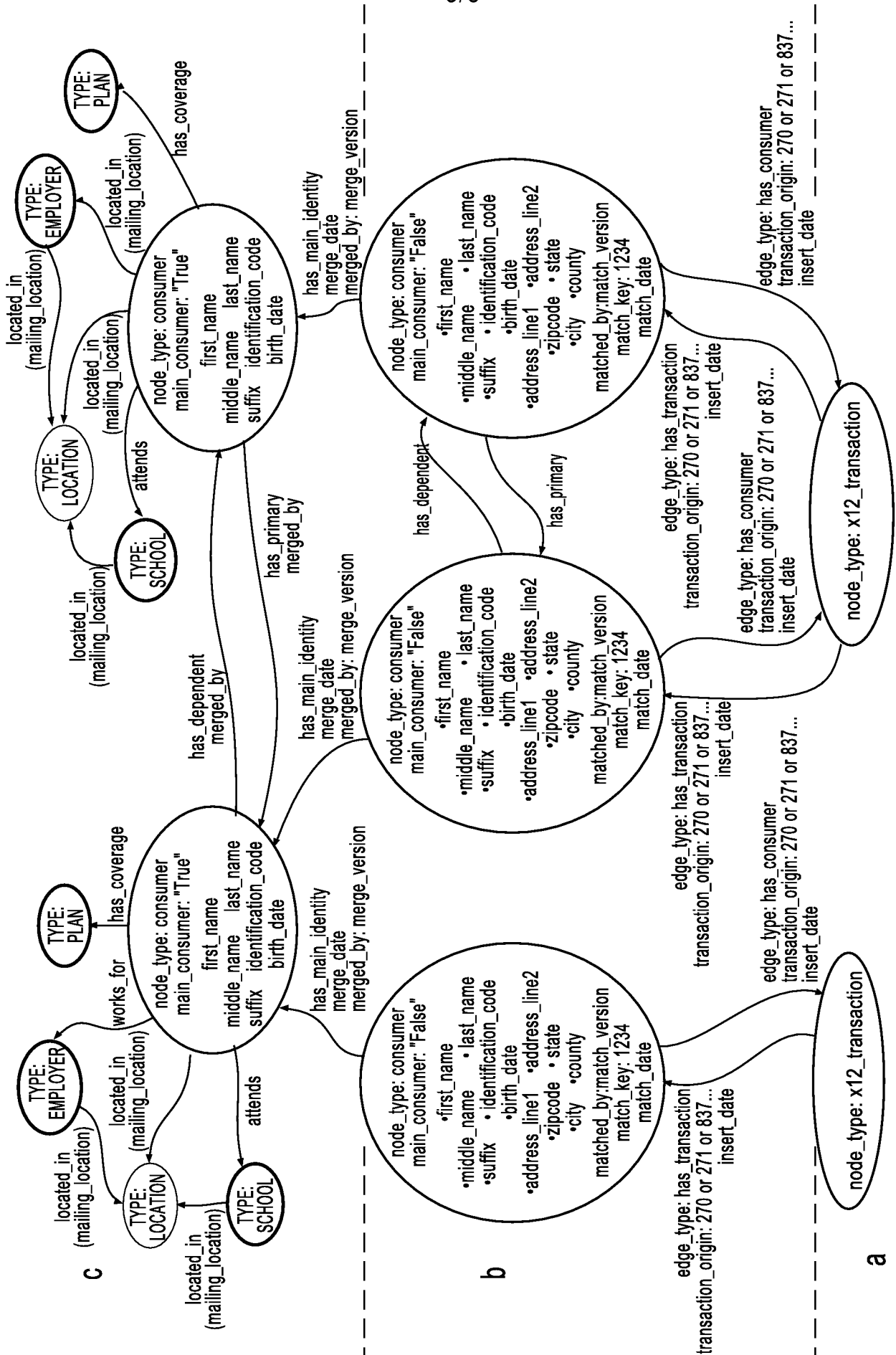


FIGURE 3

4/8

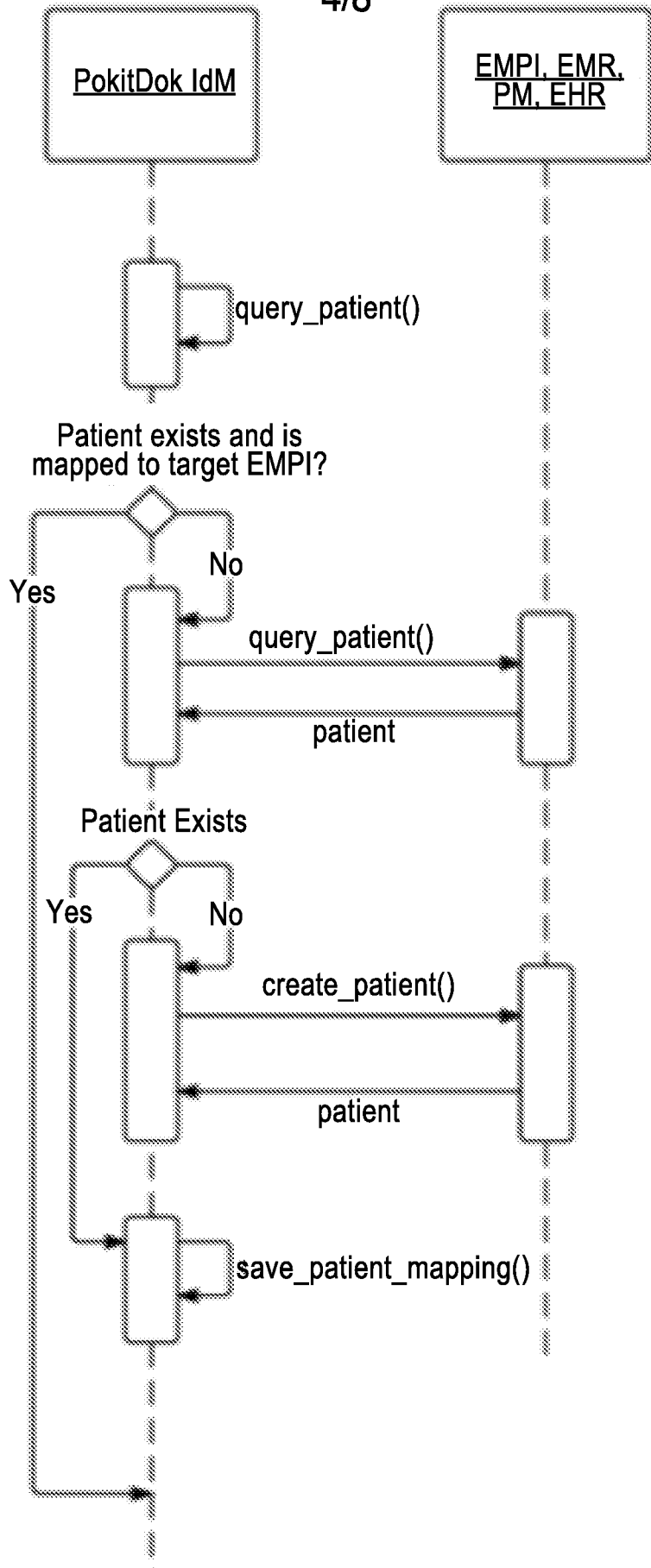


FIGURE 4

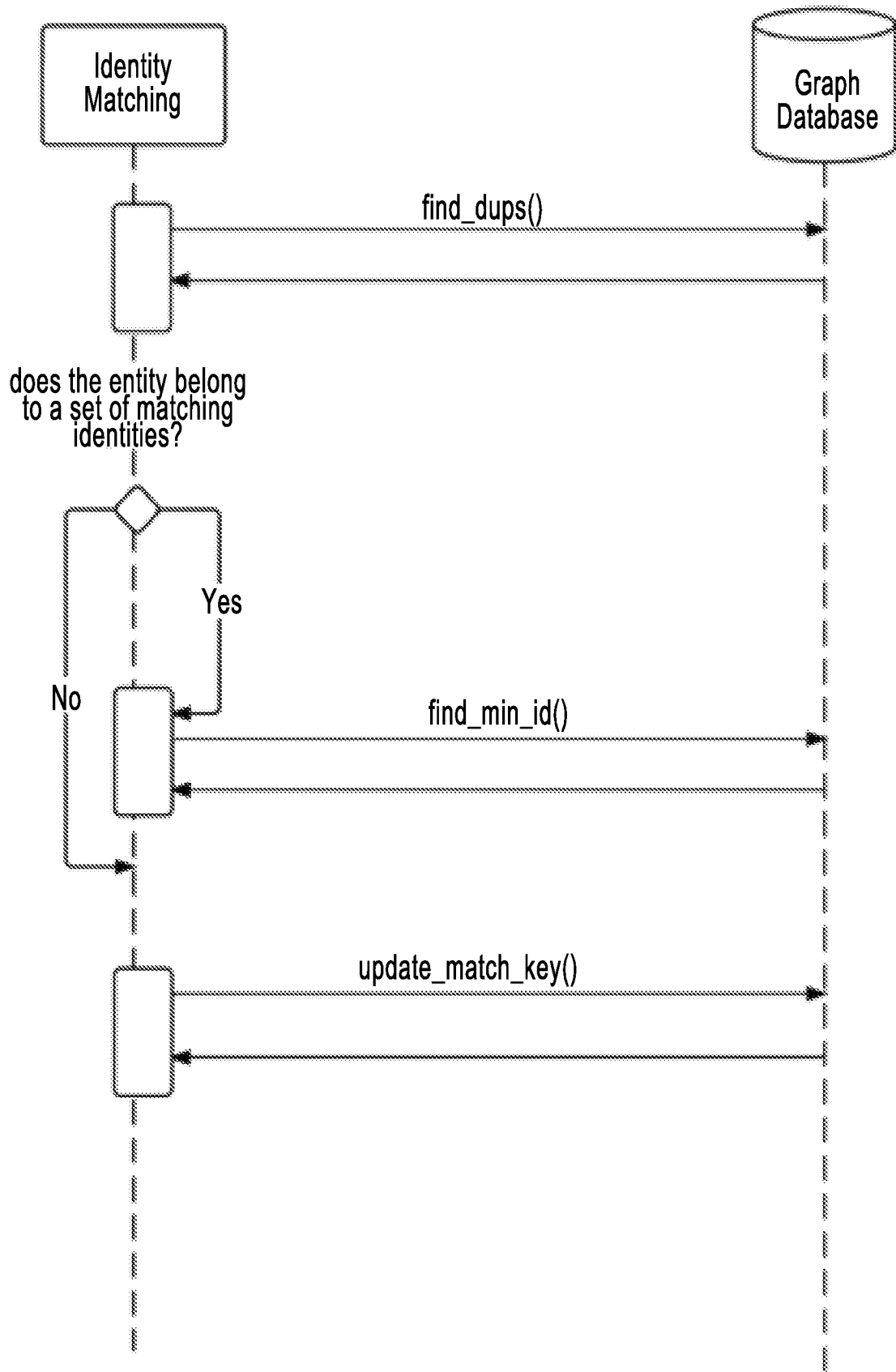


FIGURE 5

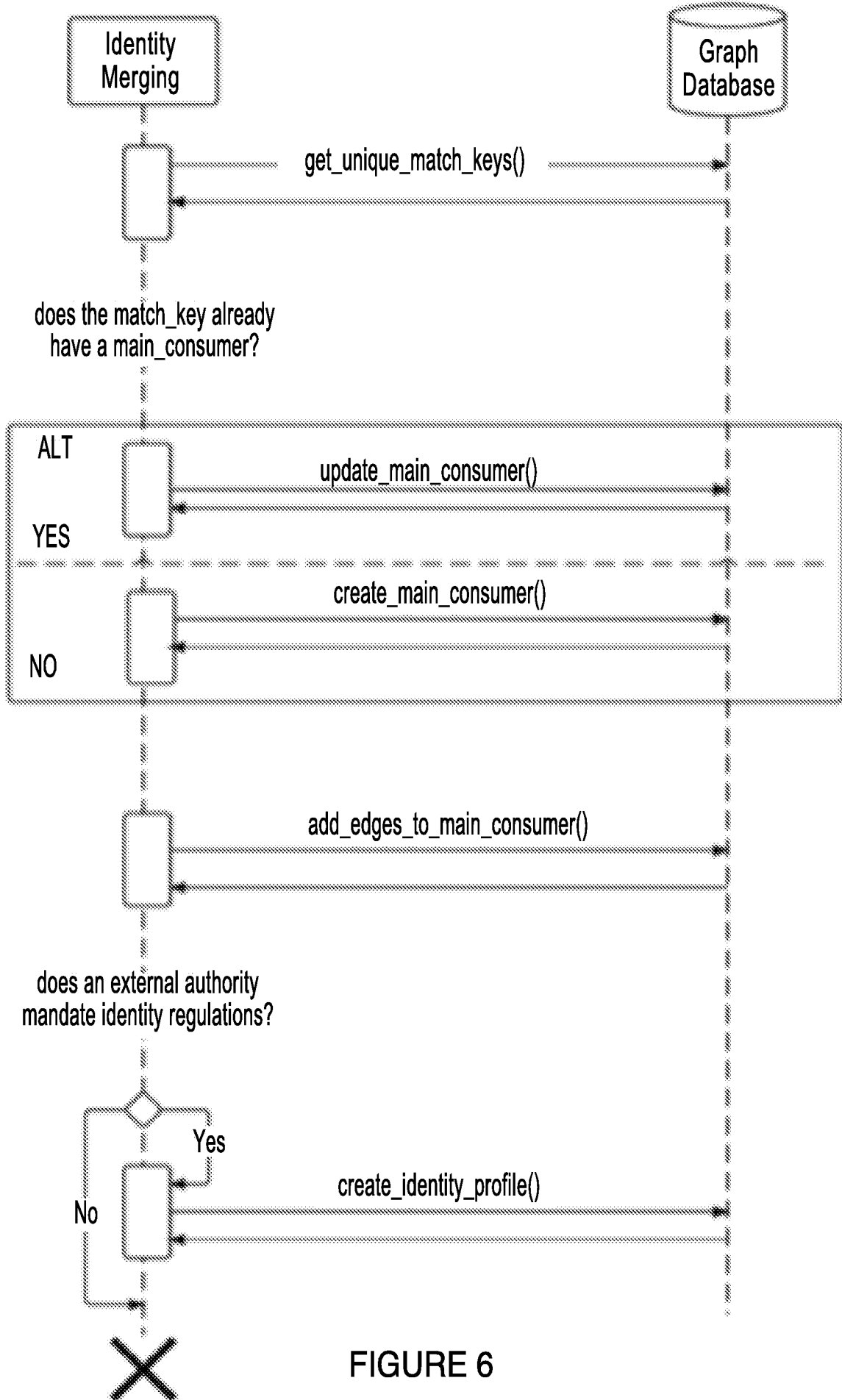


FIGURE 6

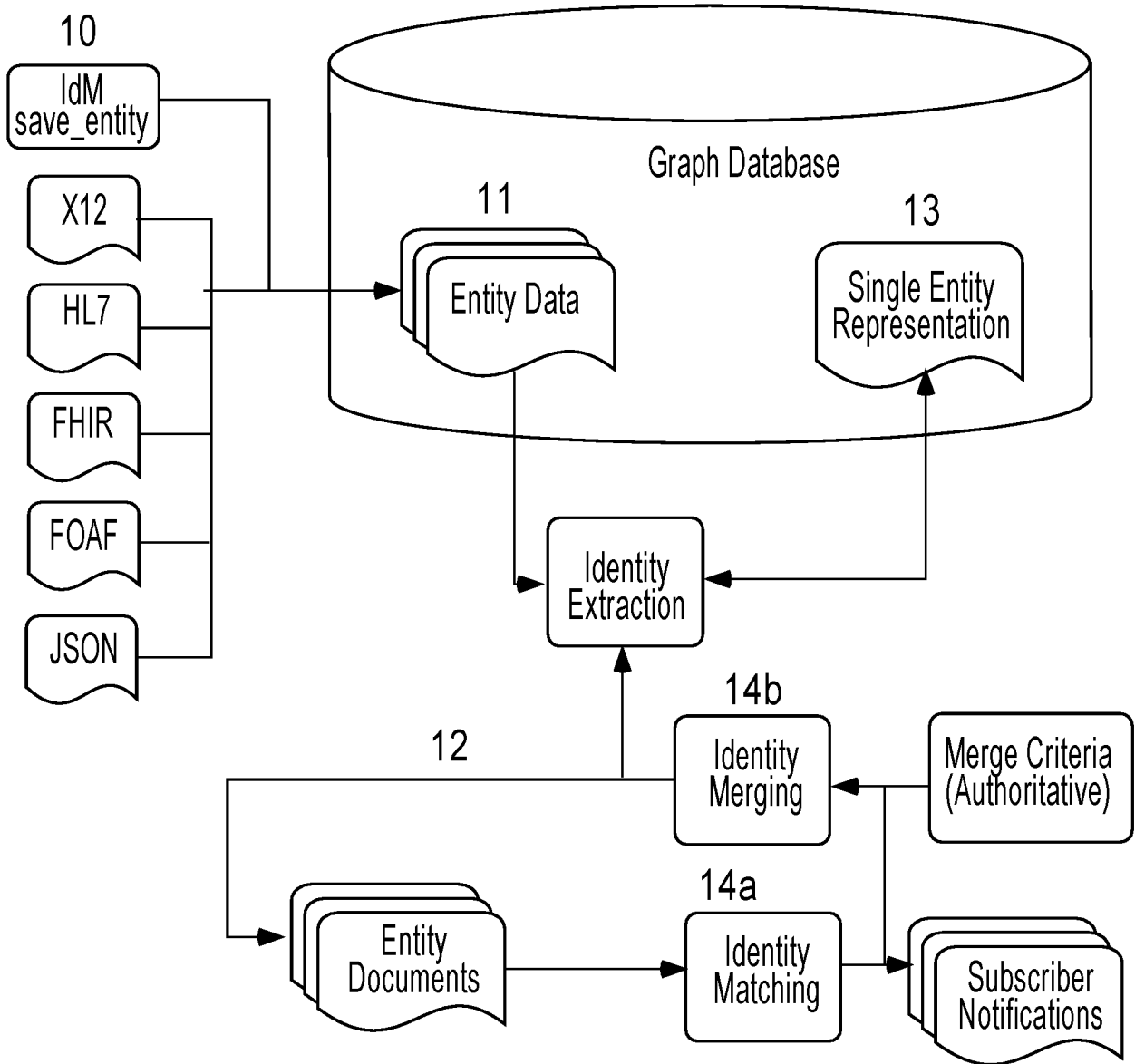


FIGURE 7

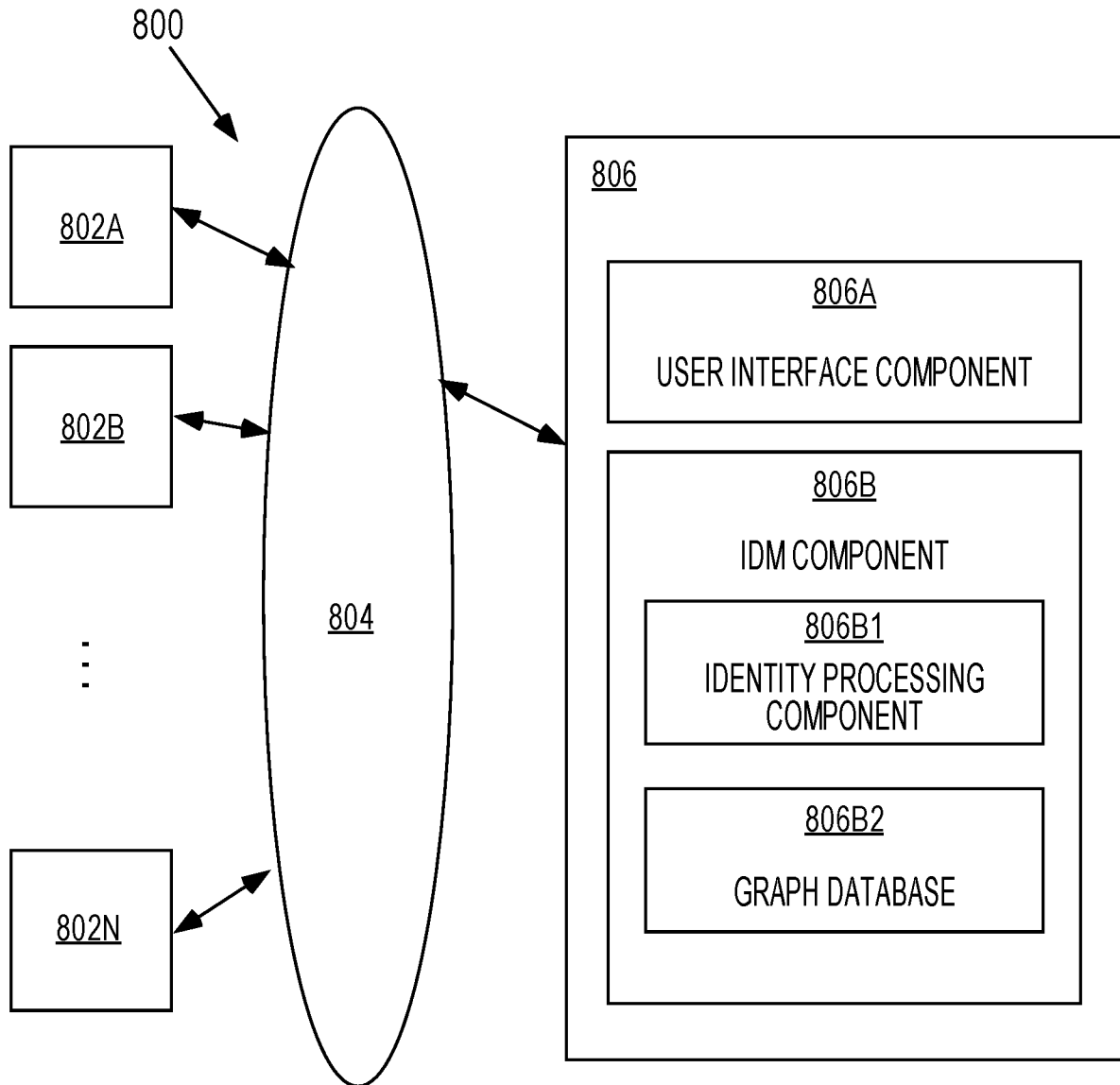


FIGURE 8

<b>INTERNATIONAL SEARCH REPORT</b>		International application No. PCT/US16/56458		
<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC: <b>G06Q 10/00( 2012.01)</b>  USPC: <b>705/2</b> According to International Patent Classification (IPC) or to both national classification and IPC				
<b>B. FIELDS SEARCHED</b>				
Minimum documentation searched (classification system followed by classification symbols) U.S. : 705/2				
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched				
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)				
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>				
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.		
Y	US 2007/0276858 A1 (Cushman, II et al.) 29 November 2007, see entire document.	1-14		
Y	US 2014/0244300 A1 (BESS et al.) 28 August 2014, see entire document.	1-14		
Y	US 2010/0082620 A1 (Jennings, III et al.) 01 April 2010, see entire document.	2, 9		
Y	US 2003/0097359 A1 (Ruediger) 22 May 2003, see entire document.	3, 10		
Y	Chen et al., Adaptive Graphical Approach to Entity Resolution, June 18-23, 2007, Proceedings of the 7th ACM/IEEE-CS joint conference of Digital Libraries, pp.204-213.	6, 13		
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.				
* Special categories of cited documents: <table style="width: 100%; border: none;"> <tr> <td style="width: 50%; border: none;">               "A" document defining the general state of the art which is not considered to be of particular relevance                "E" earlier application or patent 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             </td> <td style="width: 50%; border: none;">               "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                "&amp;" document member of the same patent family             </td> </tr> </table>			"A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent 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
"A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent 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 01 November 2016 (01.11.2016)		Date of mailing of the international search report <b>17 NOV 2016</b>		
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (571) 273-8300		Authorized officer  John Nguyen  Telephone No. 571-272-4300		