US 20240037132A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2024/0037132 A1**

YANNUZZI et al. (43) **Pub. Date: Feb. 1, 2024**

(54) **MAPPING OF APPLICATION DATA**

(71) Applicant: **Cisco Technology, Inc.**, San Jose, CA (US)

(72) Inventors: **Marcelo YANNUZZI**, Vufflens-La-Ville (CH); **Hervé MUYAL**, Gland (CH); **Jean Andrei DIACONU**, Gaillard (FR); **Jelena KLJUJIC**, Deinze (BE); **Carlos GONCALVES PEREIRA**, Carlsbad, CA (US)

(21) Appl. No.: **17/877,529**

(22) Filed: **Jul. 29, 2022**

Publication Classification

(51) **Int. Cl.**

G06F 16/36 (2006.01)

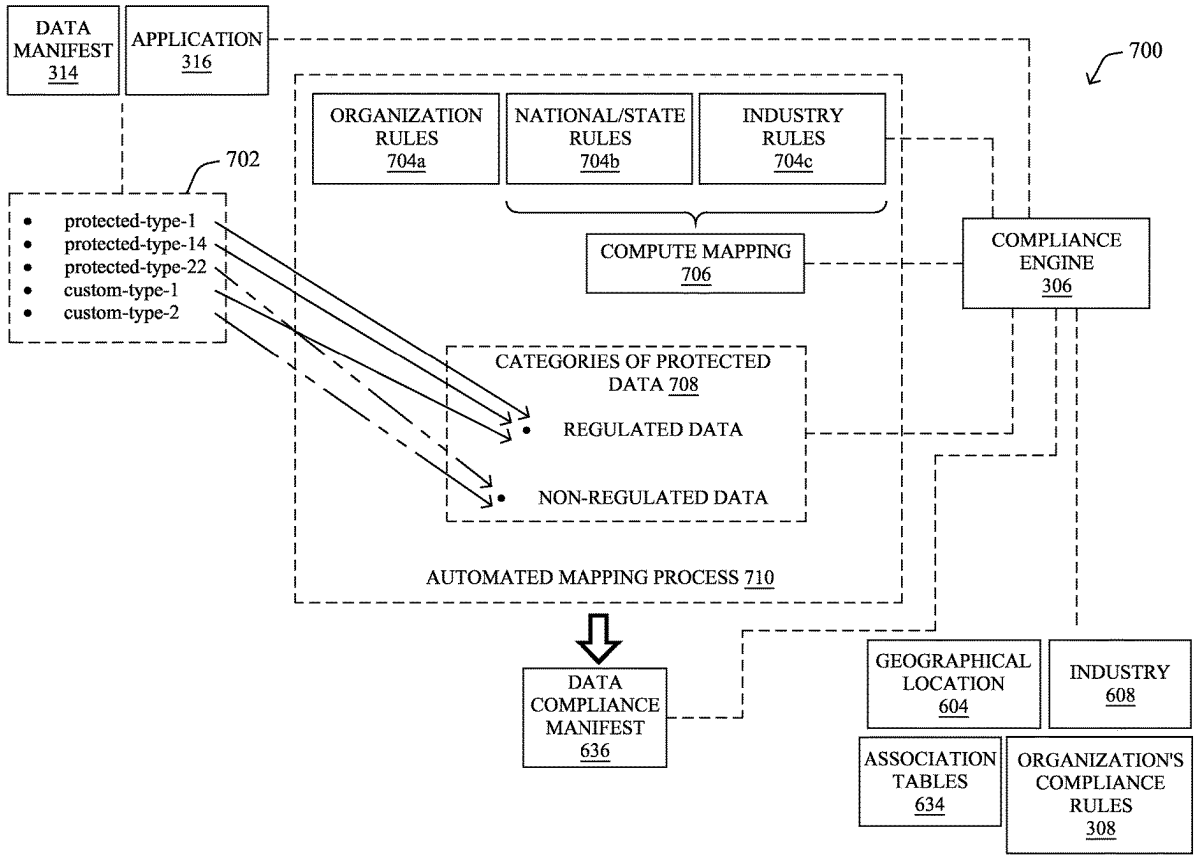
G06F 16/84 (2006.01)

(52) **U.S. Cl.**

CPC ***G06F 16/367*** (2019.01); ***G06F 16/86*** (2019.01)

(57) **ABSTRACT**

In one embodiment, a device obtains an ontology derived from a data usage restriction document and indicative of a category of protected data. The device obtains metadata indicative of a type of data handled by an application. The device creates a mapping between the type of data handled by the application and the category of protected indicated by the ontology. The device generates, based on the mapping, a data compliance manifest used by a workload engine to constrain use of the type of data during execution of the application or used to constrain use of the type of data during deployment of the application.



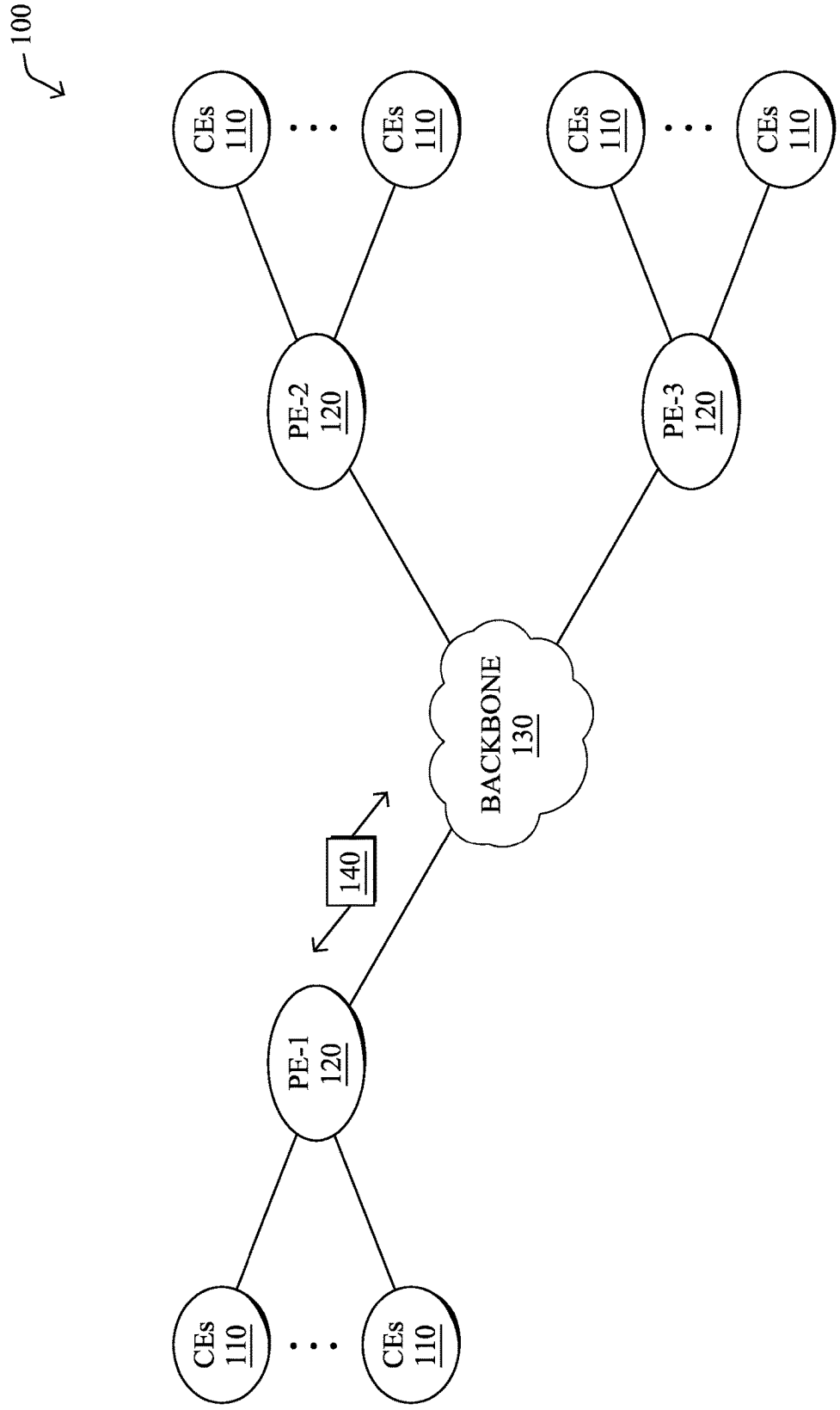

100
CEs 110
CEs 110
CEs 110
CEs 110
PE-2 120
PE-3 120
BACKBONE 130
140
PE-1 120
CEs 110
CEs 110

FIG. 1A

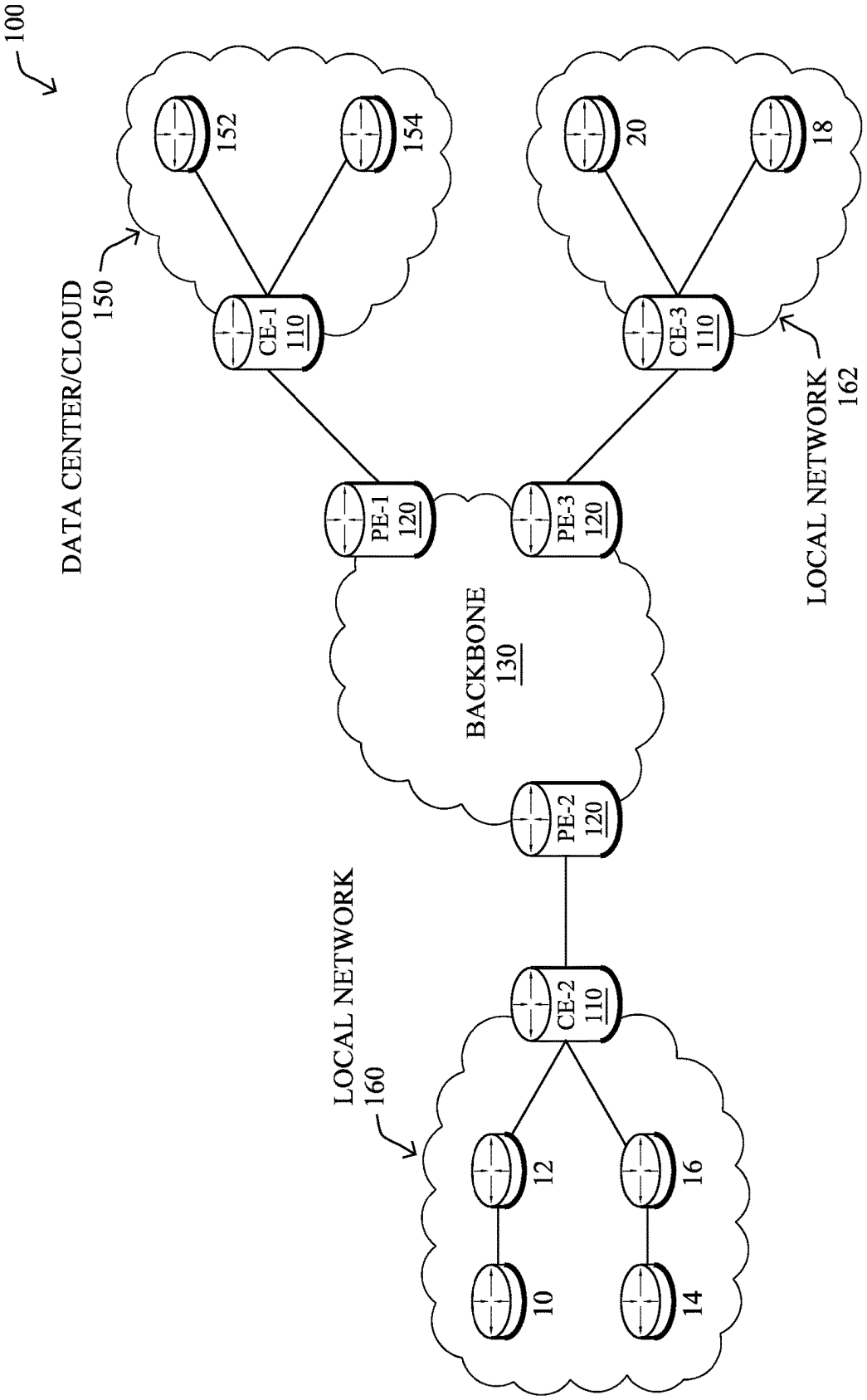

100
DATA CENTER/CLOUD
150
152
154
CE-1
110
20
18
CE-3
110
LOCAL NETWORK
162
PE-1
120
PE-3
120
BACKBONE
130
PE-2
120
LOCAL NETWORK
160
CE-2
110
12
16
10
14

FIG. 1B

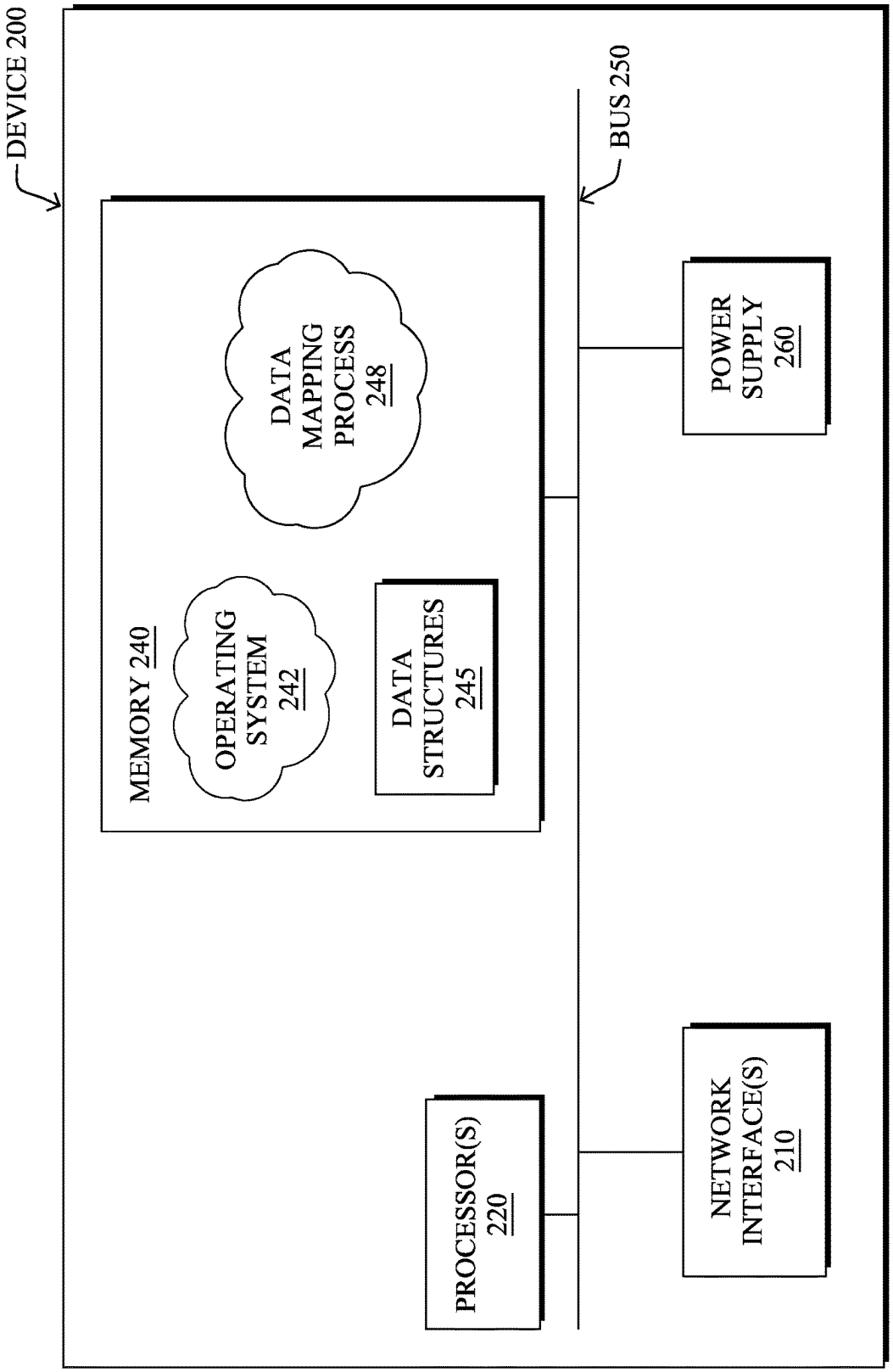

DEVICE 200
BUS 250
MEMORY 240
OPERATING SYSTEM 242
DATA STRUCTURES 245
DATA MAPPING PROCESS 248
POWER SUPPLY 260
PROCESSOR(S) 220
NETWORK INTERFACE(S) 210

FIG. 2

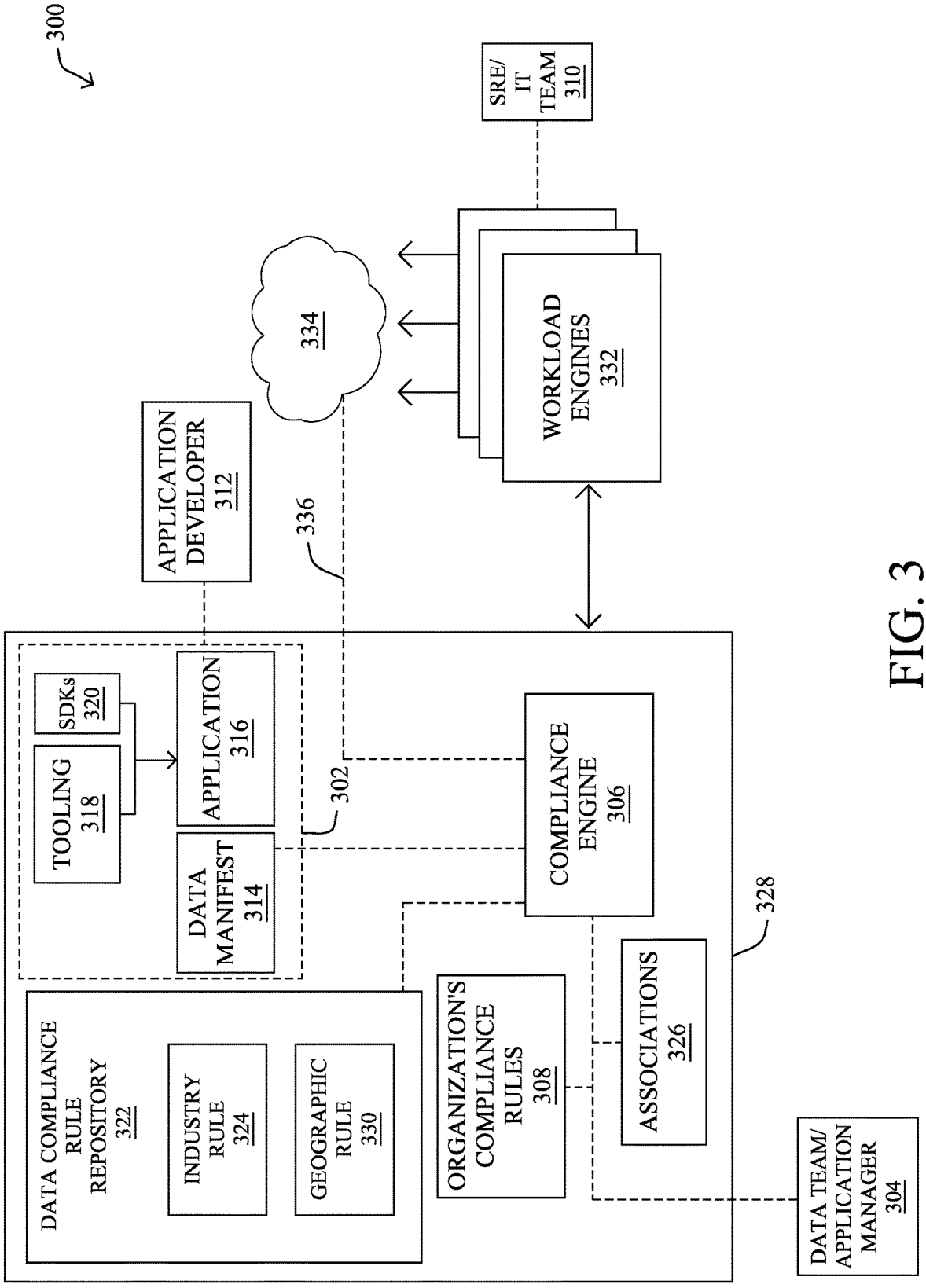

300
SRE/ IT TEAM 310
334
WORKLOAD ENGINES 332
APPLICATION DEVELOPER 312
336
SDKs 320
TOOLING 318
APPLICATION 316
DATA MANIFEST 314
302
COMPLIANCE ENGINE 306
328
DATA COMPLIANCE RULE REPOSITORY 322
INDUSTRY RULE 324
GEOGRAPHIC RULE 330
ORGANIZATION'S COMPLIANCE RULES 308
ASSOCIATIONS 326
DATA TEAM/ APPLICATION MANAGER 304

FIG. 3

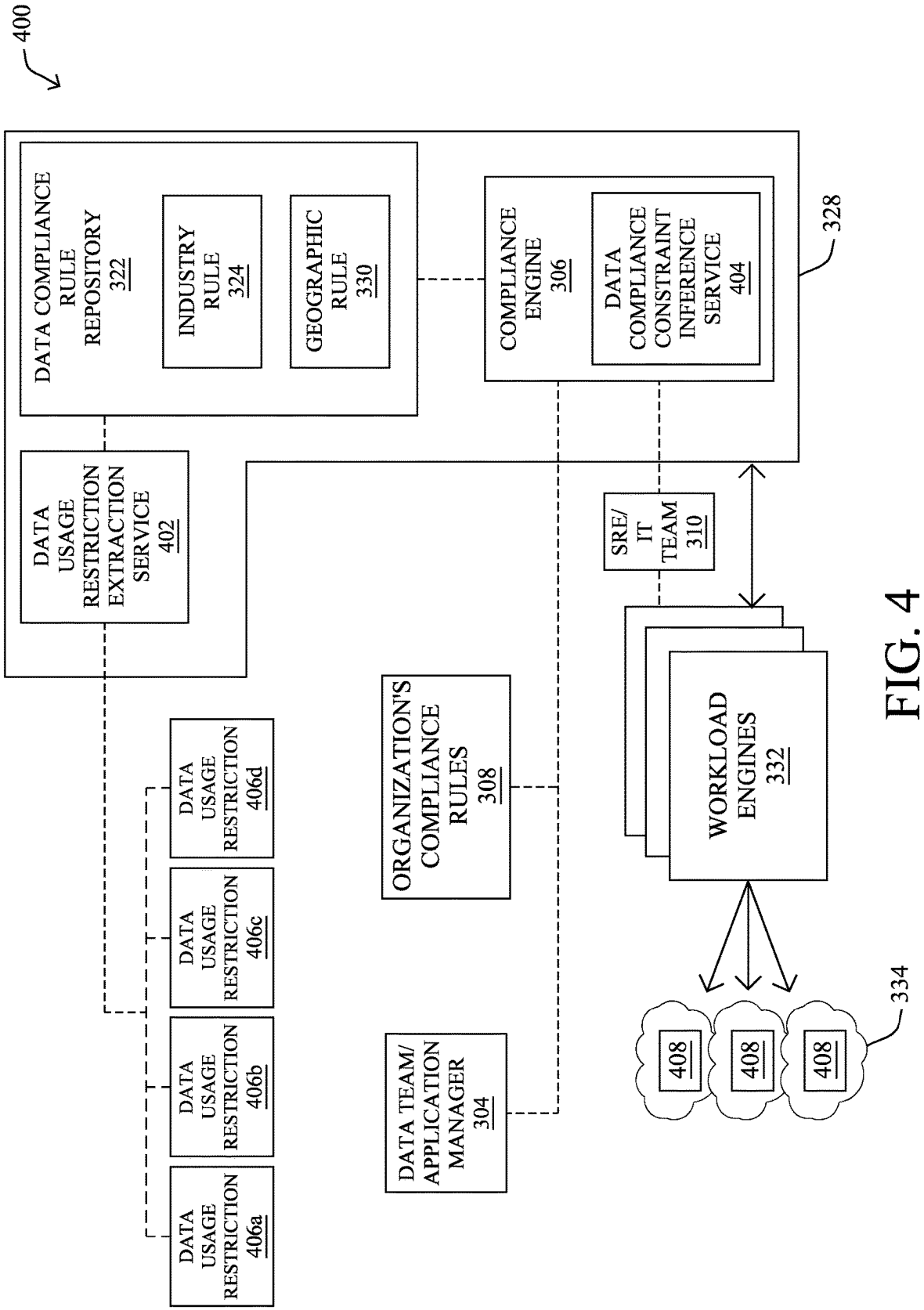

400
DATA COMPLIANCE RULE REPOSITORY 322
INDUSTRY RULE 324
GEOGRAPHIC RULE 330
COMPLIANCE ENGINE 306
DATA COMPLIANCE CONSTRAINT INFERENCE SERVICE 404
328
DATA USAGE RESTRICTION EXTRACTION SERVICE 402
SRE/ IT TEAM 310
WORKLOAD ENGINES 332
ORGANIZATION'S COMPLIANCE RULES 308
DATA USAGE RESTRICTION 406a
DATA USAGE RESTRICTION 406b
DATA USAGE RESTRICTION 406c
DATA USAGE RESTRICTION 406d
DATA TEAM/ APPLICATION MANAGER 304
408
408
408
334

FIG. 4

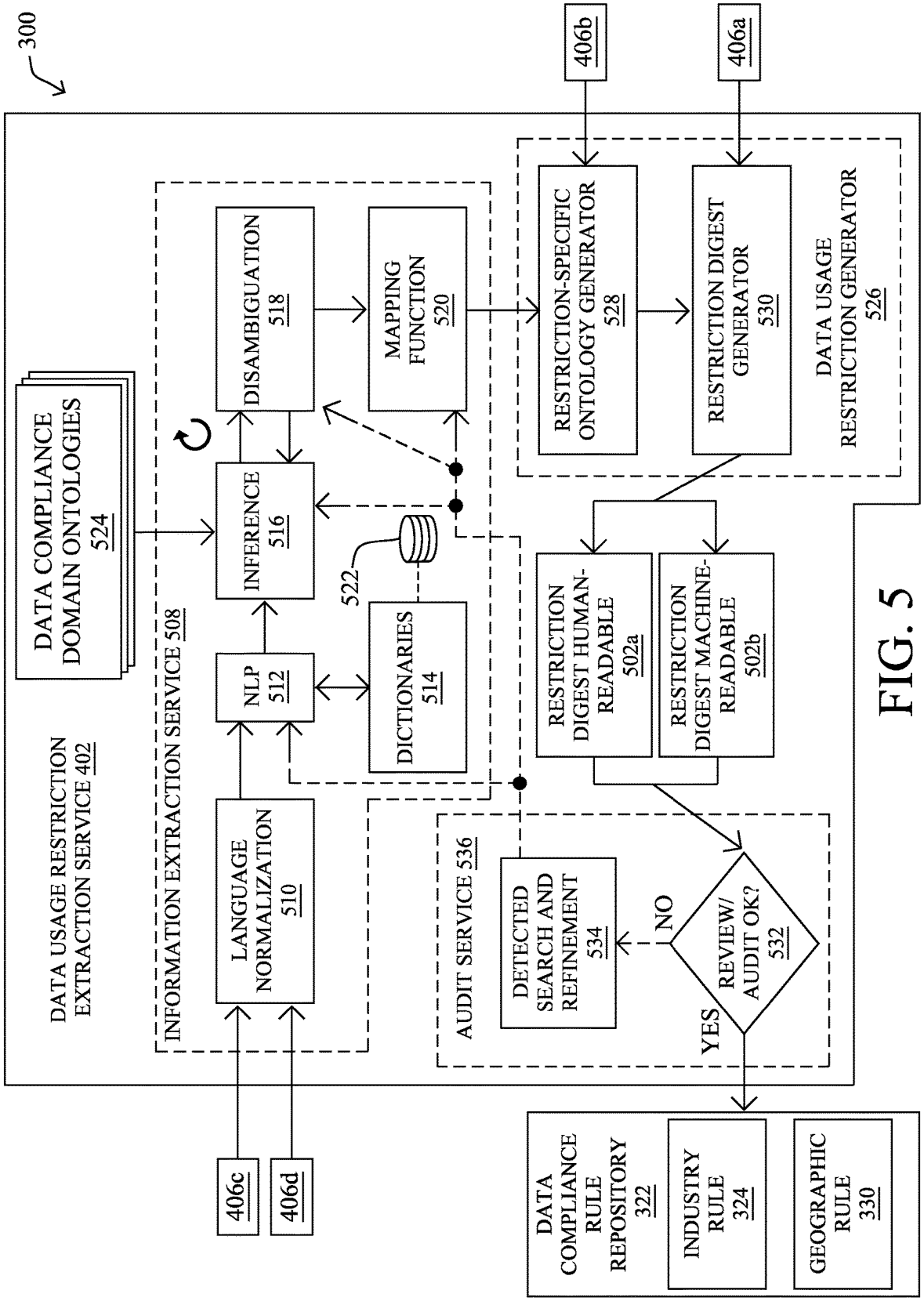

300
406b
406a
DATA USAGE RESTRICTION EXTRACTION SERVICE 402
DATA COMPLIANCE DOMAIN ONTOLOGIES 524
INFORMATION EXTRACTION SERVICE 508
DISAMBIGUATION 518
MAPPING FUNCTION 520
INFERENCE 516
NLP 512
LANGUAGE NORMALIZATION 510
DICTIONARIES 514
522
RESTRICTION-SPECIFIC ONTOLOGY GENERATOR 528
RESTRICTION DIGEST GENERATOR 530
DATA USAGE RESTRICTION GENERATOR 526
RESTRICTION DIGEST HUMAN-READABLE 502a
RESTRICTION DIGEST MACHINE-READABLE 502b
AUDIT SERVICE 536
DETECTED SEARCH AND REFINEMENT 534
REVIEW/ AUDIT OK? 532
NO
YES
406c
406d
DATA COMPLIANCE RULE REPOSITORY 322
INDUSTRY RULE 324
GEOGRAPHIC RULE 330

FIG. 5

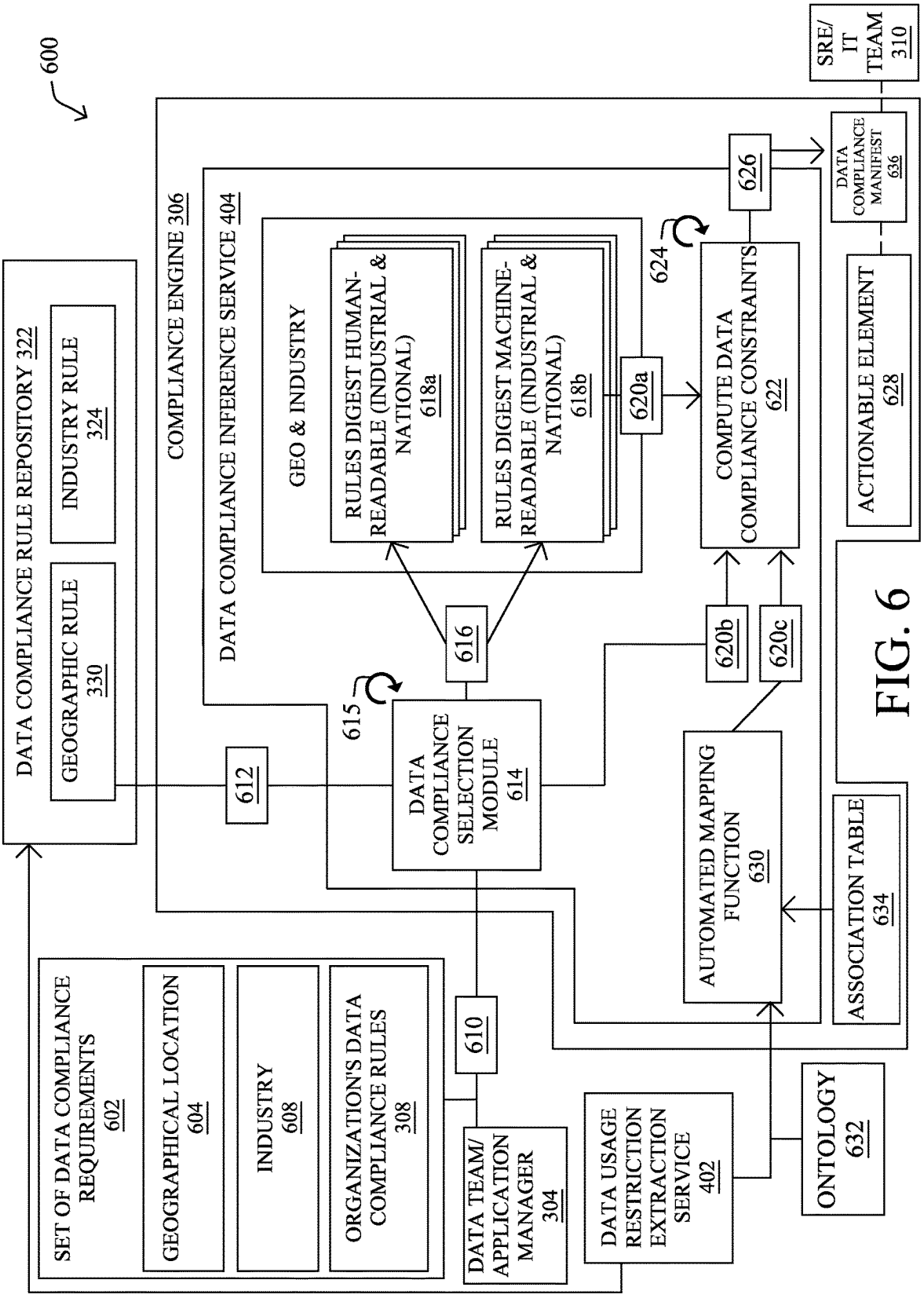

600
SRE/ IT TEAM 310
DATA COMPLIANCE MANIFEST 636
626
624
COMPLIANCE ENGINE 306
DATA COMPLIANCE INFERENCE SERVICE 404
GEO & INDUSTRY
RULES DIGEST HUMAN-READABLE (INDUSTRIAL & NATIONAL) 618a
RULES DIGEST MACHINE-READABLE (INDUSTRIAL & NATIONAL) 618b
620a
COMPUTE DATA COMPLIANCE CONSTRAINTS 622
ACTIONABLE ELEMENT 628
DATA COMPLIANCE RULE REPOSITORY 322
GEOGRAPHIC RULE 330
INDUSTRY RULE 324
612
615
DATA COMPLIANCE SELECTION MODULE 614
616
620b
620c
AUTOMATED MAPPING FUNCTION 630
ASSOCIATION TABLE 634
SET OF DATA COMPLIANCE REQUIREMENTS 602
GEOGRAPHICAL LOCATION 604
INDUSTRY 608
ORGANIZATION'S DATA COMPLIANCE RULES 308
610
DATA TEAM/ APPLICATION MANAGER 304
DATA USAGE RESTRICTION EXTRACTION SERVICE 402
ONTOLOGY 632

FIG. 6

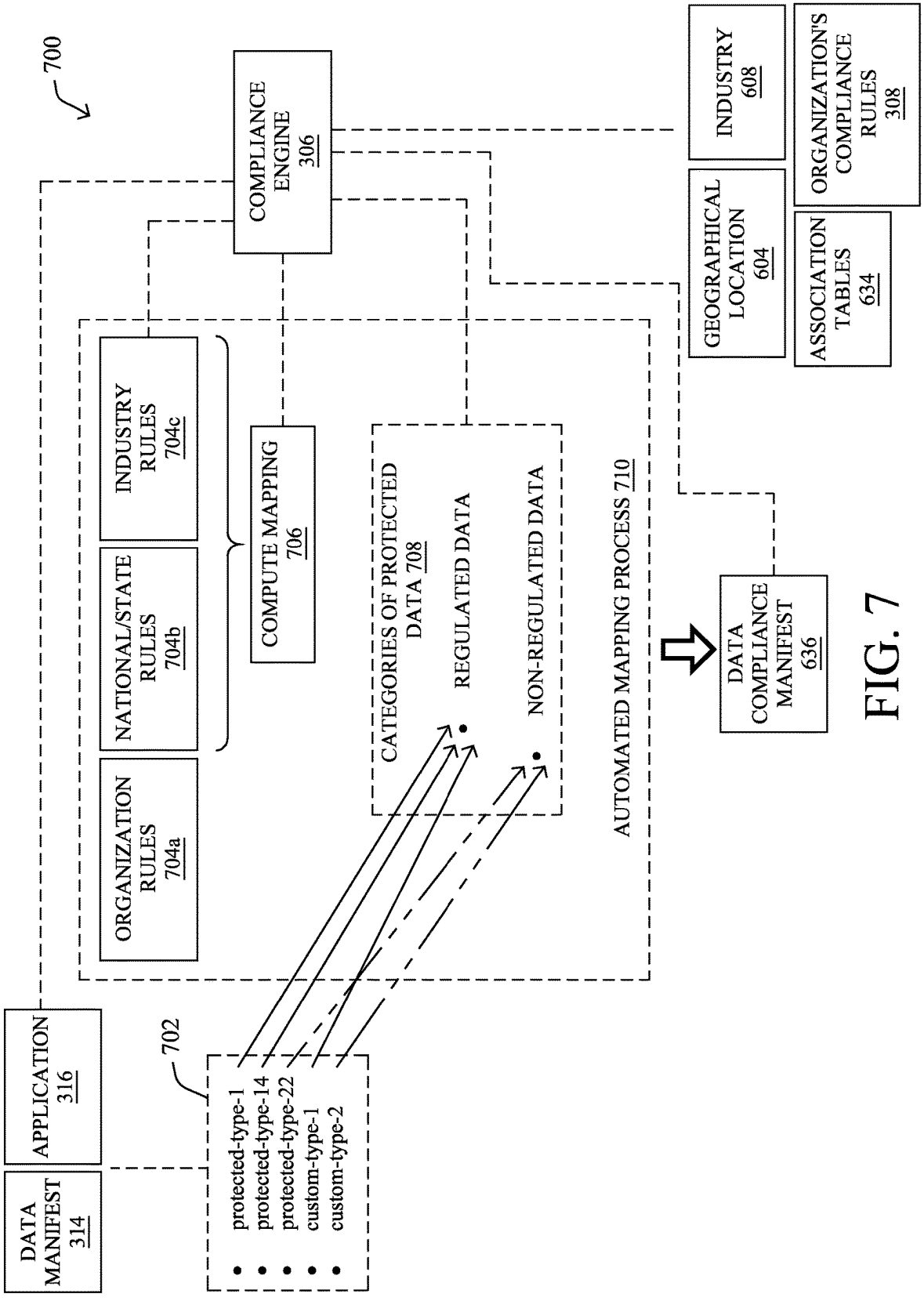

700
COMPLIANCE ENGINE 306
INDUSTRY 608
ORGANIZATION'S COMPLIANCE RULES 308
GEOGRAPHICAL LOCATION 604
ASSOCIATION TABLES 634
ORGANIZATION RULES 704a
NATIONAL/STATE RULES 704b
INDUSTRY RULES 704c
COMPUTE MAPPING 706
CATEGORIES OF PROTECTED DATA 708
REGULATED DATA
NON-REGULATED DATA
AUTOMATED MAPPING PROCESS 710
DATA COMPLIANCE MANIFEST 636
DATA MANIFEST 314
APPLICATION 316
702
protected-type-1
protected-type-14
protected-type-22
custom-type-1
custom-type-2

FIG. 7

900

{} Data Manifest - 45fea578932a-4285d17.json x

Users > mayanmuz > Desktop> { Data Manifest 45faa578932a-2a5d17.json >{ }com

```
{
    "ApplicationName": "Sales Performance",
    "Application Version: 1.8,
    "ApplicationID": "45fea5789328",
    "DataManifestID": "45fea578932a-2a5d17",
    "containers" [
        {
            "name": "postgresql",
            "version": "14.2",
            "paths": []
        },
        {
            "name": "products-api",
            "version": "2.0.1",
            "paths": []
        },
        {
            "name":"sales-analysis-api",
            "version": "1.4.1",
            "paths": |
                {
                    "type":"protected-type-14",
                    "defaultAssociation": "Sales Confidential",
                    "path": "/sales/analysis.5",
                    "handles": {
                        "PII": false,
                        "Confidential": true,
                        "Restricted": true
                    }
                }
            }
        },
        {
            "name": "mysql",
            "version": "5.5.8",
            "paths": |
                {
                    "type":"protected-type-1",
                    "defaultAssociation": "Customer PII",
                    "path": "N/A",
                    "handles": {
                        "PII": true,
                        "Confidential": false,
                        "Restricted": true
                    }
                },
                {
                    "type": "protected-type-14",
                    "defaultAssociation": "Sales Confidential",
                    "path": "/",
                    "handles": {
                        "PII": false,
                        "Confidential": true,
                        "Restricted": true
                    }
```

FIG. 8A

704a

{} Organization DC Rules-45ea5789326-78dc23a87b72.json x

Users > mayasmaa > Desktop > {} Organization DC Rules-45ea5789326-78dc23a87077.json


```
                "status": "compliant"
            },
            "createTime": "2822-02-22T18:06:20.1632",
            "createUser": "daniel.johnson@enterprise.com",
            "modifyTime": "2822-82-22T18:42:24.4772",
            "approverUser": "claudia.smith@enterprise.com",
            "ancestry": {
                "parents":[]
            },
            "metaData": {},
            "DataCompliancesRules": {
              "Rule-1": {
                "ProtectedDataType": "protected-type-1",
                "Association": "Customer PII",
                "States-Provinces":[
                    "Riyadh"
                ],
                "Constraints": {
                    "DataProcessing":{
                        "processing Policy":
                            "LocalRegulation"
                        }
                    },
                    "DataRetention": {
                        "retentionPolicy": [
                            "LocalRegulation"
                        ]
                    },
                    "DataAccess": {
                        "accessPolicy": {
                        "LocalRegulation"
                    },
                    "SelectLocations": {
                        "AccessRight-1": "office Riyadh",
                        "AccessRight-2": "Data Center Riyadh",
                        "AccessRight-3": "Corpliant Public Cloud Zone"
                    },
                    "DataAccessGroups": [
                        "sales",
                        "finance"
                    ]
                }
            }
        },
        "Rule-2": {
            "ProtectedDataType": "protected-type-14",
            "Association": "Sales Confidential",
            "States-Provinces":[
                "Riyadh",
                "Makkah"
            ],
            "Constraints": {
                "DataProcessing": {
                    "processing Policy": {
                        "WithinCountry
                    }
                },
                "DataRetention": {
                    "retentionPolicy":{
                        "WithinCountry"
                    }
                },
                "DataAccess": {
                    "accessPolicy": {
```


FIG. 8B

704b

National DC Rules-79af21698.json x
Users >msyamsr > Desktop Final II National DC Rules-79af21099.json >


```
{
    "Geo":{
        "countries": ["Kingdom of Saudi Arabia (KSA)"],
        "GeoID":"79afa21698",
        "states-provinces": ["all"]
    },
    "Industry": "NA",
    "Law": "POPL - Personal Data Protection Law,
    "Scope": {
        "private-sector": true,
        "public-sector": true
    },
    "DataSubjects": {
        "citizens": true,
        "residents non-citizens": true
    },
    "ForeignOrganizations": {
        "If": {
            "Handles PersonalData": true
        },
        "Then": {
            "POPLApplies": true
        }
    },
    "EffectiveDate": "2022-03-23T00:00:00",
    "EnforcementDate": "2023-03-23T00:00:00",
    "Supervisory Authority": {
        "Entity": "Saudi Data & Artificial Intelligence Authority (SDAIA)",
        "Period1": "2023-03-23T00:00:00-2025-03-22123:59:59",
        "Entity2":"National Data Management Office (NDMO)",
        "Period2": 2025-03-23T00:00:00"
    },
    "Legal": {
        "legalTags": {
            "revised": true,
            "status": "validated"
            }
        },
        "createTime": "2021-12-02T13:16:21.261?",
        "createUser": "system internal",
        "modifyTime": "2022-01-12T18:28:19.2352",
        "ValidatedBy": "legal@enterprise.com",
        "DataComplianceRules": {
            "Rule": {
                "DataSet": "Personal Data",
                    "Constraints": {
                        "DataProcessingOutsideGeo": {
                            "processing Policy": "Denied",
                            "LocalRegulation": "Within Country"
                        },
                        "DataRetentionOutsideGeo": {
                            "processingPolicy":"Denied",
                            "LocalRegulation": "Within Country"
                        },
                        "DataAccessOutsideGeo": {
                            "processing Policy": "Denied",
                        "LocalRegulation": "Within Country"
                            }
                        }
                    }
        },
        "EncodingFormatTypeID": "namespace:data-constraints:national-DC-rules-79afa21698-json:",
        "Checksum", "12a5e4a21d81e30134281238b2af923b",
        "ChecksumAlgorithm":"SHA-256"
}
```


FIG. 8C

704d

{} Industry DC Rules-79afa21698-31dea17495.json

Users > mayanmoz > Desktop > Final > {} Industry DC Rules -79afa21698-31dea17495.json ...


```
{
    "Geo":{
        "countries": ["Kingdom of Saudi Arabia (KSA)"],
        "GeoID": "79af21698",
        "states-provinces": ["all"]
    },
    "Industry":{
        "segments":[
            "Retail",
            "e-Commerce"
    },
    "IndustryID": "79afa21698-31dea17495"
},
"Law": "NA",
    "Scope": "NA",
    "DataSubjects": "NA",
    "ForeignOrganizations": "NA",
    "EffectiveDate": "NA",
    "EnforcementDate": "NA",
    "SupervisoryAuthority": {
        "Entity1": "NA",
        "Period1": "NA",
        "Entity2": "NA",
        "Period2": "NA"
    },
    "Legal": {
        "legaltags": {
            "revised": true,
            "status": "validated"
        }
    },
    "createTime": "2021-12-02T13:17:22.1782",
    "createUser": "system internal",
    "modifyTime": "2022-01-12T11:35:34.5432",
    "ValidatedBy": "legalgenterprise.com",
    "DataCompliance Rules": {
        "Rule": {
            "DataSet": "NA",
                "Constraints": {
                    "DataProcessingOutsideGeo": {
                        "processing Policy": "Na",
                        "LocalRegulation": "NA"
                    },
                    "DataRetentionOutsideGeo": {
                        "processing Policy": "NA",
                        "LocalRegulation": "NA"
                    },
                    "DataAccessOutsideGeo": {
                        "processing Policy": "NA",
                        "LocalRegulation": "NA"
                    }
                }
        }
    },
    "EncodingFormatTypeID": "namespace:data-constraints: industry-DC-rules-79afa21698-31dea17495-json:",
    "Checksum": "45b6a2a37a10e748d9825a41df1b568ac",
    "ChecksumAlgorithm":"SHA-256"
}
```


FIG. 8D

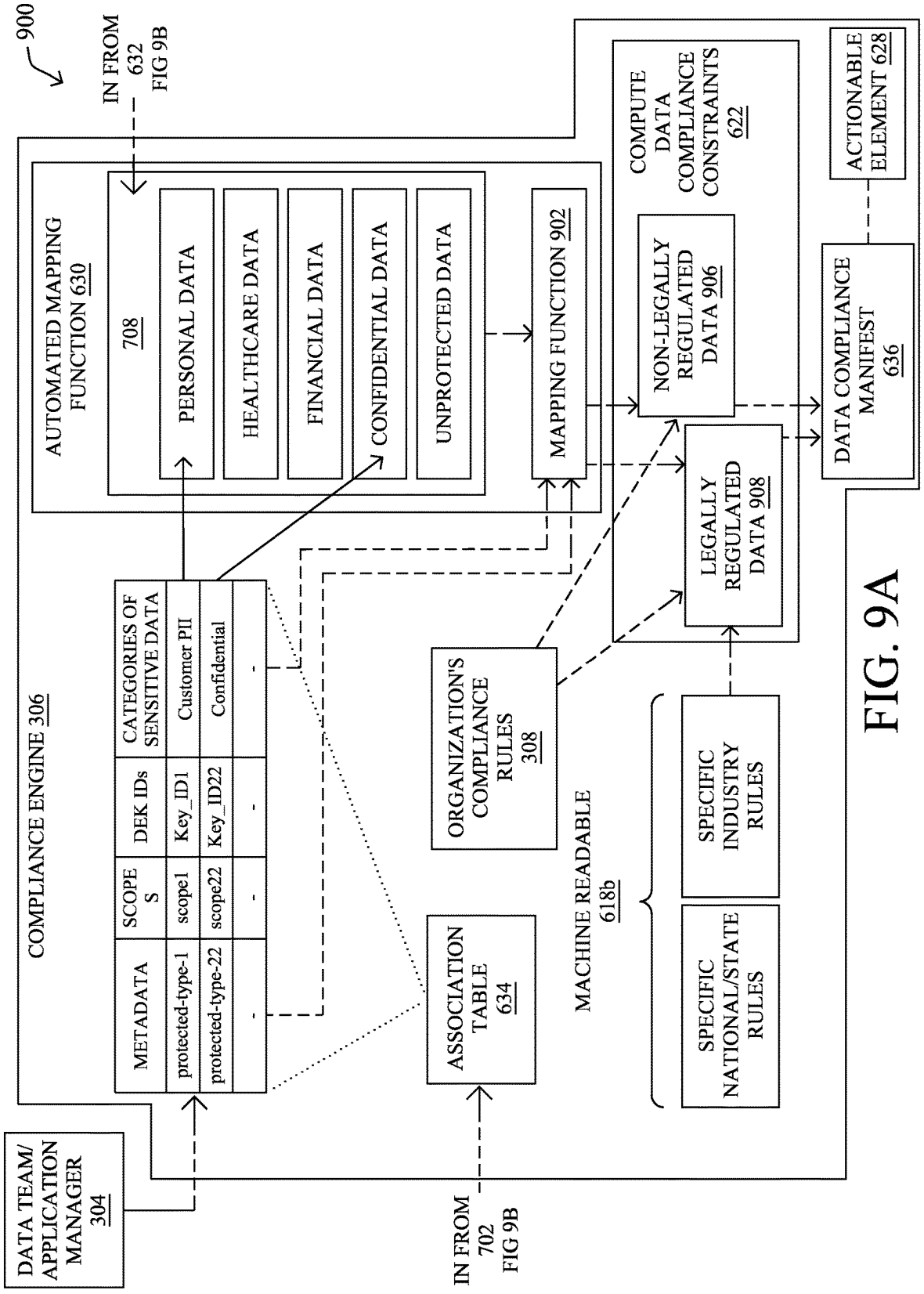

900
IN FROM 632 FIG 9B
AUTOMATED MAPPING FUNCTION 630
708
PERSONAL DATA
HEALTHCARE DATA
FINANCIAL DATA
CONFIDENTIAL DATA
UNPROTECTED DATA
MAPPING FUNCTION 902
COMPUTE DATA COMPLIANCE CONSTRAINTS 622
NON-LEGALLY REGULATED DATA 906
LEGALLY REGULATED DATA 908
ACTIONABLE ELEMENT 628
DATA COMPLIANCE MANIFEST 636
COMPLIANCE ENGINE 306
METADATA
SCOPE S
DEK IDs
CATEGORIES OF SENSITIVE DATA
protected-type-1
protected-type-22
scope1
scope22
Key_ID1
Key_ID22
Customer PII
Confidential
ASSOCIATION TABLE 634
ORGANIZATION'S COMPLIANCE RULES 308
MACHINE READABLE 618b
SPECIFIC INDUSTRY RULES
SPECIFIC NATIONAL/STATE RULES
DATA TEAM/ APPLICATION MANAGER 304
IN FROM 702 FIG 9B

FIG. 9A

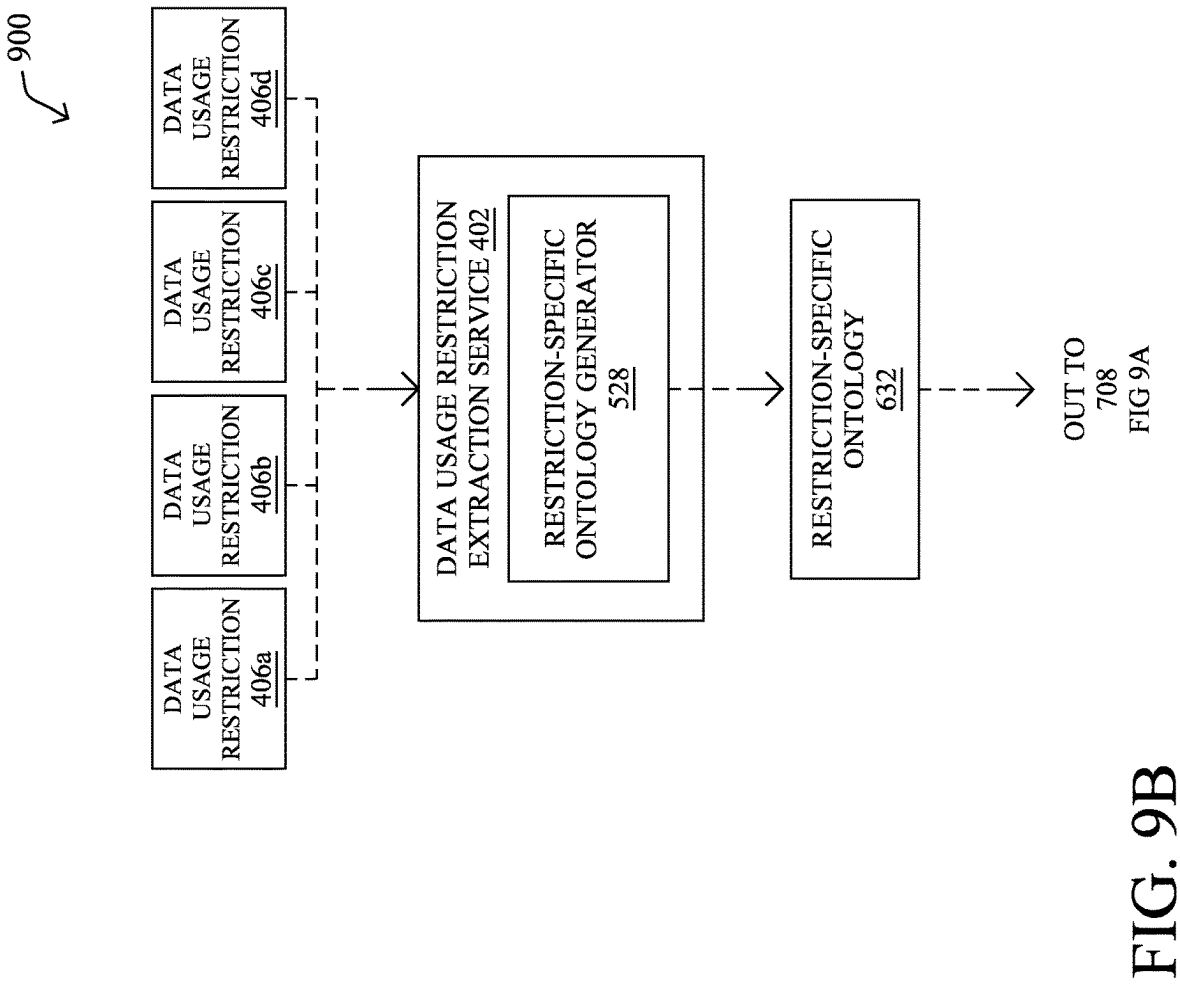

900
DATA USAGE RESTRICTION 406a
DATA USAGE RESTRICTION 406b
DATA USAGE RESTRICTION 406c
DATA USAGE RESTRICTION 406d
DATA USAGE RESTRICTION EXTRACTION SERVICE 402
RESTRICTION-SPECIFIC ONTOLOGY GENERATOR 528
RESTRICTION-SPECIFIC ONTOLOGY 632
OUT TO 708 FIG 9A
METADATA SOURCE 702
PROTECTED TYPES
CUSTOM TYPES
DATA CATALOGS
TAGGED DATA
SOURCE DCaC
OTHER SOURCES
IN FROM 634 FIG 9A
OUT TO 634 FIG 9A

FIG. 9B

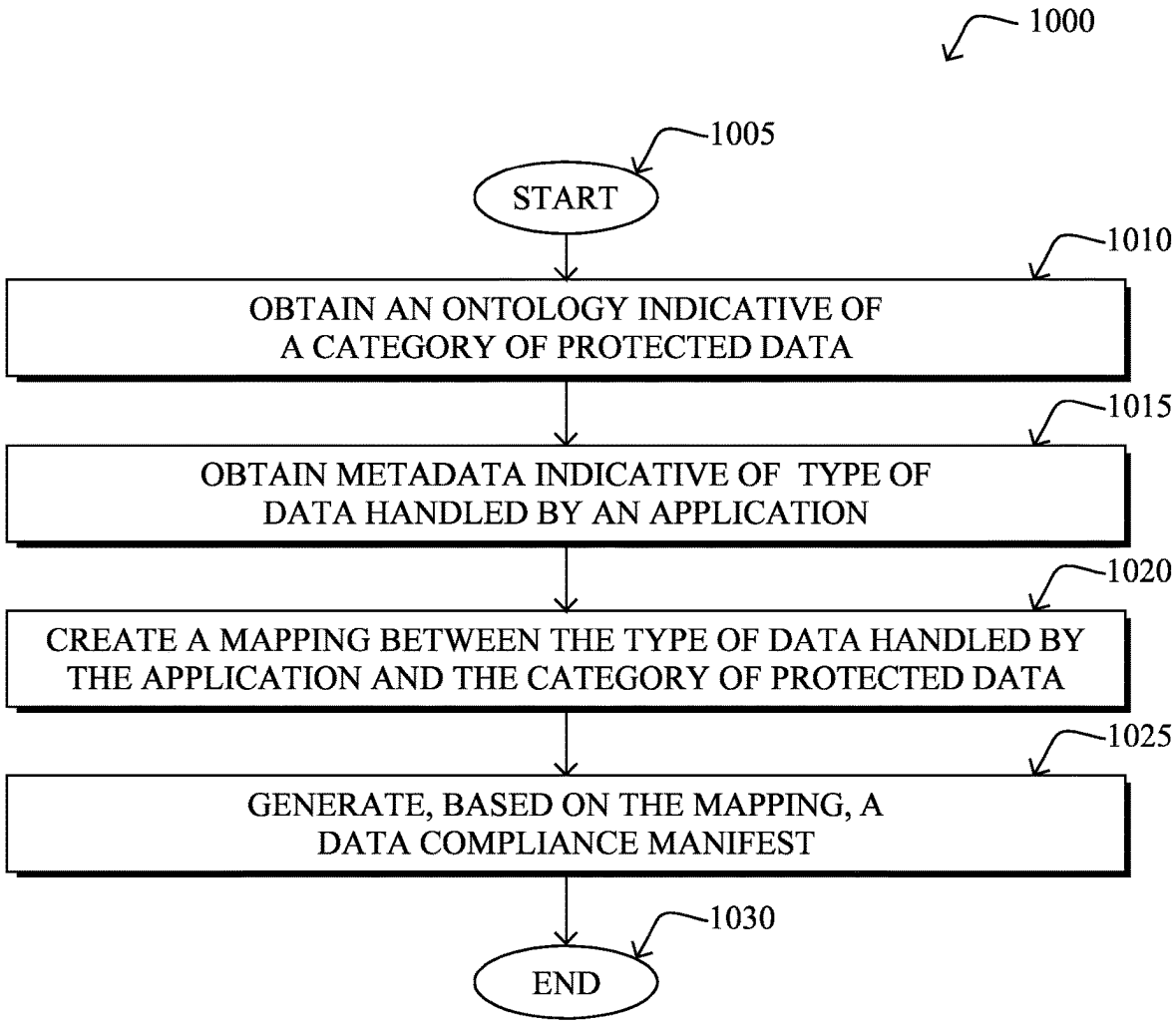

1000
1005
START
1010
OBTAIN AN ONTOLOGY INDICATIVE OF A CATEGORY OF PROTECTED DATA
1015
OBTAIN METADATA INDICATIVE OF TYPE OF DATA HANDLED BY AN APPLICATION
1020
CREATE A MAPPING BETWEEN THE TYPE OF DATA HANDLED BY THE APPLICATION AND THE CATEGORY OF PROTECTED DATA
1025
GENERATE, BASED ON THE MAPPING, A DATA COMPLIANCE MANIFEST
1030
END

FIG. 10

MAPPING OF APPLICATION DATA

TECHNICAL FIELD

[0001] The present disclosure relates generally to the mapping of application data.

BACKGROUND

[0002] Applications operate by handling data. For instance, executing an application can involve the storage, communication, processing, etc. of various types of data. The various types of data may include data whose handling is subject to various regulations. For example, data handling regulations at national, federal, state, industry, and/or organizational levels may be applicable to the data handled by an application.

[0003] The enforcement of data compliance has only been made more complex, and potential violations made more likely, as applications are increasingly being developed as sets of distributed services running across a mix of multi-cloud and edge infrastructures that handle a mix of data types differentially subject to various regulations. Unfortunately, enforcement of data compliance requirements largely occurs non-specifically and in a programmatic blind-spot. Given the current regulatory environment and trends, continuing to treat data compliance as an afterthought in this manner will likely yield increased violations of data compliance regulations which may result in substantial fines, penalties, and/or other negative impacts to data handlers.

BRIEF DESCRIPTION OF THE DRAWINGS

[0004] The embodiments herein may be better understood by referring to the following description in conjunction with the accompanying drawings in which like reference numerals indicate identically or functionally similar elements, of which:

[0005] FIGS. **1**A-**1**B illustrate an example communication network;

[0006] FIG. **2** illustrates an example network device/node;

[0007] FIG. **3** illustrates an example architecture for data compliance;

[0008] FIG. **4** illustrates an example architecture to extract data usage restrictions and generate data compliance constraints;

[0009] FIG. **5** illustrates another example architecture to extract data usage restrictions and generate data compliance constraints;

[0010] FIG. **6** illustrates an example architecture and procedure for automated inference of data compliance constraints; and

[0011] FIG. **7** illustrates an example architecture to perform an automated mapping of application metadata to a category of protected data;

[0012] FIG. **8**A-**8**D illustrate an example of a data manifest and data compliance rules utilizable to perform an automated mapping of application metadata to a category of protected data;

[0013] FIGS. **9**A-**9**B illustrate an example architecture to perform an automated mapping of application metadata to categories of protected data indicated in a restriction-specific ontology; and

[0014] FIG. **10** illustrates an example simplified procedure for mapping application data.

DESCRIPTION OF EXAMPLE EMBODIMENTS

Overview

[0015] According to one or more embodiments of the disclosure, a device obtains an ontology derived from a data usage restriction document and indicative of a category of protected data. The device obtains metadata indicative of a type of data handled by an application. The device creates a mapping between the type of data handled by the application and the category of protected indicated by the ontology. The device generates, based on the mapping, a data compliance manifest that may be used by a workload engine to constrain use of the type of data during execution of the application or used to constrain use of the type of data during deployment of the application.

DESCRIPTION

[0016] A computer network is a geographically distributed collection of nodes interconnected by communication links and segments for transporting data between end nodes, such as personal computers and workstations, or other devices, such as sensors, etc. Many types of networks are available, with the types ranging from local area networks (LANs) to wide area networks (WANs). LANs typically connect the nodes over dedicated private communications links located in the same general physical location, such as a building or campus. WANs, on the other hand, typically connect geographically dispersed nodes over long-distance communications links, such as common carrier telephone lines, optical lightpaths, synchronous optical networks (SONET), or synchronous digital hierarchy (SDH) links, or Powerline Communications (PLC) such as IEEE 61334, IEEE P1901.2, and others. The Internet is an example of a WAN that connects disparate networks throughout the world, providing global communication between nodes on various networks. The nodes typically communicate over the network by exchanging discrete frames or packets of data according to predefined protocols, such as the Transmission Control Protocol/Internet Protocol (TCP/IP). In this context, a protocol consists of a set of rules defining how the nodes interact with each other. Computer networks may be further interconnected by an intermediate network node, such as a router, to extend the effective "size" of each network.

[0017] Smart object networks, such as sensor networks, in particular, are a specific type of network having spatially distributed autonomous devices such as sensors, actuators, etc., that cooperatively monitor physical or environmental conditions at different locations, such as, e.g., energy/power consumption, resource consumption (e.g., water/gas/etc. for advanced metering infrastructure or "AMI" applications) temperature, pressure, vibration, sound, radiation, motion, pollutants, etc. Other types of smart objects include actuators, e.g., responsible for turning on/off an engine or perform any other actions. Sensor networks, a type of smart object network, are typically shared-media networks, such as wireless or PLC networks. That is, in addition to one or more sensors, each sensor device (node) in a sensor network may generally be equipped with a radio transceiver or other communication port such as PLC, a microcontroller, and an energy source, such as a battery. Often, smart object networks are considered field area networks (FANs), neighborhood area networks (NANs), personal area networks (PANs), etc. Generally, size and cost constraints on smart

object nodes (e.g., sensors) result in corresponding constraints on resources such as energy, memory, computational speed and bandwidth.

[0018] FIG. **1**A is a schematic block diagram of an example computer network **100** illustratively comprising nodes/devices, such as a plurality of routers/devices interconnected by links or networks, as shown. For example, customer edge (CE) routers **110** may be interconnected with provider edge (PE) routers **120** (e.g., PE-1, PE-2, and PE-3) in order to communicate across a core network, such as an illustrative network backbone **130**. For example, routers **110**, **120** may be interconnected by the public Internet, a multiprotocol label switching (MPLS) virtual private network (VPN), or the like. Data packets **140** (e.g., traffic/messages) may be exchanged among the nodes/devices of the computer network **100** over links using predefined network communication protocols such as the Transmission Control Protocol/Internet Protocol (TCP/IP), User Datagram Protocol (UDP), Asynchronous Transfer Mode (ATM) protocol, Frame Relay protocol, or any other suitable protocol. Those skilled in the art will understand that any number of nodes, devices, links, etc. may be used in the computer network, and that the view shown herein is for simplicity.

[0019] In some implementations, a router or a set of routers may be connected to a private network (e.g., dedicated leased lines, an optical network, etc.) or a virtual private network (VPN), such as an MPLS VPN thanks to a carrier network, via one or more links exhibiting very different network and service level agreement characteristics. For the sake of illustration, a given customer site may fall under any of the following categories:

[0020] 1.) Site Type A: a site connected to the network (e.g., via a private or VPN link) using a single CE router and a single link, with potentially a backup link (e.g., a 3G/4G/5G/LTE backup connection). For example, a particular CE router **110** shown in network **100** may support a given customer site, potentially also with a backup link, such as a wireless connection.

[0021] 2.) Site Type B: a site connected to the network by the CE router via two primary links (e.g., from different Service Providers), with potentially a backup link (e.g., a 3G/4G/5G/LTE connection). A site of type B may itself be of different types:

[0022] 2a.) Site Type B1: a site connected to the network using two MPLS VPN links (e.g., from different Service Providers), with potentially a backup link (e.g., a 3G/4G/5G/LTE connection).

[0023] 2b.) Site Type B2: a site connected to the network using one MPLS VPN link and one link connected to the public Internet, with potentially a backup link (e.g., a 3G/4G/5G/LTE connection). For example, a particular customer site may be connected to network **100** via PE-3 and via a separate Internet connection, potentially also with a wireless backup link.

[0024] 2c.) Site Type B3: a site connected to the network using two links connected to the public Internet, with potentially a backup link (e.g., a 3G/4G/5G/LTE connection).

[0025] Notably, MPLS VPN links are usually tied to a committed service level agreement, whereas Internet links may either have no service level agreement at all or a loose service level agreement (e.g., a "Gold Package" Internet service connection that guarantees a certain level of performance to a customer site).

[0026] 3.) Site Type C: a site of type B (e.g., types B1, B2 or B3) but with more than one CE router (e.g., a first CE router connected to one link while a second CE router is connected to the other link), and potentially a backup link (e.g., a wireless 3G/4G/5G/LTE backup link). For example, a particular customer site may include a first CE router **110** connected to PE-2 and a second CE router **110** connected to PE-3.

[0027] FIG. **1**B illustrates an example of network **100** in greater detail, according to various embodiments. As shown, network backbone **130** may provide connectivity between devices located in different geographical areas and/or different types of local networks. For example, network **100** may comprise local/branch networks **160**, **162** that include devices/nodes **10-16** and devices/nodes **18-20**, respectively, as well as a data center/cloud environment **150** that includes servers **152-154**. Notably, local networks **160-162** and data center/cloud environment **150** may be located in different geographic locations.

[0028] Servers **152-154** may include, in various embodiments, a network management server (NMS), a dynamic host configuration protocol (DHCP) server, a constrained application protocol (CoAP) server, an outage management system (OMS), an application policy infrastructure controller (APIC), an application server, etc. As would be appreciated, network **100** may include any number of local networks, data centers, cloud environments, devices/nodes, servers, etc.

[0029] In some embodiments, the techniques herein may be applied to other network topologies and configurations. For example, the techniques herein may be applied to peering points with high-speed links, data centers, etc.

[0030] According to various embodiments, a software-defined WAN (SD-WAN) may be used in network **100** to connect local network **160**, local network **162**, and data center/cloud environment **150**. In general, an SD-WAN uses a software defined networking (SDN)-based approach to instantiate tunnels on top of the physical network and control routing decisions, accordingly. For example, as noted above, one tunnel may connect router CE-2 at the edge of local network **160** to router CE-1 at the edge of data center/cloud environment **150** over an MPLS or Internet-based service provider network in backbone **130**. Similarly, a second tunnel may also connect these routers over a 4G/5G/LTE cellular service provider network. SD-WAN techniques allow the WAN functions to be virtualized, essentially forming a virtual connection between local network **160** and data center/cloud environment **150** on top of the various underlying connections. Another feature of SD-WAN is centralized management by a supervisory service that can monitor and adjust the various connections, as needed.

[0031] FIG. **2** is a schematic block diagram of an example node/device **200** (e.g., an apparatus) that may be used with one or more embodiments described herein, e.g., as any of the computing devices shown in FIGS. **1**A-**1**B, particularly the PE routers **120**, CE routers **110**, nodes/device **10-20**, servers **152-154** (e.g., a network controller/supervisory service located in a data center, etc.), any other computing device that supports the operations of network **100** (e.g., switches, etc.), or any of the other devices referenced below. The device **200** may also be any other suitable type of device depending upon the type of network architecture in place, such as IoT nodes, etc. Device **200** comprises one or more

network interfaces **210**, one or more processors **220**, and a memory **240** interconnected by a system bus **250**, and is powered by a power supply **260**.

[0032] The network interfaces **210** include the mechanical, electrical, and signaling circuitry for communicating data over physical links coupled to the network **100**. The network interfaces may be configured to transmit and/or receive data using a variety of different communication protocols. Notably, a physical network interface **210** may also be used to implement one or more virtual network interfaces, such as for virtual private network (VPN) access, known to those skilled in the art.

[0033] The memory **240** comprises a plurality of storage locations that are addressable by the processor(s) **220** and the network interfaces **210** for storing software programs and data structures associated with the embodiments described herein. The processor **220** may comprise necessary elements or logic adapted to execute the software programs and manipulate the data structures **245**. An operating system **242** (e.g., the Internetworking Operating System, or IOS®, of Cisco Systems, Inc., another operating system, etc.), portions of which are typically resident in memory **240** and executed by the processor(s), functionally organizes the node by, inter alia, invoking network operations in support of software processors and/or services executing on the device. These software processors and/or services may comprise data mapping process **248**, as described herein, any of which may alternatively be located within individual network interfaces.

[0034] FIG. **3** illustrates an example architecture **300** for data compliance, according to various embodiments. The architecture **300** may include a data compliance as code process **328**. Data compliance as code process **328** may be utilized to provide configuration, observability, and enforcement of data compliance rules. Data compliance as code process **328** may accomplish these functions utilizing Data Compliance as Code (DCaC).

[0035] DCaC may include integrating a data compliance mechanism into the program code of an application and/or application service. For example, data compliance as code process **328** may be utilized to build data compliance into the application development process, supported by automated code annotations, bindings between such annotations and categories of sensitive data, and controls at code, build, and pre-deploy time. Data compliance as code process **328** may provide a mechanism whereby application developers proactively assist data teams, application managers, and legal departments with data compliance, while ensuring that developers may remain oblivious to specific regulations, data related obligations, or compliance requirements that organizations might have across different regions.

[0036] For example, data compliance as code process **328** may include data annotating process **302**. Data annotating process **302** may facilitate application developer **312** automatically adding metadata to program code of an application **316** during the development of the application **316**. In various embodiments, this may be performed by automated annotations of data fields in the program code and by the creation of references to such annotations at code-build time. These references to annotated code may be automatically rendered in the form of machine-readable data manifest **314**.

[0037] More specifically, data annotating process **302** may provide a mechanism for automated annotations of the program code of application **316**, including classes, application programming interfaces (APIs), and the resulting data at code/build time (e.g., by implementing a Low-Code/No-Code approach supported by software development kits (SDKs) **320** and tooling **318**). Application developers may utilize SDKs **320** and tooling **318** to automatically label data topics, data producers, data consumers, data processors, data holders, etc. For instance, developers may label certain data by annotating it with a data type identifier. For example, a developer may annotate certain data as "protected-type-1," or other data as "protected-type-2," and so on.

[0038] SDKs **320** in data annotating process **302** may provide a set of predefined data types out-of-the-box, including associations by default to specific categories of sensitive data. Sensitive data may include a type of data that may be considered personal, private, sensitive, confidential, protected, secret, restricted, personally identifiable information (PII), etc. In some examples, sensitive data may include data that is subject to regulation. For example, Table 1 lists examples of predefined protected data types and default associations to some examples of categories of sensitive data.

TABLE 1

PROTECTED DATA TYPE	DEFAULT ASSOCIATION
protected-type-1	Customer PII
protected-type-2	Employee PII
. . .	. . .
protected-type-23	Patient Analysis Results
. . .	. . .
protected-type-41	Sales Confidential
. . .	. . .
protected-type-56	Restricted HR
. . .	. . .
unprotected	NA

[0039] A list of the associations, such as the example illustrated in Table 1, may provide associations by default to several categories of sensitive data, including but not limited to PII, confidential, restricted, and unprotected data. In some embodiments, the set of predefined protected data types might be standardized or rely on an existing taxonomy.

[0040] SDKs **320** in data annotating process **302** may also provide a mechanism to define and use custom data types in annotating program data of the application **316**. For example, custom data types may be utilized, which identify protected data types that are not covered by any of those available by default in SDKs **320**. For example, "custom-type-1" might be a custom data type associated to a category of sensitive data such as "Restricted Employee Poll." In various embodiments, the generation and/or insertion of the annotations into the program code of the application **316** may be accomplished by an automated process (e.g., a programmatic identification of data of a particular data type triggering an automated insertion of an annotation of the data as the particular data type, etc.), a partially automated process (e.g., a programmatic flagging of data of a particular data type with a supervised or manual annotation of the data as the particular data type, etc.), and/or a manual process (e.g., a manual flagging of data of a particular data type and/or a manual annotation of the data as the particular data type, etc.).

[0041] In various embodiments, associations between protected data types and categories of sensitive data may be assigned and/or instrumented by different organizations and

at different moments in time. In some cases, the association between protected data types and categories of sensitive data may be assigned by application developers **312** at code/build time. This might be the case when the team of application developers **312** is part of, or develops for, the organization that may use or manage the application **316**. In such cases, the team of application developers **312** might have sufficient knowledge about the data and their use, so that they may either use the associations provided by default or create custom ones.

[0042] In additional instances, application developers **312** of application **316** and/or the users of the application **316** might belong to different organizations. For example, this may be the case when application developers **312** are a DevSecOps team that develops an application **316** that may be used across different organizations, industries, etc. In such cases, application developers **312** may be unaware of the categories of data that should be assigned by a data team and/or application manager **304** in another organization (e.g., precisely what data is confidential and what data is not with respect to that organization and its use of the application **316**). In these instances, application developers **312** may leverage SDKs **320** and tooling **318** to approach data labeling and association in a manner that sidesteps the knowledge deficit while still instilling the functionality. For example, the application developers **312** may leverage SDKs **320** and tooling **318** to automatically add the different classes of protected data type at code and build time (e.g., utilizing predefined and custom protected data types). Additionally, or alternatively, the application developers **312** may leverage SDKs **320** and tooling **318** to automatically insert references in the form of machine-readable descriptions for the protected data types that may be used to generate data manifest **314** bound to application **316** at build time.

[0043] The protected data type annotations and their corresponding references may be utilized by a data team and/or application manager **304** in another organization to select and/or create automated associations **326** between categories of annotated data in the application **316** (e.g., metadata provided by application developers **312**) and specific categories of sensitive data (e.g., personal data, private data, sensitive data, confidential data, protected data, secret data, restricted data, PII, etc.). For instance, each protected data type might be bonded to a class of tokens (e.g., JSON Web Tokens with a specific scope), which in turn might represent different categories of sensitive data for a data team and/or application manager **304**.

[0044] In a specific example, an API call for application **316** may be labeled by application developers **312** with a data type identifier such as "custom-type-7" at code/build time. The "custom-type-7" labeled API call may attempt to access certain data using its bound token (e.g., "Token 7") with a scope defined by, for example, a data team and/or application manager **304** before application **316** was deployed. From the data team and/or application manager **304** perspective, the attempt to access this data may translate to a request to access, for instance, "Confidential Partner" data. As such, the data type labels, and their associations may be utilized as an automated data mapping between the programmatic operations of application **316** and the sensitive data implicated in those operations. In various embodiments, these associations and functionalities may be supported by compliance engine **306** based on the selection, configuration, and automation of data compliance rules before application **316** is deployed and/or post-deployment.

[0045] In some examples, application developers **312**, which again may be a DevSecOps team, might opt for a hybrid approach to generating these associations. For example, this may be the case when making some custom associations between data types and categories of sensitive data or using those predefined in the system (e.g., "protected-type-1" to "Customer PII") might not only be trivial for the application developers **312** but also may facilitate the task of a data team and/or application manager **304** in defining associations. However, other associations might not be apparent to application developers **312**. Hence, certain data in application **316** may be labeled as "protected types" along with their corresponding machine-readable descriptions in data manifest **314**, though they may remain unassigned to a specific category of sensitive data, so they can be associated later by a data team and/or application manager **304** before the application is deployed, or by an automated data lineage, classification, and tagging process at run time (e.g., during the testing phase, that is, before the application is deployed in production).

[0046] In some embodiments, a data team and/or application manager **304** may be provided with a mechanism to change the associations created by application developers **312** or even associate more than one category of sensitive data to a given data type (e.g., a data team and/or application manager **304** may associate certain data with both "Employee PIT" and "Confidential Data"). Hence, two categories of data compliance policies (e.g., one for "Employee PIT" and another for "Confidential Data") may apply and restrict even further the access to this category of data. In general, a data team and/or application manager **304** may be able to Create, Read, Update, or Delete (CRUD) any association between the metadata provided by application developers **312** and categories of sensitive data.

[0047] In various embodiments, a data team and/or application manager **304** may proactively create a set of custom data types. A data team and/or application manager **304** may provide the set of custom data types to application developers **312**. Application developers **312** may then utilize the set of custom data types so that application **316** is annotated at development based on guidelines (e.g., the set of custom data types, etc.) provided beforehand by the data team and/or application manager **304**.

[0048] In additional embodiments, application developers **312** and a data team and/or application manager **304** may collaborate to annotate application **316**. For example, application developers **312** and a data team and/or application manager **304** may iterate in the annotation and association processing in an agile manner. For example, the iteration may be performed as part of a continuous integration/continuous delivery (CI/CD) pipeline (e.g., at testing, staging, and production).

[0049] In some examples, application **316** may be composed of several services developed with different programming languages. Therefore, application **316** may utilize different SDKs **320**. In some instances, the annotation methods and terminology applied to application **316** may vary depending on the programming language (e.g., usually referred to as attributes in C #, decorators in Python, annotations in Golang, etc.). In such cases, tooling **318** of data annotating process **302** may examine the different predefined and custom data types used with different SDKs

320, perform checks, and ensure consistency in the annotations and enumeration across the different services at build time. For example, these consistency checks may ensure that a given "custom-type-X" data type identifier represents the same type of data across services programmed using different programming languages even if they were programmed by different developers. Overall, the data annotating process 302 may provide different degrees of freedom to application developers 312, data teams and/or application managers 304, and the number of protected data types used, and their corresponding associations may vary depending on the type of application 316.

[0050] Data annotating process 302 may, as described above, be utilized in generating automated data references. Specifically, data annotating process 302 may automatically render a data manifest 314 bonded to application 316 at build time. Data manifest 314 may provide machine-readable descriptions of the predefined and/or custom data types used by application 316. A combination of SDKs 320 and tooling 318 may facilitate the instrumentation and automation of the program code at build time, including the automated rendering of data manifest 314. In some cases, application 316 may be composed of various containers. Each container may be built and packaged with its own data manifest, such that the final data manifest rendered for application 316 may be a composition of the individual data manifests. In some cases, application 316 may include dependencies on external services, such as a MySQL database. Such dependencies may be captured as a dependency manifest. Data fed, processed, retained, or retrieved from these external services may also be annotated and automatically captured in application 316 data manifest 314.

[0051] Data annotating process 302 may, as described above, be utilized for decoupling data compliance from the business logic of application 316. For example, SDKs 320 and tooling 318 of data annotating process 302 may provide automated mechanisms for decoupling the configuration, observability, and enforcement of data compliance rules from the business logic of application 316. In some instances, application 316 may be a cloud/edge native application, which may be implemented as a set of workloads composing a service mesh. The decoupling of data compliance from the business logic may be especially relevant for applications of this type, as geographically dispersed and/or variably deployed workloads may implicate increased data compliance complexity.

[0052] Various possible embodiments for decoupling data compliance from the business logic of application 316 may be utilized. For instance, a sidecar model, where the services that implement the business logic of application 316 are deployed together with sidecar proxies associated to each of those services, may be utilized. The sidecar proxies may be utilized to enforce horizontal functions that are independent of the business logic, such as routing, security, load balancing, telemetry, retries, etc. As such, the sidecars may be well-positioned to decouple, observe, and control data compliance. For example, a combination of distributed data compliance controllers and sidecar proxies may be used to configure, observe, and enforce data compliance rules across different geographies, and distributed multi-cloud and edge infrastructures 334.

[0053] Instead of, or in addition to, using sidecars, various embodiments may use client libraries, daemons working in tandem with the application-specific services, or sandboxed programs in the OS kernel, e.g., using the Extended Berkeley Packet Filter (eBPF). Further embodiments may use an agentless approach or embed such functionality in Kubernetes itself. In any case, the functionality introduced herein may enable the portability and reuse of observability and enforcement of data compliance functions across not only different applications but also cloud and edge environments.

[0054] The above-described data annotating process 302 may yield a portable annotated application 316 that is geared with built-in annotations for different types of protected data. In addition, the yielded annotated application 316 may be structured to operate while remaining agnostic of any state, country, industry, organization-specific regulation and/or data policy requirements that a data team and/or application manager 304 might have. As a result, data annotating process 302 may be leveraged as a new model of building applications including DCaC by not only data teams and/or application managers 304, but also software as a service (SaaS) providers and others.

[0055] Data compliance as code process 328 may provide configuration, observability, and enforcement of data compliance rules. As described above, associations 326 between categories of annotated data in application 316 and specific categories of sensitive data may be instrumented prior to a deployment of application 316. The associations 326 may be used to control the processing and use of data during and after the deployment of application 316. More specifically, compliance engine 306 may utilize associations 326 together with current data compliance regulations governing data handling in each region where application 316 may be used, as well as a specific organization's compliance rules 308 for/while using application 316, to enforce compliance with them. Such controls may apply to data access requests, data storage and retention policies, data processing requirements, etc. of application 316 both at deploy and execution time, etc.

[0056] To this end, data compliance as code process 328 may include data compliance rule repository 322. Data compliance rule repository 322 may provide a repository of data compliance rules. A data compliance rule may include a data usage restriction such as an industry/company/legal standard, an industry/company/legal regulation, an industry/company/legal preference, an industry/company/legal policy, an industry/company/legal obligation (contractual or otherwise), a statute, a court ruling, and/or any other expression of an industry/company/legal data usage restriction. The data usage restriction may be geography-specific, jurisdiction-specific, company-specific, client-specific, industry-specific, data type-specific, etc.

[0057] For example, data compliance rules repository 322 may include a repository of industry rules 324 which may be applicable to the use of application 316 within a specific industry. For example, with respect to instances where application 316 is used by a healthcare provider, data compliance rule repository 322 may include industry rules 324 such as Health Insurance Portability and Accountability Act of 1996 (HIPAA) regulations applicable to handling of data in the healthcare industry. In other examples, data compliance rule repository 322 may include a repository of geographic rules 330 which may be applicable to the use of application 316 within a specific geographical location. For example, with respect to instances where application 316 is based in a member state of the E.U., data compliance rules

repository **322** may include geographic rules **330** such as the GDPR applicable to handling of data in the E.U.

[0058] The data compliance regulations included in data compliance rules repository **322** may be consumed by a data team and/or application manager **304** as a service (aaS). Data compliance rules repository **322** may support input, expression, collection, approval, visualization, and/or use of data compliance policies covering multiple categories of rules. The data compliance policies may include restrictions to the use of particular types of data. For example, data compliance rules repository **322** may store data compliance policies that are specific to an industry, those that may apply at a national, multi-national, federal, state, and industry levels, etc. For instance, an organization (e.g., a multi-national company) may leverage a data compliance rules repository **322** service of a data compliance as code process **328** and utilize the rules already available in data compliance rules repository **322**, which may cover regulations across several industries and countries out-of-the-box. An organization may select the target state, country or region, the industry if needed, and select the data compliance regulations that may be applicable at the organizational level (e.g., organization's compliance rules **308**).

[0059] Compliance engine **306** may offer APIs and a user-friendly user interface (UI) through which a data team and/or application manager **304** may select and define data compliance requirements. For instance, if application **316**, which handles Customer PII data, needs to be deployed in British Columbia, Canada, a data team and/or application manager **304** may simply select "Customer PII→Apply Local Rules" to constrain the processing, storage, retention, and access to Customer PII data according to the regulations in British Columbia as retrieved from data compliance rules repository **322**. To this end, compliance engine **306** may compute and handle the resulting constraints that apply to Customer PII data in British Columbia transparently to data teams and/or application managers **304**. More specifically, the set of data compliance constraints may be captured in a machine-readable format from data compliance regulation repository **322**, and therefore, used by compliance engine **306** programmatically.

[0060] In some examples, compliance engine **306** may be utilized as a pluggable module working in tandem with one or more workload engines **332**, such as Cisco Intersight, Hashicorp Nomad, or other automation tools offered by cloud and edge solution providers. As used herein, "workload engines" may include any software and/or hardware that are responsible for executing or deploying the application in a computing system and the term is not limited to a particular type of software and/or hardware managing any particular stage of the application lifecycle (e.g., deployment, execution, etc.). Workload engines **332** may manage the deployment and/or execution of application **316**, subject to the rules and constraints provided by compliance engine **306**.

[0061] In various embodiments, compliance engine **306** may operate either in a push or a pull model. For instance, in a pull model, a workload engine **332** may receive a request to deploy application **316** in a given region (e.g., a request from a site reliability engineering (SRE) and/or information technology (IT) team **310**). In such a case, workload engine **332** may issue a request to compliance engine **306**, to compute and return data compliance rules and constraints that must be applied for their specific deployment. Alternatively, in a push model, a data team and/or application manager **304** may select the compliance rules required and a declarative intent for application deployment may be issued from compliance engine **306** to one or more workload engines **332**. Such deployments may involve multi-cluster service meshes, which may run across multi-cloud and edge infrastructures **334**. In various embodiments, the above-described sidecar proxies in the service mesh may not only provide monitoring and observability of data compliance to compliance engine **306** but also may receive configuration and compliance updates in real-time **336**. In additional embodiments, the same functionality may be implemented utilizing client libraries, daemons, eBPF, an agentless approach, or Kubernetes itself. In addition, some embodiments may support the techniques described herein without utilizing a service mesh.

[0062] As noted above, the number and scope of laws and regulations regarding the processing, storage, and use of certain types of data are continually increasing across the world. For instance, the General Data Protection Regulation (GDPR) in Europe places strict requirements on how a user's personal data is collected and processed throughout its lifecycle. The GDPR has extraterritorial reach (e.g., it applies to entities outside of the EU) and includes broad definitions of personal data and equally broad definitions of processing activities. These and other requirements have spawned independent efforts across several countries to ensure that modern applications comply with specific data regulation at national, federal, or state level, particularly, those that are cloud delivered.

[0063] Legal uncertainty in the eye of accelerated digital transformation has resulted in concerns around cross-border data transfers and has given rise to "data localization" trends including its most severe form "data sovereignty". This dynamic is posing complex challenges to the organizations that process and manage the application data, since legal obligations and constraints often vary from country to country. The challenge is even greater since data compliance requirements are usually not limited to data sovereignty obligations.

[0064] Depending on the type of application, data compliance may involve the amalgamation of other requirements. For example, data compliance may involve industry-specific rules (e.g., complying with HIPAA obligations in the healthcare industry in the US), or organization-specific rules (e.g., on how to handle confidential information, i.e., data that might not be regulated by the law).

[0065] Similarities in terminology and semantic relatedness may exist between the various concepts that are subject to data protection and/or other data regulations across the world. However, determining the set of data compliance constraints that must be applied when application data is jointly subject to data sovereignty obligations, industry-specific regulation, and organization-specific rules remains a complex, manual, and labor-intensive task.

[0066] For instance, data teams in multinational corporations may work with legal counsel to constantly interpret any relevant regulation or gain specialized expertise on the resulting constraints. The collaborative interpretations may be relied upon to continuously keep such knowledge up-to-date as new regulations arise or existing ones are updated. This costly and labor-intensive approach to keeping abreast of regulatory landscapes often fails to produce a reliable regulatory knowledgebase as the volume of information to

be processed and diversity of expertise, knowledge, and organizational familiarity involved in building and modifying a holistic regulatory knowledgebase may be outside the capacity of an individual or team.

[0067] Further, organizations suffer from poor automation and lack the capacity to programmatically adapt to the different and ever-changing data compliance regulations. Indeed, extracting information programmatically from data compliance regulations and turning such knowledge into actionable constraints that workload engines can parse, and use in order to instrument the deployment of an application subject to data sovereignty obligations and other rules across different regions, is an art in itself. Multi-cloud and edge functionalities pose additional layers of complexity. Indeed, modern applications are typically composed of a set of services that may run in containerized environments across various public and/or private clouds and edge. While some techniques exist to classify and tag the data of an application, there are no existing mechanism available for mapping such classifications to specific data regulations, legal obligations and organization-specific rules. Moreover, there are no mechanisms available for turning such mappings into actionable outcomes.

Mapping of Application Data

[0068] The techniques herein introduce mechanisms for the automated mapping of application data to categories of protected data to ensure compliance with data compliance sovereignty laws, regulations, rules, policies, or the like. These techniques may provide an automated process of programmatically extracting data usage (e.g., sending, receiving, storing, processing, transforming, etc.) restrictions from various possible sources. Further, these techniques may be utilized to create mappings of application metadata (e.g., DCaC annotations) to those categories of protected data present in the restrictions in a manner that allows for automated compliance actions. That is, the mappings created utilizing these techniques may be consumable and actionable programmatically.

[0069] Illustratively, the techniques described herein may be performed by hardware, software, and/or firmware, such as in accordance with data mapping process **248**, which may include computer executable instructions executed by the processor **220** (or independent processor of interfaces **210**) to perform functions relating to the techniques described herein.

[0070] Specifically, according to various embodiments, a device may obtain an ontology derived from a data usage restriction document and indicative of a category of protected data. The device may obtain metadata indicative of a type of data handled by an application. The device may create a mapping between the type of data handled by the application and the category of protected indicated by the ontology. The device may generate, based on the mapping, a data compliance manifest that may be used by a workload engine to constrain use of the type of data during execution of the application or used to constrain use of the type of data during deployment of the application

[0071] Operationally, and according to various embodiments, the techniques described herein may be utilized to extract information programmatically from data usage restrictions and turn such knowledge into actionable constraints that workload engines and/or other systems can parse and use.

[0072] FIG. **4** illustrates an example architecture **400** to extract data usage restrictions and generate data compliance constraints, in accordance with one or more embodiments described herein. The architecture **400** may include a data compliance as code (DCaC) process **328**. As described above, DCaC process **328** may include data compliance rule repository **322** supporting compliance engine **306** in programmatically generating data compliance constraints to be applied to workload engines **332** in spinning up clusters supporting application workloads **408** across multi-cloud and/or edge infrastructures **334**.

[0073] As described above, compliance engine **306** may enable data team and/or application manager **304** to enter and/or manage data compliance requirements of an organization. For instance, if an application that handles Personally Identifiable Information (PII) needs to be deployed in Brazil, data team and/or application manager **304** may simply select "PII→Apply Local Regulation", and compliance engine **306** may compute and generate the resulting constraints that apply to PII in Brazil transparently to data team and/or application manager **304**. More specifically, the set of data compliance constraints may be captured in a machine-readable manner, and therefore, may be used by compliance engine **306** programmatically.

[0074] Data compliance rule repository **322** may provide the repository of rules, both in machine-readable and human-readable formats, which may be consumed by data team and/or application manager **304**, legal counselors, external systems as a service (aaS), etc. through DCaC process **328**. In addition to any rules already available in the data compliance rule repository **322**, which may cover regulations across several industries and/or geographic locations out-of-the-box, data compliance rule repository **322** may be populated with data compliance rules (e.g., data usage restrictions) programmatically extracted from various sources.

[0075] In various embodiments, a data usage restriction extraction service **402** may programmatically process data usage restrictions and/or other regulatory data from a variety of sources. For example, data usage restriction extraction service **402** may programmatically process any of a variety of expressions of data usage restrictions **406**. Expressions of data usage restrictions **406** may include records of data usage restrictions (e.g., rules, regulations, statutes, laws, treaties, legislative records, court decisions, guidelines, policies, preferences, best-practices, contracts, communications, obligations, etc.). The records may include audio records, visual records, text records, digital records, documents, files, posts, etc. In some examples, the records may include bodies of text such as words, sentences, paragraphs, pages, volumes, etc.

[0076] For example, expressions of data usage restrictions **406** may, in some examples, be industry-specific regulation documents (e.g., expression of data usage restriction **406***d*). An industry-specific regulation may include a regulation specific to industries such as third-party data suppliers, education institutions, financial services, medical/healthcare institutions, energy/utility providers, etc.

[0077] Expressions of data usage restrictions **406** may include data type-specific regulation documents (e.g., expression of data usage restriction **406***c*). For example, data type-specific regulation documents may include documentation of regulations covering the handling of personal data. A personal data regulation document may include documen-

tation of data usage restrictions (e.g., GDPR, CCPA, an omnibus personal data regulation document, etc.) for handling personally identifiable information (PII) and/or other personally distinguishable sensitive data.

[0078] In addition to being specific to an industry or data type, expressions of data usage restrictions **406** may be geolocation specific. For example, expressions of data usage restrictions **406** may be specific to an industry within a geographic location, specific to using a data type within a geographic location, etc.

[0079] Expressions of data usage restrictions **406** may, in some examples, be ontology-based online forms (e.g., expression of data usage restriction **406***b*). The online forms may capture industry specific, data type specific, etc. data usage restrictions.

[0080] Additionally, or alternatively, expressions of data usage restrictions **406** may, in some examples, be an ontology (e.g., expression of data usage restriction **406***a*) capturing industry specific, data type specific restrictions, etc. data usage restrictions.

[0081] Data usage restriction extraction service **402** may extract the corresponding data usage restrictions from expressions of data usage restrictions **406**. For example, data usage restriction extraction service **402** may utilize various data sources and programmatic techniques which may employ various statistical and/or non-statistical mechanisms of data extraction to extract the data usage restrictions.

[0082] Data usage restriction extraction service **402** may store the extracted data usage restrictions (e.g., as digests of extracted data usage restrictions) in a human-readable format and/or in a machine-readable format. The data usage restriction digests may be stored in data compliance rule repository **322**.

[0083] The data usage restriction digests stored in data compliance rule repository **322** may be utilized, in addition to other compliance rules that might be required at the organizational level (e.g., organization's compliance rules **308**), by workload engines **332** to deploy an application service. For example, compliance engine **306** may be utilized as a pluggable module working in tandem with workload engines **332**, such as Cisco Intersight, Hashicorp Nomad, or other automation tools offered by cloud and edge solution providers. These workload engines **332** may manage the deployment of the application service, subject to the rules and constraints provided by compliance engine **306**.

[0084] Compliance engine **306** might operate either in a push or pull model. For instance, in a pull model, workload engines **332** may receive a request to deploy an application in a given geographic region (e.g., from the SRE or IT team **310**). In such examples, workload engines **332** may issue a request to compliance engine **306** to compute and return the data compliance constraints that must be applied for that specific deployment.

[0085] The inference and response of such data compliance constraints in an actionable manner may be supported by a data compliance constraint inference service **404** of compliance engine **306**. Alternatively, in a push model a data team and/or application manager **304** may select the compliance rules required, and a declarative intent for deploying the application might be issued from compliance engine **306** to one or more of workload engines **332**. If the application deployment performed by workload engines **332** is successful, this may result in the spin up of several clusters (e.g., a set of K8s or K3s clusters) supporting the application workloads **408** across multi-cloud and edge infrastructures **334**.

[0086] FIG. **5** illustrates another example architecture **500** for extraction of data usage restrictions. For example, architecture **500** may be utilized to enable a programmatic extraction of data usage restrictions from various sources (e.g., expressions of data usage restrictions **406**) and their storage in data compliance rule repository **322**. The programmatic extraction of data usage restrictions may be performed utilizing statistical and/or non-statistical information extraction techniques.

[0087] In various embodiments, such as those employing statistical techniques to extract data usage restrictions, data usage restriction extraction service **402** may receive expressions of data usage restrictions **406** as inputs. The expressions of data usage restrictions **406** may include an industry-specific regulation document (e.g., expression of data usage restriction **406***d*) and/or a data type-specific regulation document (e.g., expression of data usage restriction **406***c*).

[0088] Data usage restriction extraction service **402** may process each of the expressions of data usage restrictions **406** independently. For each of the expressions of data usage restrictions **406**, the resulting output may include one or more digest of data usage restrictions (e.g., data usage restriction digest **502**) extracted from the expressions of data usage restrictions **406**. For example, the output for expressions of data usage restrictions **406** may include a data usage restriction digest in a human-readable format (e.g., human-readable data usage restriction digest **502***a*) and/or a data usage restriction digest in a machine-readable format (e.g., machine-readable data usage restriction digest **502***b*).

[0089] Each data usage restriction digest **502** may include descriptions of data usage restrictions such as those extracted from the expressions of data usage restrictions **406**. Data usage restriction digest **502** may be specific to and/or include data usage restrictions applicable to a specific geolocation and/or industry of an application service. Data usage restriction extraction service **402** may generate as many data usage restriction digest **502** as needed depending on the geolocation and/or industries of interest (e.g., the geographic location where an application service will operate and/or an industry associated with the application service).

[0090] Data usage restriction extraction service **402** may include information extraction service **508**. Information extraction service **508** may include language normalization module **510**. Language normalization module **510** may translate and bring different data usage restrictions from the expressions of data usage restrictions **406** to a common language. However, some embodiments may skip this step and/or exclude this module, instead relying on external means to achieve this translation. For instance, information extraction service **508** may be fed data usage restrictions in a common language (e.g., English).

[0091] Data usage restriction extraction service **402** may utilize natural language processing (NLP) techniques (e.g., executed by NLP module **512**). NLP techniques may be used to perform natural language processing and transformation of the expressions of data usage restrictions **406**. NLP techniques may be utilized to summarize and/or reformulate expressions of data usage restrictions **406** (e.g., one or more bodies of text therein) by leveraging one or more dictionaries **514**.

[0092] Dictionaries **514** may capture and/or define the reference terminology and/or domain-specific lexicon of the subject expressions of data usage restrictions **406**. Dictionaries may store relevant definitions on data regulation at state, province, country, multi-national, and industry-specific levels to access them from data structure **522**.

[0093] A combination of inference process **516** and disambiguation process **518** may utilize a set of domain-specific data compliance ontologies (e.g., data compliance domain ontologies **524**) and use techniques such as lexical matching, clustering, and semantic relatedness to identify the key concepts and relationships between such concepts in expressions of data usage restrictions **406**. The pool of data compliance domain ontologies **524** may include one reference ontology for data protection regulation of a particular geographic area (e.g., at national level, state level, city level, etc.), another ontology for data regulation in a particular industry (e.g., the finance industry), another for data regulation in another industry (e.g., the healthcare industry), etc. These reference ontologies may generalize and/or comprehensively capture the set of concepts and possible relationships that may apply among such concepts for each knowledge domain, independently of any specific data usage restriction in that domain.

[0094] Such ontologies may comprise multiple layers and taxonomies, such as a "data types" or "data categories" layer, a "data actors" layer, a "data resource" layer, a "data operations" layer, a "data permissions" layer, a "data location" layer, etc. For example, in one ontology a concept in the "data types" or "data categories" layer (e.g., "PII"), might be semantically related to a concept in the "data operations" layer (e.g., "Data Processing") through the property "CONSTRAINED TO". Likewise, the concept "Data Processing" in the "data operations" layer might be semantically related to a concept in the "data location" layer (e.g., "Country") through the property "WITHIN". Hence, once possible set of relationships among concepts captured by the ontology could be: (PII)→CONSTRAINED TO→(Data Processing)→WITHIN→(Country). Mapping function **520** may identify and/or map the concepts and relationships that would be obtained from a specific data usage restriction that is being processed by the information extraction service **508** to those that semantically match in the reference (domain-specific) ontology (e.g., the corresponding ontology from data compliance domain ontologies **524**).

[0095] Data usage restriction extraction service **402** may include data usage restriction generator **526**. Data usage restriction generator **526** may utilize the outputs (e.g., from mapping function **520**) of information extraction service **508** to generate an ontology specific to a data usage restriction being processed and/or to generate digests of data usage restrictions (e.g., data usage restriction digest **502**). For example, data usage restriction generator **526** may render both human-readable data usage restriction digest **502***a* and machine-readable data usage restriction digest **502***b*.

[0096] For example, the mapping yield by mapping function **520** may be subsequently used by restriction-specific ontology generator **528** to prune the reference ontology (e.g., the corresponding ontology from data compliance domain ontologies **524**) and produce an ontology that models the specific data usage restriction being processed by information extraction service **508**. Each of the ontologies generated by restriction-specific ontology generator **528** might be represented as a sub-graph of the reference (domain-specific) ontology (e.g., data compliance domain ontologies **524**). Hence, the newly generated ontology may capture the knowledge about the specific data usage restrictions contained in the expression of data usage restrictions **406** that is being processed.

[0097] The ontology computed by the restriction-specific ontology generator **528** may serve as an input to a restriction digest generator **530**. Restriction digest generator **530** may render both human-readable data usage restriction digest **502***a* and machine-readable data usage restriction digest **502***b* from the data usage restrictions extracted from expressions of data usage restrictions **406**. Data usage restriction digests **502** may include descriptions of and/or instructions for implementing their data usage restrictions as a constraint to a particular data type.

[0098] The techniques above described may utilize commonalities and semantic relatedness between the various concepts that are subject to data protection and regulations across different regions and industries. For example, GDPR is widely considered as a benchmark data handling regulation, so many of the existing data usage restrictions and upcoming ones may be inspired by a common set of principles within the GDPR. Indeed, many of them leverage the taxonomies, terminology, lexical semantics, and rules (or variants of them) present in GDPR. Similarly, industry-specific data usage restrictions also have clear commonalities across different regions. For instance, this data could be utilized as the basis to populate data structure **522** with reference terminology and data usage restriction data.

[0099] In addition, the solution space for data compliance constraints may be discrete, bounded, and sparse. For instance, a considerable number of data localization restrictions may be clustered into three groups regarding personal data: G1) those where data can be processed, stored, and accessed both within and outside the region of interest; G2) those where data must be stored within the region of interest, but they can be processed and accessed both within and outside the region of interest under certain data retention conditions; and/or G3) those where data must be processed, stored, and accessed exclusively within the region of interest. These characteristics may facilitate considerably the tasks of programmatically searching for and extracting data usage restrictions from expressions of data usage restrictions **406** and generating the corresponding data usage restriction digest **502**.

[0100] Data usage restriction extraction service **402** may include audit service **536**. Audit service **536** may operate in a semi-autonomous and/or autonomous mode. Audit service **536** may elicit and/or receive feedback from a user regarding the data usage restriction digest **502** generated from the extracted data usage restrictions.

[0101] For example, audit service **536** may send a human-readable data usage restriction digest **502***a* to a user for their review and/or provision of feedback. The feedback may include approval of data usage restriction digest **502**, rejection of data usage restriction digest **502**, modification of data usage restriction digest **502**, removal of data usage restriction digest **502**, further restriction of data usage restriction digest **502**, lessening restriction of data usage restriction digest **502**, commenting on data usage restriction digest **502**, etc.

[0102] The feedback may be utilized to revise the data usage restriction digest **502**. For example, the feedback may be utilized to steer specific modifications (e.g., more restric-

tive, less restrictive, more inclusive, less inclusive, etc.) to the data usage restriction digest **502** and/or to approve them before they are stored in data compliance rule repository **322**.

[0103] When operating in a semi-autonomous mode, audit service **536** may send at least the human-readable data usage restriction digest **502***a* to a user for the user to provide feedback at **532**. For example, audit service **536** may require a user such as an authorized legal counselor to approve the data usage restriction digest **502** generated by data usage restriction generator **526** and/or to supervise and/or request specific refinements to the obligations identified (e.g., via directed search and refinement module **534**).

[0104] As previously noted, data compliance rule repository **322** may provide a comprehensive catalog of data usage restriction digests for different geographical locations, capturing both personal data and industry specific obligations selectively employable as data compliance constraints to application services. Since the resulting data usage restriction digests stored in data compliance rule repository **322** may be uniformly formatted, they may be available and delivered in machine-readable format. Therefore, the resulting data usage restriction digests may be inserted into data compliance rule repository **322** and consumed and/or used programmatically by compliance engine **306** or third-party systems. For example, the resulting rule digests and/or their component data usage restrictions may be consumable in machine-readable format as a service from data compliance rule repository **322**. Thus, compliance engine **306** may pull a particular data usage restriction from data compliance rule repository **322** and utilize it to compute a data compliance constraint to be applied (e.g., via a workload engine **332**) to the operation of an application workload **408**.

[0105] In some examples, audit service **536** may automatically find data usage restriction digests already present in data compliance rule repository **322** from the same and/or similar geographies, industries, etc. and compare their classifications to the classifications of the data usage restriction digests **502**. For example, audit service **536** may determine that the same and/or similar data usage restriction digests already present in data compliance rule repository **322** are either classified differently from the machine-readable data usage restriction digest **502***b*, constituting a conflicting classification, and triggering a reconciliation process which may involve notifying a user of the conflict and requesting additional feedback. Alternatively, audit service **536** may determine that the same and/or similar data usage restriction digests already present in data compliance rule repository **322** are either classified similarly or identically to machine-readable data usage restriction digest **502***b*, constituting a confirmation of the classification which may be utilized as permission to save the machine-readable data usage restriction digest **502***b* to data compliance rule repository **322** and/or to make it consumable therefrom as a data compliance constraint.

[0106] Various embodiments may include additional instrumentation and/or enforcement refinements to the data usage restrictions identified by information extraction service **508** and/or data usage restriction generator **526**. For instance, audit service **536** may enable a user to edit the human-readable digest **502***a* and/or modify specific fields thereof. These modifications may be fed back (e.g., via directed search and refinement module **534**) to information extraction service **508** and/or data usage restriction generator **526**. The modifications may be utilized not only to regenerate data usage restriction digest **502** but also to train and/or refine information extraction service **508** in a semi-supervised fashion.

[0107] When operating in an autonomous mode, audit service **536** may operate as a combination of algorithmic and machine learning models that may fully automate the techniques of data usage restriction extraction service **402**. For instance, audit service **536** may be fully automated to programmatically and/or autonomously review the machine-readable data usage restriction digest **502***b*. For instance, audit service **536** may be trained to detect deviations from a set of bounded outcomes (e.g., digests cannot confidently fit into any of the groups G1, G2 and G3 described above) and/or to find replications or contradictions in the digests (e.g., similar to finding conflicting or confirming classifications among existing restriction digests in data compliance rule repository **322**, as described above). In such cases, a directed search and refinement module **534** may be triggered, enabling closed-loop operation and automated control between information extraction service **508**, data usage restriction generator **526**, and/or audit service **536** until a viable set (e.g., a non-conflicting and/or confirmed set) of data usage restrictions is found. Additional embodiments may rely on fully self-supervised training methods and backward feedback between modules until a viable set of obligations is found.

[0108] In contrast to the embodiments employing statistical techniques to extract data usage restrictions, various embodiments may alternatively, or additionally, employ deterministic (e.g., non-statistical) techniques to extract data usage restrictions from expressions of data usage restrictions **406**. For instance, various embodiments may utilize the data compliance domain ontologies **524** themselves (i.e., the set of reference ontologies) along with non-statistical techniques to simplify the data usage restriction extraction service **402**.

[0109] For instance, a user, such as an authorized legal counselor, may enter the data usage restrictions directly in the data usage restriction extraction service **402** by filling in an online form (e.g., expression of data usage restriction **406***b*). The form or forms that the user fills in may be internally structured according to the corresponding reference (domain-specific) ontology (e.g., an ontology from data compliance domain ontologies **524**). Hence, once a form is duly completed and submitted, restriction-specific ontology generator **528** may directly process the corresponding fields and data entered, and automatically generate a regulation-specific ontology without requiring any statistical inference. More specifically, data usage restriction extraction service **402** may determine the rules deterministically, instead of statistically. Information extraction service **508** may still be used to prefill the form and/or to provide suggestions to assist and simplify the data entry tasks involved in manually completing such forms.

[0110] In some embodiments, information extraction service **508** may be simplified and reduced to only NLP module **512**. NLP module **512** may utilize Pre-trained Language Models (PLMs), which may process the expressions of data usage restrictions (e.g., expression of data usage restriction **406***c* and expression of data usage restriction **406***d*) directly. In various embodiments, the outputs of this processing by NLP module **512** may be used to generate pre-filled forms (e.g., e.g., expression of data usage restriction **406***b*), which

may be ready to be completed and reviewed by users, such as authorized legal counselors.

[0111] Additional embodiments may enable a Regulatory Entity (RE) or an authorized organization to proactively provide a corresponding ontology (e.g., expression of data usage restriction **406***a*) jointly with the expression of data usage restriction (e.g., expression of data usage restriction **406***c*, expression of data usage restriction **406***d*, etc.) that it is to be utilized to extract the data usage restrictions from. Such ontology may capture, in machine-readable format, the data compliance obligations in a specific data usage restriction. In addition, the ontology may be cryptographically signed by the RE and may be directly processable by data usage restriction generator **526**.

[0112] Irrespective of whether the data usage restrictions were captured and conveyed via an ontology-based form (e.g., expression of data usage restriction **406***b*) or an ontology (e.g., expression of data usage restriction **406***a*) directly, data usage restriction extraction service **402** may be streamlined to solely data usage restriction generator **526** and a streamlined version of audit service **536**. Hence, data usage restriction extraction service **402** may be utilized for programmatic extraction of data usage restrictions and their corresponding storage in data compliance rule repository **322**, by combining both statistical and non-statistical information extraction techniques. Those skilled in the art would understand that the described use of forms and/or ontologies simply represent examples of possible embodiments. Alternative methods, including data models, files, or protocols could be used as well.

[0113] Once the data usage restrictions are extracted and stored in the data compliance rule repository **322** (e.g., in day −1), an organization may provide their organization's compliance rules **308** to compliance engine **306**. Then, data compliance constraint inference service **404** may automatically compute the corresponding data compliance constraints.

[0114] FIG. **6** illustrates an example architecture **600** to perform automated inference and generation of data compliance constraints, in accordance with one or more embodiments described herein. Compliance engine **306** may leverage the data compliance rule repository **322** to perform an automated inference of a data compliance constraint (e.g., in day+1).

[0115] At **610**, a data team and/or application manager **304** may submit a request for a deployment of an application service to compliance engine **306**. For example, once the data usage restrictions are generated and stored in data compliance rule repository **322** (e.g., in day −1), data teams and/or application managers **304** may provide their data compliance requirements to compliance engine **306** and let data compliance constraint inference service **404** automatically compute the corresponding data compliance constraints applicable to a deployment of an application service. For instance, a data team and/or application manager **304** may enter a set of data compliance requirements **602** for a given application deployment. The set of data compliance requirements **602** may include the geographical location **604** where the application needs to be deployed, the industry **608** in which the application will operate, and/or a set of rules that the organization might require (e.g., organizations' compliance rules **308**), including data processing, storage, retention, and access policies.

[0116] At **612**, the request may be processed by data compliance selection module **614** of compliance engine **306**. Data compliance selection module **614** may identify the geographical location and the corresponding industry of the requested application service deployment. Data compliance selection module **614** may use these identified characteristics as inputs to retrieve the data usage restrictions applicable to the requested deployment. For example, data compliance selection module **614** may identify machine-readable restriction digests and/or human-readable restriction digests stored within data compliance repository **322** that restrict data usage having characteristics that correspond to and/or match the identified characteristics of the requested application service.

[0117] At **615**, a user, such as a legal representative of an organization requesting a deployment of an application service, may revise the identified human-readable data usage restriction digests and/or machine-readable restriction digests. For example, the user may approve or edit the identified human-readable restriction digests and the system may automatically update the corresponding machine-readable restriction digests to further restrict the data usage restrictions (e.g., to reduce the exposure and potential risk of the organization to future fines in case of a regulation infringement). For instance, while the identified machine-readable restriction digests and/or human-readable restriction digests may stipulate that certain data types must be stored within the region of interest, but they can be processed and accessed both within and outside the region of interest under certain data retention conditions (G2), a user may restrict this even further, and recommend to process, store, retain, and access the data exclusively within the region of interest (G3).

[0118] At **616**, the revised and/or approved restriction digests **618** output from data compliance selection module **614** may be caused to be stored by data compliance selection module **614**. The revised and/or approved restriction digests **618** may be stored both in a machine-readable format (e.g., machine-readable restriction digest **618***b*) and/or in a human-readable format (e.g., human-readable restriction digest **618***a*). The revised and/or approved restriction digests **618** may be those directly obtained (e.g., unmodified) from data compliance rule repository **322** or may represent a more restrictive or less restrictive version of those obtained from data compliance rule repository **322**.

[0119] At **620**, a variety of inputs from architecture **600** may be provided to and/or utilized by data compliance inference service **404** to compute data compliance constraints. For example, at **620***a* the machine-readable data usage restrictions specified in machine-readable restriction digest **618***b* may be used by compute data compliance constraints module **622**, together with the organization's compliance rules **308** retrieved from the set of data compliance requirements **602**, at **620***b*, as well as the output of automated mapping function **630**, at **620***c*, to compute the specific set of data compliance constraints that should be applied to the application. For instance, for "Personal Data", such data compliance constraints may stipulate that the processing, storage, retention, and access of this category of data must be restricted to "within the country".

[0120] Automated mapping function **630** may utilize various inputs to generate the output utilized to compute the specific set of data compliance constraints that should be applied to the application. For example, automated mapping

function **630** may utilize ontology **632** as an input. Ontology may include a plurality of concepts (e.g., elements or components of a data usage restriction) and/or their relationships (e.g., how the concepts are related to each other within a data usage restriction). Ontology may be a data usage restriction-specific ontology **632**. For example, restriction-specific ontology **632** may be an ontology that was programmatically extracted (e.g., as illustrated in FIG. **5**) from a particular one or set of data usage restriction documents (e.g., expressions of data usage restrictions such as expression of data usage restriction **406***b*, expression of data usage restriction **406***c*, expression of data usage restriction **406***d*, etc.). Since the restriction-specific ontology **632** in such examples may be extracted from an expression of a particular one or set of data usage restriction(s) (e.g., a specific rule, regulation, statute, law, treaty, legislative record, court decision, guideline, policies, preference, best-practice, contract, communication, obligations, etc.), restriction-specific ontology **632** may be a data usage restriction-specific ontology that is specific to the data usage restriction described in the document whence it is extracted. In various embodiments, restriction-specific ontology **632** may be generated by restriction-specific ontology generator **528** within data usage restriction extraction service **402**.

[0121] Additionally, automated mapping function **630** may utilize a set of associations and/or protection bindings between metadata linked to the application (e.g., such metadata may be provided by application developers using DCaC or through third-party data tagging, data catalogs, and/or data inventories) and specific categories of sensitive data handled by the application as an input. These associations and/or protection bindings may be characterized in association tables **634**. Association tables **634** may be generated by and/or retrieved from compliance engine **306**. Association tables **634** may include a list of the associations, such as the example illustrated in Table 1.

[0122] Automated mapping function **630** may identify and/or extract the relevant categories of protected data according to organization-specific, industry-specific, and/or personal data obligations in a given region. Automated mapping function **630** may map such categories of protected data to the specific categories of sensitive data and the corresponding metadata associated with the application, as defined by data team and/or application manager **304** in association tables **634**.

[0123] In various embodiments, compliance engine **306** may have access to a catalog of candidate infrastructures that may be used for the application deployment. For example, the catalog of candidate infrastructures may include a pool of public cloud zones, a pool of public edge zones, a set of private cloud infrastructures, etc. Such information may be accessible as part of a "pull" request coming from a workload engine **332**, or it may be configured beforehand by an SRE/IT team **310**, or it may be available through other means. In any case, at **624**, the constraints computed by compute data compliance constraints module **622** may be mapped to the available infrastructure to filter and identify those infrastructure candidates that are compliant with the constraints, and subsequently generate a declarative intent to deploy the application.

[0124] At **626**, the declarative intent computed as **624** may be captured in the form of actionable data compliance manifest **636**. Actionable data compliance manifest **636** may include the constraints to the use of particular types of data handled by the application. As described, data compliance manifest **636** may be packaged as an actionable element **628** which may be sent to, obtained by, and/or programmatically utilized by workload engines **332** and/or other systems both at deploy and execution/run time. For example, workload engines **332** may utilize actionable element **628** to configure the deployment and/or execution of a corresponding workload on compliant infrastructure.

[0125] FIG. **7** illustrates an example architecture **700** to perform an automated mapping of application metadata to a category of protected data, in accordance with one or more embodiments described herein. Compute data compliance constraints module **622** and/or automated mapping function **630** may be collectively utilized to perform the automated mapping within compliance engine **306**. The automated mapping process may be utilized to generate a set of data compliance constraints (e.g., captured as an actionable data compliance manifest **636**).

[0126] Application developers may create application **316** and/or its corresponding data manifest **314**. Data manifest **314** may be available for data teams and/or application managers through compliance engine **306**.

[0127] Data manifest **314** may include metadata **702** associated with the different categories of data handled by application **316**. Metadata **702** may include data tags associated with particular data handled by application **316**, which may be obtained as part of DCaC and/or imported from a data catalog. Such data tags may be utilized in application **316** and/or may be described in the data manifest **314** associated to application **316**. For example, metadata **702** may be in the form of data tags to application data indicating protected-types and/or custom-types.

[0128] Data teams and/or application managers may validate, update, and/or create associations or protection bindings between the metadata **702** contained in the data manifest **314** and the categories of sensitive data handled by application **316** and store such associations in association table **634**. Data teams and/or application managers may also select the required organization's compliance rules **308**. In addition, data teams and/or application managers may select the industry **608** in which application **316** will operate as well as the geographical location **604** where it needs to be deployed.

[0129] Based on these inputs, compliance engine **306** may identify specific data compliance rules **704** that are to be applied at a national (e.g., data compliance rule **704***b*), industry (e.g., data compliance rule **704***c*), and/or organizational level (e.g., data compliance rule **704***a*) for the specified geographic location **604** and/or application deployment. Compliance engine **306** may also compute mapping **706**. Computing the mapping may include determining a mapping for a specific data usage restriction to be applied to the application **316** in light of the inputs. For example, computing the mapping may include determining a mapping for an application handling data of users in the Kingdom of Saudi Arabia to the applicable data compliance rules **704**.

[0130] Compliance engine **306** may infer how different categories of data indicated in metadata **702** of application **316** map onto categories of protected data **708**. Categories of protected data **708** may cover both regulated data (e.g., PII data) and non-regulated data (e.g., confidential data).

[0131] Based on the mappings, compliance engine **306** may generate a set of data compliance constraints. The set of data compliance constraints may be an actionable data

compliance manifest **636** configured to be programmatically utilized by, e.g., a workload engine and/or by an SRE/IT lead to manage the configuration and deployment of application **316** according to the mappings. For example, compliant infrastructure and/or resources may be selected, and the application may be deployed and configured to execute in a compliant manner.

[0132] FIGS. **8**A-**8**D illustrate an example of data manifest **314** and data compliance rules **704** to perform an automated mapping of application metadata to a category of protected data, in accordance with one or more embodiments described herein. For example, data manifest **314** and data compliance rules **704** at an organizational level (e.g., data compliance rule **704***a*), at a national level (e.g., data compliance rule **704***b*), and/or at an industry level (e.g., data compliance rule **704***c*) may be utilized as inputs to the automated mapping process (e.g., automated mapping process **710** of FIG. **7**).

[0133] As previously described, metadata may be used to tag sensitive data in the application and described in the data manifest **314** associated with the application. These tags may then be tied to compliance rules **704** that are derived from the various legal, industry, and/or organizational requirements for the data. The compliance engine **306** may then generate the set of data compliance constraints expressed as an actionable data compliance manifest **636**.

[0134] FIGS. **9**A-**9**B illustrate an example architecture **900** to perform an automated mapping of application metadata to categories of protected data indicated in a restriction-specific ontology **632**, in accordance with one or more embodiments described herein. As described above, a programmatic extraction of data usage restrictions from various sources (e.g., expressions of data usage restrictions **406**) may be performed. For example, an ontology (e.g., expression of data usage restriction **406***a*), an ontology-based form (e.g., expression of data usage restriction **406***b*), a data type-specific regulation document (e.g., expression of data usage restriction **406***c*), and/or an industry-specific regulation document (e.g., expression of data usage restriction **406***d*) may be processed and/or utilized as inputs to data usage restriction extraction service **402** to generate a restriction-specific ontology **632**. For example, restriction-specific ontology generator **528** may extract a restriction-specific ontology **632** that is specific to the data usage restriction expressed in the expression of data usage restriction utilized as an input. Restriction-specific ontology **632** may include a plurality of concepts and their relations extracted from a data usage restriction document, an ontology, an ontology-based form, etc. Restriction-specific ontology **632** may be utilized by automated mapping function **630** in generating the mapping between application metadata and categories of protected data indicated in a restriction-specific ontology **632**.

[0135] In addition, automated mapping function **630** may utilize metadata **702** in generating the mapping between application metadata and categories of protected data indicated in a restriction-specific ontology **632**. For example, metadata **702** such as protected data types and/or custom data types obtained from DCaC sources such as DCaC annotations in the application code may be utilized as well as data catalogs and/or other tagged data obtained from other sources may be captured in association tables **634**. As noted above, association tables **634** may include default associations, custom associations, third party associations, etc. between metadata, categories of sensitive data, scopes, encryption/decryption key IDs, etc. The contents of association tables **634** may be assigned, created, supplemented, modified, approved, etc. by a data team and/or application manager **304**.

[0136] Automated mapping function **630** may operate on the restriction-specific ontology **632** and/or the association table **634** to perform an automated mapping of application metadata from association table **634** to categories of protected data **708** indicated in a restriction-specific ontology **632**. The mapping may be performed by applying a mapping function **902** to the inputs. For example, the mapping may include mapping metadata-indicated data type such as "Customer PIT" to a "personal data" category of protected data **708** indicated in a restriction-specific ontology **632** and/or a "confidential" metadata-indicated data type to a "confidential data" category of protected data **708** indicated in a restriction-specific ontology **632**.

[0137] The output of automated mapping function **630** may be utilized, along with machine-readable data usage restrictions specified in a machine-readable restriction digest **618***b* and/or with the organization's compliance rules **308** retrieved from the set of data compliance requirements, by compute data compliance constraints module **622** to compute the specific set of data compliance constraints that should be applied for the application. The compliance constraints may be computed for both non-legally regulated data **906** of the application and for legally regulated data **908** of the application.

[0138] The computed compliance constraints may be captured as in data compliance manifest **636**. Data compliance manifest **636** may be packaged as an actionable element **628** which may be sent to, obtained by, and/or programmatically utilized by workload engines and/or other systems. For example, workload engines may utilize actionable element **628** to configure the deployment and/or execution of a corresponding workload on compliant infrastructure.

[0139] FIG. **10** illustrates an example simplified procedure (e.g., a method) for mapping application data, in accordance with one or more embodiments described herein. For example, a non-generic, specifically configured device (e.g., device **200**), may perform procedure **1000** by executing stored instructions (e.g., data mapping process **248**).

[0140] The procedure **1000** may start at step **1005**, and continues to step **1010**, where, as described in greater detail above, a device may obtain an ontology derived from a data usage restriction document, which may include an ontology, an ontology-based form, etc. The data usage restriction document may be a law, regulation, or industry rule, etc.

[0141] The ontology may comprise a plurality of concepts and their relations extracted from the data usage restriction document. For example, the ontology may include indications of a category of protected data in the data usage restriction document. The ontology may be based at least in part on data supplied by a user via a user interface form (e.g., an ontology-based form).

[0142] At step **1015**, as detailed above, a device may obtain metadata indicative of a type of data handled by an application. In various embodiments, the metadata may include data tags associated with application data and/or their associations captured in an association table associated with the application.

[0143] At step **1020**, as detailed above, the device may create a mapping between the type of data handled by the application and the category of protected data indicated by

the ontology. Creating the mapping may include matching the category of protected data from the ontology to the type of data handled by the application.

[0144] At step **1025**, as detailed above, the device may generate, based on the mapping, a data compliance manifest. The data compliance manifest may include a data compliance constraint that is based on the ontology. In various embodiments, the data compliance manifest constrains the use of the type of data handled by the application by indicating how that type of data can be utilized. For example, the data compliance manifest may constrain the use of the type of data by the application by indicating one or more geolocations in which the type of data can be stored or retained by the application.

[0145] The data compliance manifest may be utilized by a workload engine to constrain the use of the type of data handled by the application during execution of the application or used to constrain use of the type of data during deployment of the application. The workload engine may configure a workload of the application that uses the type of data at infrastructure that satisfies the data compliance constraint. The mapping and/or the data compliance manifest may be updated, based on a change in the data usage restriction document.

[0146] Procedure **1000** then ends at step **1030**.

[0147] It should be noted that while certain steps within procedure **1000** may be optional as described above, the steps shown in FIG. **10** are merely examples for illustration, and certain other steps may be included or excluded as desired. Further, while a particular order of the steps is shown, this ordering is merely illustrative, and any suitable arrangement of the steps may be utilized without departing from the scope of the embodiments herein.

[0148] In the absence of these techniques, organizations suffer from poor automation and lack the capacity to programmatically adapt to the different and ever-changing data compliance regulations. Indeed, outside of these techniques, organizations have no means of extracting information programmatically from data usage restrictions and turning such knowledge into actionable data compliance constraints that workload engines and other systems can parse, and use, in order to instrument the deployment of an application subject to data sovereignty obligations and other rules across different regions. Multi-cloud and edge functionalities pose additional layers of complexity.

[0149] The techniques described herein, therefore, may provide these functionalities by introducing an automated mechanism for programmatically extracting data usage restrictions, classifying and/or tagging data handled by an application, and mapping the classifications to the extracted data usage restrictions. Moreover, the techniques make such mappings available in such a way that may be consumable and actionable by third-party processes. As such, the techniques introduce a new approach that enables the automated mapping between the categories of data handled by an application, and categories of data subject to various data compliance rules and regulations in a consumable and programmatically actionable manner.

[0150] While there have been shown and described illustrative embodiments to map application data, it is to be understood that various other adaptations and modifications may be made within the spirit and scope of the embodiments herein. For example, while certain embodiments are described herein with respect to using the techniques herein for certain purposes, the techniques herein may be applicable to any number of other use cases, as well.

[0151] The foregoing description has been directed to specific embodiments. It will be apparent, however, that other variations and modifications may be made to the described embodiments, with the attainment of some or all of their advantages. For instance, it is expressly contemplated that the components and/or elements described herein can be implemented as software being stored on a tangible (non-transitory) computer-readable medium (e.g., disks/CDs/RAM/EEPROM/etc.) having program instructions executing on a computer, hardware, firmware, or a combination thereof. Accordingly, this description is to be taken only by way of example and not to otherwise limit the scope of the embodiments herein. Therefore, it is the object of the appended claims to cover all such variations and modifications as come within the true spirit and scope of the embodiments herein.

What is claimed is:

1. A method comprising:

obtaining, by a device, an ontology derived from a data usage restriction document and indicative of a category of protected data;

obtaining, by the device, metadata indicative of a type of data handled by an application;

creating, by the device, a mapping between the type of data handled by the application and the category of protected data indicated by the ontology; and

generating, by the device and based on the mapping, a data compliance manifest used by a workload engine to constrain use of the type of data during execution of the application or used to constrain use of the type of data during deployment of the application.

2. The method as in claim **1**, wherein the data compliance manifest includes a data compliance constraint that is based on the ontology.

3. The method as in claim **2**, wherein the workload engine configures a workload of the application that uses the type of data at infrastructure that satisfies the data compliance constraint.

4. The method as in claim **1**, wherein the ontology comprises a plurality of concepts and their relations extracted from the data usage restriction document.

5. The method as in claim **1**, wherein the ontology is based in part on data supplied by a user via a user interface form.

6. The method as in claim **1**, wherein the data compliance manifest constrains the use of the type of data by the application by indicating one or more geolocations in which the type of data can be stored or retained by the application.

7. The method as in claim **1**, wherein creating the mapping comprises:

matching the category of protected data from the ontology to the type of data.

8. The method as in claim **1**, wherein the data compliance manifest constrains the use of the type of data by the application by indicating how the type of data can be utilized.

9. The method as in claim **1**, further comprising:

updating the mapping and the data compliance manifest, based on a change in the data usage restriction document.

10. The method as in claim **1**, wherein the data usage restriction document comprises a law, regulation, or industry rule.

11. An apparatus, comprising:

one or more network interfaces;

a processor coupled to the one or more network interfaces and configured to execute one or more processes; and

a memory configured to store a process that is executable by the processor, the process when executed configured to:

obtain an ontology derived from a data usage restriction document and indicative of a category of protected data;

obtain metadata indicative of a type of data handled by an application;

create a mapping between the type of data handled by the application and the category of protected data indicated by the ontology; and

generate, based on the mapping, a data compliance manifest used by a workload engine to constrain use of the type of data during execution of the application or used to constrain use of the type of data during deployment of the application.

12. The apparatus as in claim **11**, wherein the data compliance manifest includes a data compliance constraint that is based on the ontology.

13. The apparatus as in claim **12**, wherein the workload engine configures a workload of the application that uses the type of data at infrastructure that satisfies the data compliance constraint.

14. The apparatus as in claim **11**, wherein the ontology comprises a plurality of concepts and their relations extracted from the data usage restriction document.

15. The apparatus as in claim **11**, wherein the ontology is based in part on data supplied by a user via a user interface form.

16. The apparatus as in claim **11**, wherein the data compliance manifest constrains the use of the type of data by the application by indicating one or more geolocations in which the type of data can be stored or retained by the application.

17. The apparatus as in claim **11**, wherein to create the mapping further comprises to:

match the category of protected data from the ontology to the type of data.

18. The apparatus as in claim **11** wherein the data compliance manifest constrains the use of the type of data by the application by indicating how the type of data can be utilized.

19. The apparatus as in claim **11**, wherein the process when executed is further configured to:

update the mapping and the data compliance manifest, based on a change in the data usage restriction document.

20. A tangible, non-transitory, computer-readable medium storing program instructions that cause a device to execute a process comprising:

obtaining, by the device, an ontology derived from a data usage restriction document and indicative of a category of protected data;

obtaining, by the device, metadata indicative of a type of data handled by an application;

creating, by the device, a mapping between the type of data handled by the application and the category of protected data indicated by the ontology; and

generating, by the device and based on the mapping a data compliance manifest used by a workload engine to constrain use of the type of data during execution of the application or used to constrain use of the type of data during deployment of the application.

* * * * *