(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 March 2022 (24.03.2022)

WIPO | PCT

(10) International Publication Number
WO 2022/061270 A1

(51) International Patent Classification:
*G06F 21/55* (2013.01) *G06F 21/62* (2013.01)
*G06F 21/57* (2013.01) *G06N 20/00* (2019.01)

(21) International Application Number:
PCT/US2021/051217

(22) International Filing Date:
21 September 2021 (21.09.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/080,917 21 September 2020 (21.09.2020) US

(71) Applicant: **ONETRUST, LLC** [US/US]; 1200 Abernathy Road, Atlanta, Georgia 30328 (US).

(72) Inventors: **BRANNON, Jonathan Blake**; c/o OneTrust, LLC, 1200 Abernathy Road, Atlanta, Georgia 30328 (US). **WHITNEY, Patrick**; c/o OneTrust, LLC, 1200 Abernathy Road, Atlanta, Georgia 30328 (US).

(74) Agent: **HAGGERTY, Christopher S.**; Brient IP Law, LLC, 1175 Grimes Bridge Road, Suite 100, Roswell, Georgia 30075 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW,

(54) Title: DATA PROCESSING SYSTEMS AND METHODS FOR AUTOMATICALLY DETECTING TARGET DATA TRANSFERS AND TARGET DATA PROCESSING

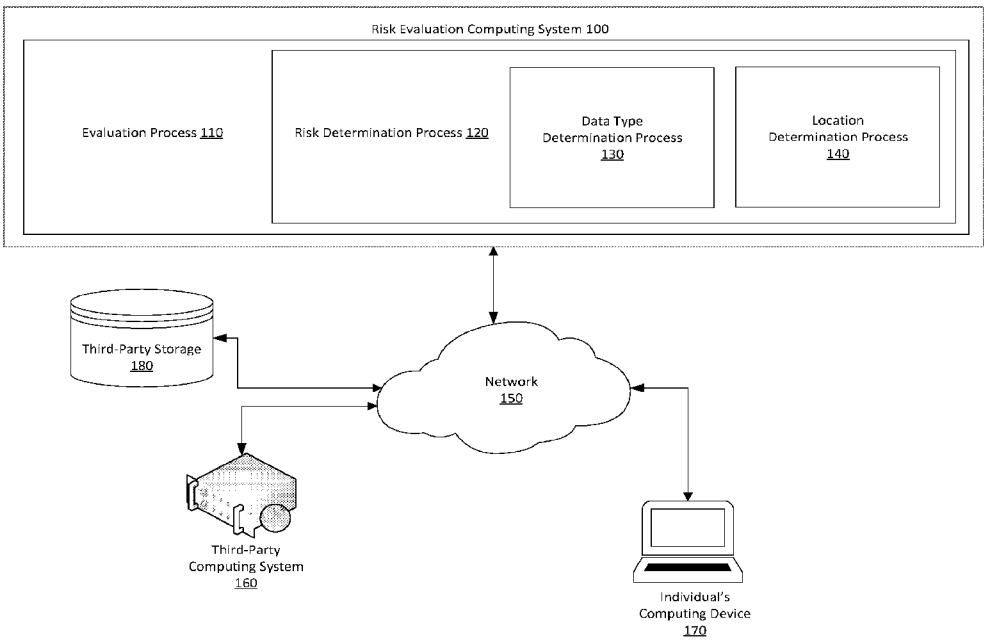


FIG. 1

(57) Abstract: Aspects of the present disclosure provide methods, apparatuses, systems, computing devices, computing entities, and/or the like for protection of system software, or data from destruction, unauthorized modification, and/or unauthorized disclosure securing by, for example, detecting the transfer and/or processing of target data. Accordingly, a method is provided that involves: scanning a software application to identify functionality configured for processing target data; identifying fields associated with the functionality; identifying metadata associated with a field; generating, from the metadata, an identification of a type of data associated with the field; determining a location based on the processing of the target data by the functionality; determining a risk associated with the functionality processing the target data based on the location and the type of data; determining that the risk satisfies a threshold level of risk; and

WO 2022/061270 A1

*[Continued on next page]*

SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

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

**Declarations under Rule 4.17:**

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

**Published:**

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

# DATA PROCESSING SYSTEMS AND METHODS FOR AUTOMATICALLY DETECTING TARGET DATA TRANSFERS AND TARGET DATA PROCESSING

## CROSS-REFERENCE TO RELATED APPLICATIONS

**[0001]** This application claims the benefit of U.S. Provisional Patent Application Serial No. 63/080,917, filed September 21, 2020, which is hereby incorporated herein by reference in its entirety.

## TECHNICAL FIELD

**[0002]** The present disclosure involves computer-implemented systems and processes for protection of system software, or data from destruction, unauthorized modification, and/or unauthorized disclosure securing by, for example, detecting the transfer and/or processing of personal data occurring via a website, mobile application, and/or the like.

## BACKGROUND

**[0003]** Software applications such as websites, mobile applications, and/or the like often collect and/or use personal data such as users' home addresses, social security number, and/or credit card numbers. Such software applications can include functionality that requires or involves transferring the personal data to different entities that are located in various locations. Therefore, an organization's use of software applications such as websites, mobile applications, and/or the like can often expose the organization to significant risk of experiencing a data privacy incident and to having to comply with various personal data processing requirements of different jurisdictions.

**[0004]** In many instances, changes in the software application can introduce vulnerabilities or other risks that are difficult to detect due to their implementation in program code. These vulnerabilities or other risks can result from, for example, updates made to the program code to introduce or change the collection and/or use of personal data. In one example, a software engineer who is maintaining a website for an organization may not appreciate that modifying the website to collect certain personal data can result in a significant risk to the organization in potentially experiencing a data privacy incident, as well as in not being in compliance with one or more data privacy standards. Furthermore, such modifications are often made without the knowledge of

appropriate personal within the organization such a privacy officer. Therefore, a need exists for systems and methods that facilitate the ability of an organization to operate and manage software applications such as websites, mobile applications, and/or the like that use personal data by identifying the risk of the organization experiencing a data privacy incident, as well as complying with personal data processing requirements associated with various data privacy standards in different jurisdictions, and by taking one or more actions to mitigate that risk.

## SUMMARY

**[0005]** In general, aspects of the present invention provide methods, apparatus, systems, computing devices, computing entities, and/or the like for protection of system software, or data from destruction, unauthorized modification, and/or unauthorized disclosure securing by, for example, detecting the transfer and/or processing of target data. In accordance with one aspect, a method is provided. According to particular aspects, the method involves: scanning, by computing hardware, a software application to identify functionality configured for processing target data; identifying, by the computing hardware, a plurality of fields associated with the functionality; identifying, by the computing hardware, metadata associated with a field from the plurality of fields; generating, by the computing hardware and from the metadata, an identification of a type of data associated with the field using at least one of a rules-based model or a machine-learning model; determining, by the computing hardware, a location based on the processing of the target data by the functionality; determining, by the computing hardware, a risk associated with the functionality processing the target data based on the location and the type of data for the field; determining, by the computing hardware, that the risk satisfies a threshold level of risk; and responsive to determining that the risk satisfies the threshold level of risk, causing, by the computing hardware, an action to be performed to mitigate the risk.

**[0006]** For example, the software application may comprise a website and the functionality may comprise a webform found on the website in which at least one of the plurality of fields is used on the webform to collect the target data. In another example, the software application may comprise a mobile application and the functionality may comprise a graphical user interface provided through the mobile application in which at least one of the plurality of fields is used on the graphical user interface to collect the target data.

**[0007]** Accordingly to particular aspects, the action may comprise at least one of generating an electronic communication sent to personnel identifying the functionality and the risk, causing the software application to become unavailable, or disabling the functionality in the software application. In addition, the risk may comprise at least one of a risk of experiencing a data privacy incident due to the functionality processing the target data and a risk of being noncompliant with a data privacy standard due to the functionality processing the target data.

**[0008]** According to some aspects, the method may further comprise determining, by the computing hardware, a vendor associated with the functionality based on metadata associated with the functionality, wherein the location is a jurisdiction in which the vendor processes data and processing of the target data by the functionality involves transferring the target data to the location. Further, according to some aspects, determining the risk associated with the functionality processing the target data based on the type of data for the field and the location involves using at least one of a second rules-based model or a second machine learning model to generate the risk, wherein the risk represents a likelihood of experiencing at least one of a data privacy incident due to the functionality processing the target data or being noncompliant with a data privacy standard due to the functionality processing the target data.

**[0009]** In accordance with another aspect, a system comprising a non-transitory computer-readable medium storing instructions and a processing device communicatively coupled to the non-transitory computer-readable medium. Accordingly, the processing device is configured to execute the instructions and thereby perform operations comprising: scanning a software application to identify functionality configured for processing target data; identifying metadata associated with the functionality; processing the metadata using at least one of a rules-based model or a machine learning model to generate an identification of a type of data associated with the functionality; determining a location based on the processing of the target data by the functionality; determining a risk associated with the functionality processing the target data based on the type of data and the location; determining the risk satisfies a threshold level of risk; and responsive to determining the risk satisfies the threshold level of risk, causing an action to be performed to mitigate the risk.

**[0010]** According to particular aspects, the action may comprise at least one of generating an electronic communication sent to personnel identifying the functionality and the risk, causing the software application to become unavailable, or disabling the functionality in the software

application. According to particular aspects, the risk may comprise at least one of a risk of experiencing a data privacy incident due to the functionality processing the target data or a risk of being noncompliant with a data privacy standard due to the functionality processing the target data.

**[0011]** According to some aspects, the operations may further comprise determining a vendor associated with the functionality based on the metadata, wherein the location is a jurisdiction in which the vendor is located and processing of the target data by the functionality involves transferring the target data to the location. In addition, according to some aspects, determining the risk associated with the functionality processing the target data based on the type of data and the location involves processing the type of data and the location using at least one of a second rules-based model or a second machine learning model to generate the risk representing a likelihood of experiencing at least one of a data privacy incident due to the functionality processing the target data or being noncompliant with a data privacy standard due to the functionality processing the target data.

**[0012]** In accordance with yet another aspect, a non-transitory computer-readable medium is provided. The non-transitory computer-readable medium having program code that is stored thereon, the program code executable by one or more processing devices for performing operations comprising: scanning a software application to identify functionality configured for processing target data; identifying metadata associated with the functionality; identifying a type of data associated with the functionality based on the metadata; determining a location based on the processing of the target data by the functionality; processing the type of data and the location using at least one of a rules-based model or a machine learning model to generate a risk representing a likelihood of experiencing a data incident due to the functionality processing the target data; determining the risk satisfies a threshold level of risk; and responsive to determining the risk satisfies the threshold level of risk, causing an action to be performed to mitigate the risk.

**[0013]** For example, the software application may comprise a website and the functionality may comprise a webform found on the website in which the functionality is used on the webform to collect the target data. For example, the software application may comprise a mobile application and the functionality may comprise a graphical user interface provided through the mobile application in which a field is used on the graphical user interface to collect the target data. For example, the data incident may comprise at least one of a data privacy incident due to the

functionality processing the target data or a risk of being noncompliant with a data privacy standard due to the functionality processing the target data.

**[0014]** According to particular aspects, the action may comprise at least one of generating an electronic communication sent to personnel identifying the functionality and the risk, causing the software application to become unavailable, or disabling the functionality in the software application. According to particular aspects, the risk may comprise at least one of a risk of experiencing a data privacy incident due to the functionality processing the target data or a risk of being noncompliant with a data privacy standard due to the functionality processing the target data.

**[0015]** According to some aspects, identifying the type of data associated with the functionality based on the metadata involves processing the metadata using at least one of a second rules-based model or a second machine learning model to generate an identification of the type of data associated with the functionality. According to some aspects, the operations may further comprise determining a vendor associated with the functionality based on the metadata, wherein the location is a jurisdiction in which the vendor is located and processing of the target data by the functionality involves transferring the target data to the location.

## BRIEF DESCRIPTION OF THE DRAWINGS

**[0016]** In the course of this description, reference will be made to the accompanying drawings, which are not necessarily drawn to scale, and wherein:

**[0017]** FIG. 1 depicts an example of a computing environment for evaluating a risk associated with target data collection of a website in accordance with various aspects of the present disclosure;

**[0018]** FIG. 2 is a flowchart of a process for evaluating a risk associated with target data collection of a website in accordance with various aspects of the present disclosure;

**[0019]** FIG. 3 is a flowchart of a process for determining a risk associated with a webform in accordance with various aspects of the present disclosure;

**[0020]** FIG. 4 is a flowchart of a process for identifying types of data collected via a webform in accordance with various aspects of the present disclosure;

**[0021]** FIG. 5 illustrates an exemplary graphical user interface containing HTML data that may be processed according to various aspects of the present disclosure;

**[0022]** FIG. 6 is a flowchart of a process for identifying locations associated with transfers of target data in accordance with various aspects of the present disclosure;

**[0023]** FIG. 7 is a block diagram illustrating an exemplary system architecture that may be used in accordance with various aspects of the present disclosure; and

**[0024]** FIG. 8 is a schematic diagram of a computing entity that may be used in accordance with various aspects of the present disclosure.

## DETAILED DESCRIPTION

**[0025]** Various aspects for practicing the technologies disclosed herein are described more fully hereinafter with reference to the accompanying drawings, in which some, but not all aspects of the technologies disclosed are shown. Indeed, various aspects disclosed herein are provided so that this disclosure will satisfy applicable legal requirements and should not be construed as limiting or precluding other aspects applying the teachings and concepts disclosed herein. Like numbers in the drawings refer to like elements throughout.

Technical Contributions of Various Aspects

**[0026]** Entities that process (e.g., collects, receives, transmits, stores, processes, shared, and/or the like) sensitive and/or personal data (personal data) associated with particular individuals, such as personally identifiable information (PII) data, are exposed to, at some level, a risk of experiencing some type of data privacy incident involving the personal data such as a data breach leading to the unauthorized access of the personal data. Furthermore, the entity may be subject to various laws and regulations regarding the processing of such personal data.

**[0027]** Software applications such as websites, mobile applications, and/or the like often collect and/or use personal data. In addition, these software applications often transfer the personal data to entities (e.g., vendors, services, systems, and/or the like) located in various locations. Further, users who are visiting and/or making use of these software applications may be associated with (residing in or currently located in) various locations. As a result, an organization's use of software applications such as websites, mobile applications, and/or the like can often expose the organization to significant risk of experiencing a data privacy incident due to transferring and/or processing personal data in certain locations, as well as having to comply with various personal data processing requirements of different jurisdictions.

**[0028]** For instance, personnel of an organization who are responsible for constructing, maintaining, managing, and/or the like of software applications such as websites, mobile

applications, and/or the like may not appreciate risks that result from introducing the collection and/or use of personal data into the software applications. For example, a software engineer who revises a website to collect and transfer personal data to a particular system found in a location (e.g., jurisdiction) known for experiencing a high number of data breaches may not appreciate that the revision to the website can result in a significant increase in risk to the organization in potentially experiencing a data privacy incident. Therefore, the revision to the website can lead to the introduction of operational and/or system vulnerabilities for the organization. Furthermore, these operational and/or system vulnerabilities can be highly technical in nature since the vulnerabilities occur as a result of modifying program code that implements a website or other software application. Consequently, such vulnerabilities may be difficult or infeasible to detect by appropriate personnel within the organization, such as a privacy officer.

**[0029]** Accordingly, various aspects of the disclosure address several of the technical challenges associated with the use (e.g., collection, storage, transfer, and/or the like) of target data, such as personal data, in software applications such as websites, mobile applications, and/or the like by providing a risk evaluation computing system configured to detect the use of target data associated with these software applications and evaluate the risk resulting from the use of the target data. For instance, the risk evaluation computing system can scan a software application (e.g., one or more webpages of a website) to identify the collection and/or transfer of target data by the software application. The risk evaluation computing system can identify the types of target data that is being collected and/or transferred by the software application, as well as identifying entities such as vendors, services, systems, and/or the like that exchange (e.g., transfer) target data with the software application. As discussed further herein, according to some aspects, the risk evaluation computing system may use a machine-learning model to identify the types of target data being used by the software application. Further, according to some aspects, the risk evaluation computing system identifies locations associated with the entities such as, for example, locations of the entities' servers used in collecting, transferring, processing, storing, and/or the like the target data exchanged with the software application.

**[0030]** In some aspects, the risk evaluation computing system can also determine a risk associated with the use of the target data by the software application. For instance, the risk evaluation computing system can use a rules-based model, a machine-learning model, or some combination thereof to determine the risk. The risk evaluation computing system may compare

the current use of the target data by the software application with a previous use to determine whether a change has occurred with respect to the use of the target data by the software application. Accordingly, any detected change in use may be reflected in the determined risk.

**[0031]** Various aspects of the risk evaluation computing system may perform one or more operations in response to the risk satisfying a threshold (e.g., equaling or exceeding a predefined level of risk). For instance, the risk evaluation computing system may generate and send some form of communication to one or more individuals to notify the individuals of the risk involved with the software application processing the target data. Additionally or alternatively, a suitable computing system (e.g., the risk evaluation computing system or a computing system receiving a risk notification from the risk evaluation computing system) may perform some type of operation to discontinue the software application's use of the personal data. In one example, according to one aspect, such a computing system may have the software application made unavailable ("locked") so that users are unable to access the software application. In another example, such a computing system may disable the functionality of the software application that directly makes use of the target data.

**[0032]** Various aspects of the disclosure provided herein address technical disadvantages encountered in designing or implementing various software applications such as websites, mobile applications, and/or the like that use target data. Specifically, various aspects of the disclosure provide a risk evaluation computing system that can monitor various software applications and evaluate the risk involved in the applications' use of target data. Further, the risk evaluation computing system can perform operations so that the risk is addressed accordingly, such as by causing the software applications or specific functionality therein to be modified or disable. As a result, various aspects of the disclosure can increase the security, reliability, capacity, and efficiency in using software applications in conjunction with target data. In doing so, various aspects of the present disclosure make major technical contributions to improving the use of such applications. This in turn translates to more computationally reliable, secure, and/or efficient systems that process target data.

**[0033]** For purposes of this disclosure, a website is discussed as the software application making use of target data in the remainder of the disclosure in describing aspects of the disclosure. However, various aspects of the disclosure may be used in conjunction with other forms of software applications such as mobile applications, software as a service, multi-user software,

and/or the like. A location may be referred to as a jurisdiction. Accordingly, a "jurisdiction" as used herein may refer to, for example, a country, region, group of countries, legal jurisdiction, federation of countries, and/or any other area to which a set of laws and/or regulations may apply.

**[0034]** In addition, it is noted that reference is made to target data throughout the remainder of the application. However, targeted data is not necessarily limited to information that may be configured as personal and/or sensitive in nature but may also include other forms of data that may introduce risk and be of interest to an entity who is operating a software application making use of such data. For example, target data may include data that relates to an entity's business, operational procedures, legal obligations, and/or the like that are to remain out of the public eye (e.g., trade secret) that can pose a risk if exposed. Further, targeted data may not necessarily be associated with an individual but may be associated with other entities such as a business, organization, government, association, and/or the like.

Example Computing Environment

**[0035]** Referring now to the figures, FIG. 1 depicts an example of a computing environment for evaluating a risk associated with a website's use of target data according to various aspects. For example, a website may use various technologies, such as webforms and/or cookies, to request, collect, track, and/or the like the target data of individuals who visit a website. Therefore, the entity (e.g., organization) providing the website may be interested in understanding the risk involved with having the website use the various technologies in requesting, collecting, tracking, and/or the like the target data of individuals who visit the website.

**[0036]** The entity may provide (e.g., publish) the website through a third-party computing system 160 in which one or more sets of program code may reside that is executable by computing hardware (e.g., Web server and/or application server) for performing one or more functions for controlling the website, providing content on the website, providing functionality for the website, and/or the like. An individual may visit the website by navigating to an address (e.g., uniform resource locator) for the website over a network 150 (e.g., Internet) using a browser application residing on a computing device 170 being used by the individual.

**[0037]** Accordingly, the one or more sets of program code may perform functions such as controlling the operation of the computing device 170 by, for example, rendering one or more webpages for the website on a display of the computing device 170. In addition, the one or more

sets of program code may control the transferring of target data of the individual to one or more remote computing systems that collect and/or process the target data. Further, the one or more sets of program code may display a webform, for example, for collecting target data from the individual. In some instances, the entity may store the one or more sets of program code for the website in some type of third-party data storage 180 to make the set(s) of program code available for analysis. The third-party data storage 180 may reside within the third-party system 160 or externally, as shown in FIG. 1.

**[0038]** As noted, the entity may be interested in understanding the risk involved with having the website use the target data of individuals who visit the website. Accordingly, a separate entity may provide a service through a risk analysis computing system 100 for analyzing and evaluating the risk associated with the website's use of the target data. Therefore, the entity associated with the website may make the website (e.g., the one or more sets of program code for the website) available to the risk analysis computing system 100. For example, the entity may upload the website over a network 150 (e.g., the Internet) to the risk analysis computing system 100 from the third-party computing system 160 and/or the third-party data storage 180. In another example, the risk analysis computing system 100 may retrieve the website from the third-party computing system 160 and/or the third-party data storage 180. Yet, in another example, the risk analysis computer system 100 may be configured to perform the analysis on the website directly in the third-party computing system 160 and/or on the third-party data storage 180.

**[0039]** According to various aspects of the disclosure, the risk analysis computing system 100 may comprise computing hardware performing a number of different processes in conducting the analysis and evaluation of the risk involved in the website's use of target data. Specifically, according to particular aspects, the risk analysis computing system 100 performs an evaluation process 110 in evaluating the risk associated with the website's use of target data. As further detailed herein, according to some aspects, the evaluation process 110 involves scanning the website to identify uses of target data such as, for example, a webform provided in the website to allow an individual (user) to provide target data such as, for example, personal data. The evaluation process 110 continues with determining a risk associated with each use of the target data and whether the risk satisfies a threshold. If a risk for a particular use of target data satisfies the threshold, then the evaluation process 110 involves performing one or more operations to mitigate

the risk. For example, the operation may involve notifying personal of the risk. In another example, the operation may involve disabling the use of the target data for the website.

**[0040]** According to particular aspects, the risk analysis computing system 100 may determine the risk associated with a particular use of target data by the website by performing a risk determination process 120. The risk determination process 120 may involve determining the type(s) of target data being used by the functionality (e.g., the webform) of the website involved in the use and determining a vendor associated with the functionality. In addition, the risk determination process 120 may involve determining one or more locations associated with the functionality. The risk determination process 120 may then involve determining the risk for the use of the target data by using a combination of the type(s) of target data being used and the location(s). According to particular aspects, the risk determination process 120 may use a rules-based model, a machine-learning model, or a combination of both in determining the risk.

**[0041]** In addition, according to particular aspects, the risk analysis computing system 100 may determine the type(s) of target data being used by performing a data type determination process 130. The data type determination process 130 may involve identifying a type of data associated with different elements, fields, and/or the like associated with the functionality involved in the use of the target data. The data type determination process 130 may use a rules-based model, a machine-learning model, or a combination of both in identifying the type(s) of target data being used. For example, the rules-based model, machine-learning model, or combination of both may be configured to process metadata associated with the different elements, fields, and/or the like in determining the type(s) of target data being used.

**[0042]** According to particular aspects, the risk analysis computing system 100 may also determine the locations(s) associated with the functionality involved in the use of the target data by performing a location determination process 140. The location determination process 140 may involve determining IP address(es) associated with the determined vendor for the functionality involved in the use of the target data. To accomplish this, the location determination process 140 may involve initiating a series of activations of the functionality from multiple systems located in various geographical locations and analyzing the resulting communications between the functionality and any remote systems. Accordingly, the location determination process 140 may involves analyzing the resulting traffic to determine the source and destination addresses (e.g., IP addresses) of such data communications. Further detail is now provided on the configuration and

functionality of the different processes 110, 120, 130, 140 according to various aspects of the disclosure.

Evaluation of Risk

**[0043]** Turning now to FIG. 2, additional details are provided regarding an evaluation process 100 for evaluating a risk associated with a website's use of target data in accordance with various aspects of the disclosure. Accordingly, the process 110 may be implemented according to various aspects as suitable program code executed on, for example, computing hardware found in the risk analysis computing system 100 as described herein.

**[0044]** Depending on the circumstances, the evaluation process 110 may be carried out to evaluate a website's use of target data at various times. For instance, according to particular aspects, the evaluation process 110 may be carried out at a time when new content has been added to one or more webpages of the website and as a result, a new or revised use of target data may have been introduced to the one or more webpages. For example, a new code release may have been issued for the website and/or a tag manager used for the website may have been updated. While in other instances, the evaluation process 110 may be carried to evaluate a website "on the fly" at a time when a user is visiting the website and the one or more webpages are being rendered. Such a configuration can ensure that any use of target data that has been newly added or modified on the website is identified and evaluated.

**[0045]** A website may collect target data from users (e.g., visitors to the website) and transfer that data to other vendors, services, systems, and/or the like. The vendors, services, systems, and/or the like to which the target data is transferred may be located in one or more locations (jurisdictions) outside of the location (jurisdiction) in which the user is found (e.g., resides and/or from which the user is accessing the website). Therefore, the website may be sending the user's target data to a different jurisdiction. In some instances, such a transfer may represent a significant risk with respect to experiencing a data privacy incident such as a breach of the target data during the transfer. In addition, a location in which the data is received may be a jurisdiction in which the laws and regulations applicable to the target data may be different from those of the user's jurisdiction. This may further increase the risk of using the website.

**[0046]** The evaluation process 110 involves performing a scan of the website to identify any use of target data associated with the website at Step 210. For instance, the website may use a

webform or equivalent computing construct that is configured to allow a user to enter target data that is then transferred to a vendor, service, system, and/or the like for processing. The user may enter data into one or more fields or elements (e.g., HyperText Markup Language (HTML) form elements, HTML fields, etc.) in the webform. Each such field or element may have metadata associated with it, as may the webform itself. Therefore, according to various aspects, the evaluation process 110 may involve dynamically scanning the website (e.g., one or more webpages thereof) to detect a webform set having one or more webforms and the elements and/or fields associated with the detected webforms. For instance, according to particular aspects, the evaluation process 110 involves analyzing the HTML used to generate the website to identify the webform set and the associated one or more elements and/or fields.

**[0047]** The evaluation process 110 also involves selecting a first webform from the webform set at Step 215 and determining a risk associated with the collection of the target data by the webform at Step 220. Although a webform is often used on a website to collect target data from a user (e.g., visitor to the website), the webform may not always serve in such a role with respect to target data. For example, the webform may instead be configured to receive target data for the user from an external source. For instance, the webform may receive information (data) from the user such as a username and password and then retrieve target data from some other data source such as the user's employee identifier, home address, social security number, and/or the like.

**[0048]** According to various aspects, this particular step of the evaluation process 110 is performed via a risk determination process 120 shown in FIG. 3. As detailed further herein, the risk determination process 120 involves identifying the type of target data being collected, transferred, processed, and/or the like by the webform, as well as identifying the vendor(s), service(s), system(s), and/or the like that may involve a transfer of the target data and the one or more locations associated with the transfer. The risk determination process 120 then involves generating a level of risk associated with the webform's use of the target data based at least in part on the types of target data, transfers of the target data, and associated location(s) thereof.

**[0049]** The evaluation process 110 continues with determining whether another webform was identified for the website that also makes use of target data at Step 225. If so, then the evaluation process involves returning to Step 215, selecting the next webform, and determining the risk associated with the use of the target data by the newly selected webform as just described.

**[0050]** The evaluation process 110 then continues with determining whether risk associated with any of the webforms satisfies a threshold level of risk at Step 230. For example, the threshold level of risk may be set by the entity (e.g., employee of the entity) associated with the website. In another example, the threshold level of risk be set by an automated process, such as by a rules-based model, that sets the threshold based on one or more factors such as the type of target data being used, the type of use being made of the target data by the website, a retention time of the target data, whether the target data is encrypted, and/or the like. According to various aspects, the risk associated with each of the webforms may represent a likelihood of an entity (e.g., organization) operating the website experiencing a data privacy incident as a result of the webform's use of the target data. In addition, or instead, the risk associated with each of the webforms may represent a likelihood of the entity operating the website in a noncompliant manner with respect to one or more legal and/or industry standards based at least in part on the webform's use of the target data.

**[0051]** According to particular aspects, different threshold levels of risk may be established depending on the type of functionality involved in the use of the target data (e.g., webform, tracking tool, and/or the like). Furthermore, according to some aspects, the risk determination process 120 may involve generating more than one risk for each of the webforms. For example, according to one aspect, the risk determination process 120 may involve generating a first risk representing a likelihood of the entity operating the website and experiencing a data privacy incident as a result of the webform's use of target data and a second, separate risk representing a likelihood of the entity operating the website in a noncompliant manner with respect to one or more legal and/or industry standards based at least in part on the webform's use of the target data. Accordingly, the evaluation process 110 may be carried out to evaluate the multiple risks returned for a webform in determining whether a risk satisfies the threshold level of risk.

**[0052]** Therefore, any of the webforms having a risk that satisfies the threshold level of risk may signal that the webform's use of target data introduces a level of risk that is unacceptable to the entity (e.g., the organization) operating the website. If this is the case, then the evaluation process 110 according to various aspects involves performing one or more operations to mitigate the risk(s) at Step 235. For instance, according to particular aspects, one or more communications may be generated and sent to personnel to notify them of the risk(s). For example, an email or some other type of electronic communication may be generated and sent to the one or more

personnel. In another example, a graphical user interface (GUI) may be generated and/or provided that displays the risk(s) and associated information.

**[0053]** According to other aspects, the evaluation process 110 may involve performing one or more operations to disable the use of the target data by the webform(s). For example, a suitable computing system (e.g., the risk evaluation computing system 100 or the third-party computing system 160 that controls the website) can "lock" the website (and/or one or more webpages thereof) so that the website is no longer available to one or more users. For instance, the risk evaluation computing system 100 and/or the third-party computing system 160 can remove (or cause the removal of) the webpages associated with the webform(s) from the Web server hosting the website. In another example, a suitable computing system (e.g., the risk evaluation computing system 100 or the third-party computing system 160 that controls the website) can disable the functionality (e.g., the webform) so that the functionality is no longer available on the website. For instance, the risk evaluation computing system 100 and/or the third-party computing system 160 can disable the ability of website to display the webform(s) so that the target data cannot be used by the webform(s). In another example, the risk evaluation computing system 100 and/or the third-party computing system 160 can disable one or more controls (e.g., buttons) on the webform(s) so that the target data cannot be saved and/or collected. Accordingly, a combination of operations may be performed such as, for example, disabling the functionality and sending one or more communications to personal to notify them of the risk(s) imposed by the use of the target data on the webform(s). Those of ordinary skill in the art can envision other operations that may be carried out according to various aspects to mitigate the risk(s) in light of this disclosure.

**[0054]** As a result, one or more risks introduced by the website's use of target data that poses a level of risk greater than the entity (e.g., the organization) operating the website would like to tolerate can be addressed so as to bring the level of risk to an acceptable level for the entity. In addition, the evaluation process 110 can be carried out in an automated fashion to evaluate and identify risk(s) imposed by a website's use of target data that is problematic for the entity and may not otherwise be identified. For example, the evaluation process 110 may be carried out according to particular aspects to periodically (e.g., weekly, monthly, bi-monthly, and/or the like) evaluate the website. Accordingly, the evaluation process 110 may enable the identification of any risk(s) that have been introduced as a result of the website being modified to include new and/or changed use of target data.

**[0055]** Further, according to some aspects, the evaluation process 110 may involve identifying any changes that have been introduced into the website with respect to the website's use of target data. For example, although not shown in FIG. 2, the evaluation process 110 (or some other process) may involve comparing the fields and/or elements currently found in the webform with the fields and/or elements found in a previous version of the webform to identify any differences in the fields and/or elements such as new fields and/or elements being added to the webform and/or the use of a particular field and/or element being changed. Identifying such changes can help in identifying a source resulting in (causing) a change in the level of risk imposed by the webform's use of target data and may help in mitigating the change in the level of risk. For example, identifying the changes in the webform may help personnel to pinpoint what may have caused an increase in risk of the webform's use of target data.

**[0056]** Therefore, in order to ensure the evaluation process 110 is carried out on current, up to date information regarding the website, the evaluation process 110 may be performed to repeatedly (e.g., periodically, in response to receiving an instruction, in response to detecting a change in the website, a webform, and/or the like) scan the website to determine any changes in the target data used by the website (e.g., any change in the target data that is collected and/or how the target data is processed). After performing an initial, or any subsequent, scan and the related processing for the website as described herein, the data generated by the scan may be stored (e.g., element/field identification, classifications, data types, sources, destinations, etc.). The evaluation process 110 may then be performed later to re-scan the particular website and again perform the processing described herein. The evaluation process 110 may involve comparing the resulting data generated by the most recent analysis to data generated by a previous (e.g., most recent previous) analysis to determine whether there have been any changes in how the website, webform, and/or the like processes target data. For example, a determination may be made as to whether new elements, fields, and/or the like have been added to a website since the last scan, whether such new elements, fields, and/or the like are associated with target data, and/or whether the website now sends target data to a new and/or different location. According to some aspects, the evaluation process 110 may involve generating a notification to personal indicating a change in response to determining that the change has occurred to the website's use of target data.

**[0057]** Although the example of the evaluation process 110 discussed above involves analyzing webforms found on the website with respect to the risk imposed by the webforms' use

of target data, the evaluation process 110 according to various aspects may be used to evaluate other functionality found on the website that makes use of target data. For example, the evaluation process 110 may also be performed to identify other functionality such as tracking tools (e.g., scripts, cookies, web beacons, and/or the like) found on the website that use target data. For example, according to these aspects, the evaluation process 110 may involve performing a scan of a webpage of the website to identify any tracking tools associated with the webpage and/or any respective scripts that may be used to execute, load, introduce, and/or the like the tracking tools. Specifically, for example, the evaluation process 110 may involve using a scanner such as Chrome scanner to scan the webpage as the webpage is being loaded to identify the tracking tools and/or associated scripts. Accordingly, the evaluation process 110 may then involve evaluating the risk imposed by the identified tracking tools' use of target data in the same manner as discussed above with respect to the webforms. Those of ordinary skill in the art will recognize other functionality that may be found on the website and use target data that the evaluation process 110 may be carried out to evaluate according to various aspects in light of this disclosure.

**[0058]** Further, although the example of the evaluation process 110 discussed above involves analyzing websites and their use of target data, aspects of the evaluation process 110 may be carried out to evaluate other forms of software applications that may use target data such as mobile applications. For example, a mobile application may collect target data through a GUI displayed in the mobile application. According to various aspects, the evaluation process 110 may be performed to evaluate these other forms of software applications in a similar manner as described above with respect to websites. For example, Step 210 may be implemented by analyzing source code of a software application. A third-party computing system 160 can provided access to this source code via a communication channel with the risk evaluation computing system 100. Here, the evaluation process 110 may consider aspects of the software application other than GUIs used in collecting target data in evaluating the risk associated with the software application's use of the target data such as, for example, software development kits (SDKs) that have been used in implementing functionality in the software application. Certain SDKs may be known for using target data for particular applications. Therefore, identification of the use of an SDK may lead to the identification of the use of target data for a software application.

Risk Determination

**[0059]** Turning now to FIG. 3, additional details are provided regarding a risk determination process 120 for determining a risk associated with the use of target data on a website in accordance with various aspects of the disclosure. Accordingly, the process 120 may be implement according to various aspects as suitable program code executed on, for example, computing hardware found in the risk analysis computing system 100 as described herein.

**[0060]** The risk determination process 120 involves determining the type(s) of target data used (e.g., collected, processed, transferred, and/or the like) by the functionality at Step 310. Accordingly, this particular step is performed in various aspects via a data type determination process 130 shown in FIG. 4. As discussed further herein, the data type determination process 130 involves using the metadata associated with each of the elements, fields, and/or the like found in the functionality in determining the type(s) of target data used by the functionality.

**[0061]** The risk determination process 120 continues with determining a vendor associated with the functionality (e.g., the webform) at Step 315. For instance, a webform may be associated with a particular vendor or service that may or may not be the same vendor or service providing the website on which the webform is presented. For example, a particular webform may be configured to transmit collected data to a vendor software application such as Salesforce, Marketo, etc., or to a custom web server. Therefore, according to various aspects, the risk determination process 120 involves using metadata associated with the webform (e.g., the webform itself and/or one or more elements, fields, and/or the like of the webform) to determine a vendor associated with the webform.

**[0062]** For example, according to particular aspects, the risk determination process 120 may involve using information derived from the metadata associated with the webform to identify a URL to which the webform transmits collected target data. The URL may then be used to determine the associated vendor, service, system, and/or the like to which data is transferred. According to particular aspects, the risk determination process 120 may be carried out by referencing a data source such as a dictionary, catalog, and/or the like of known URLs and associated vendors, services, systems, and/or the like to locate a URL matching the URL derived from a particular webform's metadata. According to other aspects, the risk determination process 120 may be carried out by determining the domain associated with the URL and then determining

the vendor, service, system, and/or the like based on the domain (e.g., by performing an ICANN or "whois" lookup on the domain name).

**[0063]** In another example, the functionality making use of the target data may be a tracking tool. Here, the risk determination process 120 may involve identifying a source for the tracking tool by analyzing one or more flows of data, for example, between a browser rending a webpage and a server serving the webpage to the browser, or between the browser and one or more remote systems (e.g., remote computing entities that one or more scripts loading on the webpage attempt to communicate with). For instance, the risk determination process 120 may involve scanning one or more response headers to identify a source or initiator of the particular tracking tool such as, for example, scanning one or more response headers that have been sent to the browser by a host server associated with the particular tracking tool in response to the host server receiving an HTTP request. Here, the response header may include, for example, a date, size, and/or type of file that the host server is attempting to send to the browser, as well as, or instead, other data such as data about the host server itself. This header information may be used to match a source script with the particular tracking tool or otherwise determine a source script for the particular tracking tool. Depending on the functionality, other steps may be used for determining the vendor(s), service(s), system(s), and/or the like associated with the functionality making use of target data.

**[0064]** The risk determination process 120 continues with determining one or more locations associated with the functionality making use of the target data at Step 320. This particular step is performed in various aspects via a location determination process 140 shown in FIG. 6. As detailed further herein, the location determination process involves identifying the location(s) associated with the functionality based at least in part on the vendor(s), service(s), system(s), and/or the like associated with the functionality. For instance, the functionality may transfer and/or receive target data from one or more of the identified vendor(s), service(s), system(s), and/or the like associated with the functionality. Therefore, the locations (e.g., jurisdictions) in which these vendor(s), service(s), system(s), and/or the like are located may be identified as location(s) associated with the functionality.

**[0065]** The risk determination process 120 continues with determining the risk associated with the functionality at Step 325. Here, according to various aspects, the risk associated with the functionality represents the risk posed by the functionality's use of the target data in experiencing a data privacy incident and/or leading to nonconformance with one or more legal and/or industry

standards. This particular step is performed according to various aspects by using a combination of the type(s) of data collected in conjunction with the associated vendor(s), service(s), system(s), and/or the like and location(s) of the functionality to determine a risk associated with the functionality.

**[0066]** For example, a webform may be determined to collect no target data, or particularly low risk target data (e.g., name only, birthday only, etc.), and therefore represents a low risk even when the webform transmits collected target data outside of the location (e.g., jurisdiction) in which such data is collected. Alternatively, a webform may be determined to represent a high risk if the webform collects target data of a particularly personal nature (e.g., social security number, phone number, driver's license number, etc.) and transmits the target data outside of the jurisdiction in which such data was collected. In another example, a webform may be determined to represent a high risk, even if the webform does not transmit target data outside of the jurisdiction in which such data is collected, where the webform transmits such data to a known high-risk vendor, service, system, and/or the like within the jurisdiction.

**[0067]** Accordingly, the processing of particular types of target data may be regulated differently in different jurisdictions. Therefore, if a particular webform transmits a type of target data to a jurisdiction in which processing that type of target data is highly regulated (e.g., subject to high monetary fines for mishandling), then the risk determination process 120 may lead to a determination that the risk associated with that particular webform is high. Alternatively, if the particular webform transmits the type of target data to a jurisdiction in which processing that type of target data is not highly regulated (e.g., subject to minimal penalties for mishandling), then the risk determination process 120 may lead to a determination that the risk associated with that particular webform is low.

**[0068]** Here, according to particular aspects, the risk determination process 120 may involve using a rules-based model in determining the risk associated with the functionality. The rules-based model may comprise a set of rules that sets a risk for the functionality's use of the target data based at least in part on the combination of the type(s) of target data being collected, the vendor(s), service(s), system(s), and/or the like associated with the functionality, and/or the location(s) associated with the functionality. Accordingly, an entity (e.g., an organization) conducting the risk analysis may maintain the set of rules in some type of data storage, such as a database, from which the set of rules can be accessed. According to some aspects, a user interface

(e.g., graphical user interface) may be provided so that personnel of the entity may maintain the set of rules.

**[0069]** According to other aspects, the risk determination process 120 may involve using a machine-learning model in determining the risk associated with the functionality. Here, the machine-learning model may be trained using historical data on the same or similar functionality's use of the same or similar type(s) of target data, in association with the same or similar vendor(s), service(s), system(s), and/or the like, and/or location(s). For instance, the machine-learning model may determine a risk in the form of a prediction as to the likelihood of the entity to experience a data privacy incident and/or noncompliance with a legal and/or industry standard due to the functionality's use of the target data. Accordingly, the machine-learning model may be configured using a variety of different types of supervised or unsupervised trained models such as, for example, support vector machine, naive Bayes, decision tree, neural network, and/or the like.

**[0070]** According to particular aspects, the risk determination process 120 may involve using a combination of the rules-based model and the machine-learning model in determining the risk associated with the functionality's use of the target data. Further, according to particular aspects, the rules-based model and/or machine-learning model may be configured to generate separate risks, a first risk associated with the entity experiencing a data privacy incident due to the functionality's use of the target data and a second, separate risk associated with the entity being noncompliant with one or more legal and/or industry standards due to the functionality's use of the target data. In doing so, the risk determination process 120 can be used to identify the risk associated with specific aspects of the functionality's use of the target data. Yet, according to other aspects, the rules-based model and/or machine-learning model may be configured to generate separate risks at a more detailed level such as, for example, a risk with respect to the entity being noncompliant with a specific legal and/or industry standard. This may enable the entity to better recognize and address any significant risk (e.g., satisfies a threshold) posed by the functionality's use of the target data.

Data Type Determination

**[0071]** Turning now to FIG. 4, additional details are provided regarding a data type determination process 130 for determining the types of target data associated with functionality in accordance with various aspects of the disclosure. Accordingly, the process 130 may be

implemented according to various aspects as suitable program code executed on, for example, computing hardware found in the risk analysis computing system 100 as described herein.

**[0072]** The data type determination process 130 involves selecting an element, field, and/or the like for the functionality at Step 410 and identifying the type of data associated with the element, field, and/or the like at Step 415. For instance, if the functionality involves a webform, then the metadata associated with the element, field, and/or the like may be evaluated in identifying the type of data associated with the element, field, and/or the like.

**[0073]** Here, according to various aspects, the data type determination process 130 may involve using a rules-based model in identifying the type of data associated with the element, field, and/or the like. For instance, the rules-based model may comprise a set of rules that identifies a type of data for the element, field, and/or the like based at least in part on metadata associated with the element, field, and/or the like. Accordingly, an entity (e.g., an organization) conducting the risk analysis may maintain the set of rules in some type of data storage, such as a database, from which the set of rules can be accessed. According to some aspects, a user interface (e.g., graphical user interface) may be provided so that personnel of the entity may maintain the set of rules.

**[0074]** According to other aspects, the data type determination process 130 may involve using a machine-learning model in identifying the type of data associated with the element, field, and/or the like. According to some aspects, the machine-learning model may be some type of classification model (e.g., multi-label classification model) that provides a classification for the element, field, and/or the like. For example, the machine-learning model may be a supervised or unsupervised trained model that provides a prediction as to the type of data associated with the element, field, and/or the like. The machine-learning model may use a variety of different types of prediction models such as, for example, support vector machines, logistic regression, neural network, and/or the like. Here, the metadata for the element, field, and/or the like may be processed (e.g., provided as input) using the machine-learning model and the model may provide output in the form of a classification for the type of data associated with the element, field, and/or the like.

**[0075]** Accordingly, the rules-based model, the machine-learning model, or combination thereof may provide, for example, an output identifying the particular element, field, and/or the like as target data or not target data. In another example, the rules-based model, the machine-learning model, or combination thereof may provide output identifying the particular element, field, and/or the like as a level of target data such as high-level target data, medium-level target

data, low-level target data, or data that is not target data. Yet in another example, the rules-based model, machine-learning model, or combination thereof may provide output identifying the actual data associated with the particular element, field, and/or the like such as first name, last name, home address, social security number, age, gender, race, and/or the like. For example, the rules-based model and/or machine-learning model may provide the output in the form of a feature representation, such as a feature vector, in which each element found in the feature representation represents a specific type of target data and the value of each element identifies the likelihood of the particular element, field, and/or the like of being that specific type of target data. The rules-based model and/or machine-learning model may also determine and provide a confidence score in identifying the type of data along with the output.

**[0076]** For instance, briefly turning to FIG. 5, this figure illustrates the HTML and metadata of an example website. Here, the webform "validator" 500 may be identified for the website and an input field may be identified as being displayed on the webform having a name of the field as "first_name" 510. Accordingly, the rules-based model and/or machine-learning model may process the metadata for the input field and provide output identifying that the type of data associated with the field "first_name" 510 is the first name of a user who is providing input to the webform. Further, the output may provide a confidence score with respect to the identified type of data and/or indicate this particular type of data represents a type of target data.

**[0077]** The data type determination process 130 continues with determining whether there is another element, field, and/or the like associated with the functionality at Step 420. If so, then the data type determination process 130 involves returning to Step 410, selecting the next element, field, and/or the like, and identifying the data type associated with the newly selected element, field, and/or the like. The data type determination process 130 concludes with saving the identified type(s) of data at Step 425. As previously discussed, the risk determination process 120 according to various aspects involves using the type(s) of data associated with the functionality in determining a risk associated with the functionality's use of the target data.

Location Determination

**[0078]** Turning now to FIG. 6, additional details are provided regarding a location determination process 140 for determining one or more locations associated with functionality's use of target data in accordance with various aspects of the disclosure. For instance, the one or

more locations may be associated with locations to which the functionality transfers and/or receives target data or at which a computing entity such as a server is located that calls, loads, executes, and/or the like the functionality. Accordingly, the process 140 may be implemented according to various aspects as suitable program code executed on, for example, computing hardware found in the risk analysis computing system 100 as described herein.

**[0079]** The location determination process 140 involves, at Step 610, determining one or more IP addresses associated with the determined vendor(s), service(s), system(s), and/or the like for the functionality. To accomplish this, the location determination process 140 according to various aspects involves initiating a series of activations of the functionality (e.g., the webform) from multiple systems located in various geographical locations and analyzing the resulting communications between the functionality and any remote systems. Activating the functionality from the different geographical locations may cause the functionality to attempt to connect to the associated URLs from each of these geographical locations. Accordingly, the location determination process 140 involves analyzing the resulting traffic to determine the source and destination addresses (e.g., IP addresses) of such data communications at Step 615. For example, using the identified addresses, a geographical or jurisdictional location for the remote system may be determined by performing a reverse IP address look-up.

Example Technical Platforms

**[0080]** Aspects of the present disclosure may be implemented in various ways, including as computer program products that comprise articles of manufacture. Such computer program products may include one or more software components including, for example, software objects, methods, data structures, and/or the like. A software component may be coded in any of a variety of programming languages. An illustrative programming language may be a lower-level programming language such as an assembly language associated with a particular hardware architecture and/or operating system platform. A software component comprising assembly language instructions may require conversion into executable machine code by an assembler prior to execution by the hardware architecture and/or platform. Another example programming language may be a higher-level programming language that may be portable across multiple architectures. A software component comprising higher-level programming language instructions

may require conversion to an intermediate representation by an interpreter or a compiler prior to execution.

**[0081]** Other examples of programming languages include, but are not limited to, a macro language, a shell or command language, a job control language, a script language, a database query, or search language, and/or a report writing language. In one or more example aspects, a software component comprising instructions in one of the foregoing examples of programming languages may be executed directly by an operating system or other software component without having to be first transformed into another form. A software component may be stored as a file or other data storage construct. Software components of a similar type or functionally related may be stored together such as, for example, in a particular directory, folder, or library. Software components may be static (e.g., pre-established, or fixed) or dynamic (e.g., created or modified at the time of execution).

**[0082]** A computer program product may include a non-transitory computer-readable storage medium storing applications, programs, program modules, scripts, source code, program code, object code, byte code, compiled code, interpreted code, machine code, executable instructions, and/or the like (also referred to herein as executable instructions, instructions for execution, computer program products, program code, and/or similar terms used herein interchangeably). Such non-transitory computer-readable storage media include all computer-readable media (including volatile and non-volatile media).

**[0083]** According to various aspects, a non-volatile computer-readable storage medium may include a floppy disk, flexible disk, hard disk, solid-state storage (SSS) (e.g., a solid-state drive (SSD), solid state card (SSC), solid state module (SSM), enterprise flash drive, magnetic tape, or any other non-transitory magnetic medium, and/or the like. A non-volatile computer-readable storage medium may also include a punch card, paper tape, optical mark sheet (or any other physical medium with patterns of holes or other optically recognizable indicia), compact disc read only memory (CD-ROM), compact disc-rewritable (CD-RW), digital versatile disc (DVD), Blu-ray disc (BD), any other non-transitory optical medium, and/or the like. Such a non-volatile computer-readable storage medium may also include read-only memory (ROM), programmable read-only memory (PROM), erasable programmable read-only memory (EPROM), electrically erasable programmable read-only memory (EEPROM), flash memory (e.g., Serial, NAND, NOR, and/or the like), multimedia memory cards (MMC), secure digital (SD) memory cards,

SmartMedia cards, CompactFlash (CF) cards, Memory Sticks, and/or the like. Further, a non-volatile computer-readable storage medium may also include conductive-bridging random access memory (CBRAM), phase-change random access memory (PRAM), ferroelectric random-access memory (FeRAM), non-volatile random-access memory (NVRAM), magnetoresistive random-access memory (MRAM), resistive random-access memory (RRAM), Silicon-Oxide-Nitride-Oxide-Silicon memory (SONOS), floating junction gate random access memory (FJG RAM), Millipede memory, racetrack memory, and/or the like.

**[0084]** According to various aspects, a volatile computer-readable storage medium may include random access memory (RAM), dynamic random access memory (DRAM), static random access memory (SRAM), fast page mode dynamic random access memory (FPM DRAM), extended data-out dynamic random access memory (EDO DRAM), synchronous dynamic random access memory (SDRAM), double data rate synchronous dynamic random access memory (DDR SDRAM), double data rate type two synchronous dynamic random access memory (DDR2 SDRAM), double data rate type three synchronous dynamic random access memory (DDR3 SDRAM), Rambus dynamic random access memory (RDRAM), Twin Transistor RAM (TTRAM), Thyristor RAM (T-RAM), Zero-capacitor (Z-RAM), Rambus in-line memory module (RIMM), dual in-line memory module (DIMM), single in-line memory module (SIMM), video random access memory (VRAM), cache memory (including various levels), flash memory, register memory, and/or the like. It will be appreciated that where various aspects are described to use a computer-readable storage medium, other types of computer-readable storage media may be substituted for or used in addition to the computer-readable storage media described above.

**[0085]** Various aspects of the present disclosure may also be implemented as methods, apparatuses, systems, computing devices, computing entities, and/or the like. As such, aspects of the present disclosure may take the form of a data structure, apparatus, system, computing device, computing entity, and/or the like executing instructions stored on a computer-readable storage medium to perform certain steps or operations. Thus, aspects of the present disclosure also may take the form of an entirely hardware aspect, an entirely computer program product aspect, and/or an aspect that comprises combination of computer program products and hardware performing certain steps or operations.

**[0086]** Various aspects of the present disclosure are described below with reference to block diagrams and flowchart illustrations. Thus, each block of the block diagrams and flowchart

illustrations may be implemented in the form of a computer program product, an entirely hardware aspect, a combination of hardware and computer program products, and/or apparatus, systems, computing devices, computing entities, and/or the like carrying out instructions, operations, steps, and similar words used interchangeably (e.g., the executable instructions, instructions for execution, program code, and/or the like) on a computer-readable storage medium for execution. For example, retrieval, loading, and execution of code may be performed sequentially such that one instruction is retrieved, loaded, and executed at a time. In some exemplary aspects, retrieval, loading, and/or execution may be performed in parallel such that multiple instructions are retrieved, loaded, and/or executed together. Thus, such aspects can produce specially configured machines performing the steps or operations specified in the block diagrams and flowchart illustrations. Accordingly, the block diagrams and flowchart illustrations support various combinations of aspects for performing the specified instructions, operations, or steps.

*Example System Architecture*

**[0087]** FIG. 7 is a block diagram of an example system architecture 700 that may be used in accordance to various aspects of the disclosure. According to various aspects, the architecture 700 associated with a particular organization and be configured to aid in evaluating software applications with respect to the risk associated with their use of target data. As may be understood from FIG. 7, the system architecture 700 includes a risk determination system 100 that comprises one or more servers 715 for carrying out the processes described herein for conducting the risk evaluation process 110, risk determination process 120, the data type determination process 130, and the location determination process 140 as detailed herein. In addition, the system architecture 700 may include one or more storage devices 720. Here, the storage devices 620 may located within or outside (as shown) of the risk determination system 100 and may be used in storing the software applications that are evaluated. In addition, the storage devices 720 may store various rules used within rules-based models as described herein.

**[0088]** The system architecture 700 may include one or more computer networks 150 that facilitate communication between the one or more servers 715 and the storage devices 720, as well as one or more remote servers 725 to which target data may be transferred. In addition, the one or more computer networks 150 may be used for downloading or accessing a software application that is to be analyzed. For example, the software application may reside in a third-party computing

system 160 and/or third-party storage 180 (not shown in FIG. 7) from which the software application is upload to or access by the risk evaluation system 100. Here, the one or more computer networks 150 may include any of a variety of types of wired or wireless computer networks such as the Internet, a private intranet, a public switched telephone network (PSTN), or any other type of network. Accordingly, the communication link between the one or more servers 715, storage devices 720, and/or remote servers 725 may be, for example, implemented via a Local Area Network (LAN), a Wide Area Network (WAN), the Internet, and/or the like.

**[0089]** In various aspects, any of the servers 715, 725 may comprise a single server, a plurality of servers, one or more cloud-based servers, or any other suitable configuration. In addition, the storage devices 720 may be stored either fully or partially on any suitable server or combination of servers described herein. Further, the one or more servers 715 and/or the storage devices 720 may be physically located in a same (e.g., central) location, such as, for example, the headquarters of the particular organization, or in separate locations.

*Example Computing Entity*

**[0090]** FIG. 8 illustrates a diagrammatic representation of an example computing entity 800 that may be used in accordance with various aspects of the disclosure. For example, the computing entity 800 may be computing hardware such as the one or more servers 715 as described in FIG. 7. In particular aspects, the computing entity 800 may be connected (e.g., networked) to one or more other computing entities, storage devices, and/or the like via one or more networks such as, for example, a LAN, a WAN, and/or the Internet. As noted above, the computing entity 800 may operate in the capacity of a server and/or a client device in a client-server network environment, or as a peer computing device in a peer-to-peer (or distributed) network environment. According to various aspects, the computing entity 800 may be, for example, a personal computer (PC), a laptop computer, a web appliance, a server, a network router, a switch or bridge, or any other device capable of executing a set of instructions (sequential or otherwise) that specify actions to be taken by that device. Further, while only a single computing entity 800 is illustrated, the term "computing entity" shall also be taken to include any collection of computing entities that individually or jointly execute a set (or multiple sets) of instructions to perform any one or more of the methodologies discussed herein.

**[0091]** An exemplary computing entity 800 includes a processor 802, a main memory 804 (e.g., read-only memory (ROM), flash memory, dynamic random-access memory (DRAM) such as synchronous DRAM (SDRAM), Rambus DRAM (RDRAM), and/or the like), a static memory 806 (e.g., flash memory, static random-access memory (SRAM), and/or the like), and a data storage device 818, that communicate with each other via a bus 832.

**[0092]** The processor 802 may represent one or more general-purpose processing devices such as a microprocessor, a central processing unit, and/or the like. According to some aspects, the processor 802 may be a complex instruction set computing (CISC) microprocessor, reduced instruction set computing (RISC) microprocessor, very long instruction word (VLIW) microprocessor, processor implementing other instruction sets, processors implementing a combination of instruction sets, and/or the like. According to some aspects, the processor 802 may be one or more special-purpose processing devices such as an application specific integrated circuit (ASIC), a field programmable gate array (FPGA), a digital signal processor (DSP), network processor, and/or the like. The processor 802 may be configured to execute processing logic 826 for performing various operations and/or steps described herein.

**[0093]** The computing entity 800 may further include a network interface device 808, as well as a video display unit 810 (e.g., a liquid crystal display (LCD), a cathode ray tube (CRT), and/or the like), an alphanumeric input device 812 (e.g., a keyboard), a cursor control device 814 (e.g., a mouse), and/or a signal generation device 816 (e.g., a speaker). The computing entity 800 may further include a data storage device 818. The data storage device 818 may include a non-transitory computer-readable storage medium 830 (also known as a non-transitory computer-readable storage medium or a non-transitory computer-readable medium) on which is stored one or more sets of instructions 822 (e.g., software, software modules) embodying any one or more of the methodologies or functions described herein. The instructions 822 may also reside, completely or at least partially, within main memory 804 and/or within the processor 802 during execution thereof by the computing entity 800 – main memory 804 and processor 802 also constituting computer-accessible storage media. The instructions 822 may further be transmitted or received over a network 150 via the network interface device 808.

**[0094]** While the computer-readable storage medium 830 is shown to be a single medium, the terms "computer-readable storage medium" and "machine-accessible storage medium" should be understood to include a single medium or multiple media (e.g., a centralized or distributed

database, and/or associated caches and servers) that store the one or more sets of instructions. The term "computer-readable storage medium" should also be understood to include any medium that is capable of storing, encoding, and/or carrying a set of instructions for execution by the computing entity 800 and that causes the computing entity 800 to perform any one or more of the methodologies of the present disclosure. The term "computer-readable storage medium" should accordingly be understood to include, but not be limited to, solid-state memories, optical and magnetic media, and/or the like.

Exemplary System Operation

**[0095]** The logical steps described herein may be implemented (1) as a sequence of computer implemented acts or one or more program modules running on a computing system and/or (2) as interconnected machine logic circuits or circuit modules within the computing system. The implementation is a matter of choice dependent on the performance and other requirements of the computing system. Accordingly, the logical steps described herein are referred to variously as states, operations, steps, structural devices, acts, or modules. These operations, steps, structural devices, acts, and modules may be implemented in software, in firmware, in special purpose digital logic, and any combination thereof. Greater or fewer steps may be performed than shown in the figures and described herein. These steps may also be performed in a different order than those described herein.

## Conclusion

**[0096]** While this specification contains many specific aspect details, these should not be construed as limitations on the scope of any invention or of what may be claimed, but rather as descriptions of features that may be specific to particular aspects of particular inventions. Certain features that are described in this specification in the context of separate aspects also may be implemented in combination in a single aspect. Conversely, various features that are described in the context of a single aspect also may be implemented in multiple aspects separately or in any suitable sub-combination. Moreover, although features may be described above as acting in certain combinations and even initially claimed as such, one or more features from a claimed combination may in some cases be excised from the combination, and the claimed combination may be a sub-combination or variation of a sub-combination.

**[0097]** Similarly, while operations are described in a particular order, this should not be understood as requiring that such operations be performed in the particular order described or in sequential order, or that all described operations be performed, to achieve desirable results. In certain circumstances, multitasking and parallel processing may be advantageous. Moreover, the separation of various components in the various aspects described above should not be understood as requiring such separation in all aspects, and the described program components (e.g., modules) and systems may generally be integrated together in a single software product or packaged into multiple software products.

**[0098]** Many modifications and other aspects of the disclosure will come to mind to one skilled in the art to which this disclosure pertains having the benefit of the teachings presented in the foregoing descriptions and the associated drawings. Therefore, it is to be understood that the disclosure is not to be limited to the specific aspects disclosed and that modifications and other aspects are intended to be included within the scope of the appended claims. Although specific terms are employed herein, they are used in a generic and descriptive sense only and not for the purposes of limitation.

## Claims

We claim:

1. A method comprising:

scanning, by computing hardware, a software application to identify functionality configured for processing target data;

identifying, by the computing hardware, a plurality of fields associated with the functionality;

identifying, by the computing hardware, metadata associated with a field from the plurality of fields;

generating, by the computing hardware and from the metadata, an identification of a type of data associated with the field using at least one of a rules-based model or a machine-learning model;

determining, by the computing hardware, a location based on the processing of the target data by the functionality;

determining, by the computing hardware, a risk associated with the functionality processing the target data based on the location and the type of data for the field;

determining, by the computing hardware, that the risk satisfies a threshold level of risk; and

responsive to determining that the risk satisfies the threshold level of risk, causing, by the computing hardware, an action to be performed to mitigate the risk.

2. The method of Claim 1, wherein the action comprises at least one of generating an electronic communication sent to personnel identifying the functionality and the risk, causing the software application to become unavailable, or disabling the functionality in the software application.

3. The method of Claim 1, wherein the risk comprises at least one of a risk of experiencing a data privacy incident due to the functionality processing the target data and a risk of being noncompliant with a data privacy standard due to the functionality processing the target data.

4. The method of Claim 1 further comprising:

determining, by the computing hardware, a vendor associated with the functionality based on metadata associated with the functionality, wherein the location is a jurisdiction in which the vendor processes data and processing of the target data by the functionality involves transferring the target data to the location.

5. The method of Claim 1, wherein the software application comprises a website and the functionality comprises a webform found on the website in which at least one of the plurality of fields is used on the webform to collect the target data.

6. The method of Claim 1, wherein the software application comprises a mobile application and the functionality comprises a graphical user interface provided through the mobile application in which at least one of the plurality of fields is used on the graphical user interface to collect the target data.

7. The method of Claim 1, wherein determining the risk associated with the functionality processing the target data based on the type of data for the field and the location involves using at least one of a second rules-based model or a second machine learning model to generate the risk, wherein the risk represents a likelihood of experiencing at least one of a data privacy incident due to the functionality processing the target data or being noncompliant with a data privacy standard due to the functionality processing the target data.

8. A system comprising:

a non-transitory computer-readable medium storing instructions; and

a processing device communicatively coupled to the non-transitory computer-readable medium,

wherein, the processing device is configured to execute the instructions and thereby perform operations comprising:

scanning a software application to identify functionality configured for processing target data;

identifying metadata associated with the functionality;

processing the metadata using at least one of a rules-based model or a machine learning model to generate an identification of a type of data associated with the functionality;

determining a location based on the processing of the target data by the functionality;

determining a risk associated with the functionality processing the target data based on the type of data and the location;

determining the risk satisfies a threshold level of risk; and

responsive to determining the risk satisfies the threshold level of risk, causing an action to be performed to mitigate the risk.

9. The system of Claim 8, wherein the action comprises at least one of generating an electronic communication sent to personnel identifying the functionality and the risk, causing the software application to become unavailable, or disabling the functionality in the software application.

10. The system of Claim 8, wherein the risk comprises at least one of a risk of experiencing a data privacy incident due to the functionality processing the target data or a risk of being noncompliant with a data privacy standard due to the functionality processing the target data.

11. The system of Claim 8, wherein the operations further comprise:

determining a vendor associated with the functionality based on the metadata, wherein the location is a jurisdiction in which the vendor is located and processing of the target data by the functionality involves transferring the target data to the location.

12. The system of Claim 8, wherein determining the risk associated with the functionality processing the target data based on the type of data and the location involves processing the type of data and the location using at least one of a second rules-based model or a second machine learning model to generate the risk representing a likelihood of experiencing at least one of a data privacy incident due to the functionality processing the target data or being noncompliant with a data privacy standard due to the functionality processing the target data.

13. A non-transitory computer-readable medium having program code that is stored thereon, the program code executable by one or more processing devices for performing operations comprising:

scanning a software application to identify functionality configured for processing target data;

identifying metadata associated with the functionality;

identifying a type of data associated with the functionality based on the metadata;

determining a location based on the processing of the target data by the functionality;

processing the type of data and the location using at least one of a rules-based model or a machine learning model to generate a risk representing a likelihood of experiencing a data incident due to the functionality processing the target data;

determining the risk satisfies a threshold level of risk; and

responsive to determining the risk satisfies the threshold level of risk, causing an action to be performed to mitigate the risk.

14. The non-transitory computer-readable medium of Claim 13, wherein the action comprises at least one of generating an electronic communication sent to personnel identifying the functionality and the risk, causing the software application to become unavailable, or disabling the functionality in the software application.

15. The non-transitory computer-readable medium of Claim 13, wherein identifying the type of data associated with the functionality based on the metadata involves processing the metadata using at least one of a second rules-based model or a second machine learning model to generate an identification of the type of data associated with the functionality.

16. The non-transitory computer-readable medium of Claim 13, wherein the risk comprises at least one of a risk of experiencing a data privacy incident due to the functionality processing the target data or a risk of being noncompliant with a data privacy standard due to the functionality processing the target data.

17. The non-transitory computer-readable medium of Claim 13, wherein the operations further comprise:

determining a vendor associated with the functionality based on the metadata, wherein the location is a jurisdiction in which the vendor is located and processing of the target data by the functionality involves transferring the target data to the location.

18. The non-transitory computer-readable medium of Claim 13, wherein the software application comprises a website and the functionality comprises a webform found on the website in which the functionality is used on the webform to collect the target data.

19. The non-transitory computer-readable medium of Claim 13, wherein the software application comprises a mobile application and the functionality comprises a graphical user interface provided through the mobile application in which a field is used on the graphical user interface to collect the target data.

20. The non-transitory computer-readable medium of Claim 13, wherein the data incident comprises at least one of a data privacy incident due to the functionality processing the target data or a risk of being noncompliant with a data privacy standard due to the functionality processing the target data.

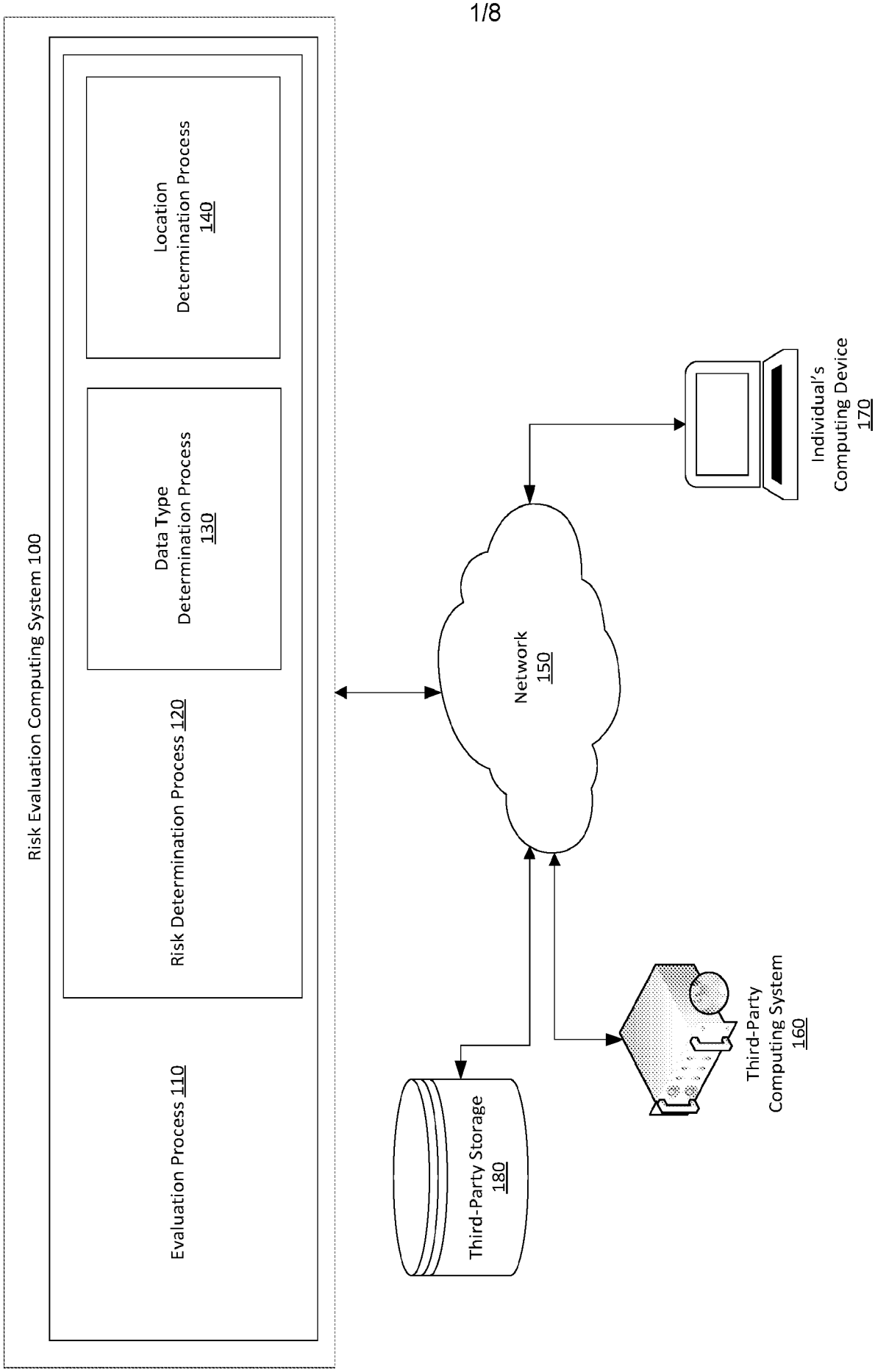
Risk Evaluation Computing System 100
Evaluation Process 110
Risk Determination Process 120
Data Type Determination Process 130
Location Determination Process 140
Network 150
Third-Party Computing System 160
Third-Party Storage 180
Individual's Computing Device 170

FIG. 1

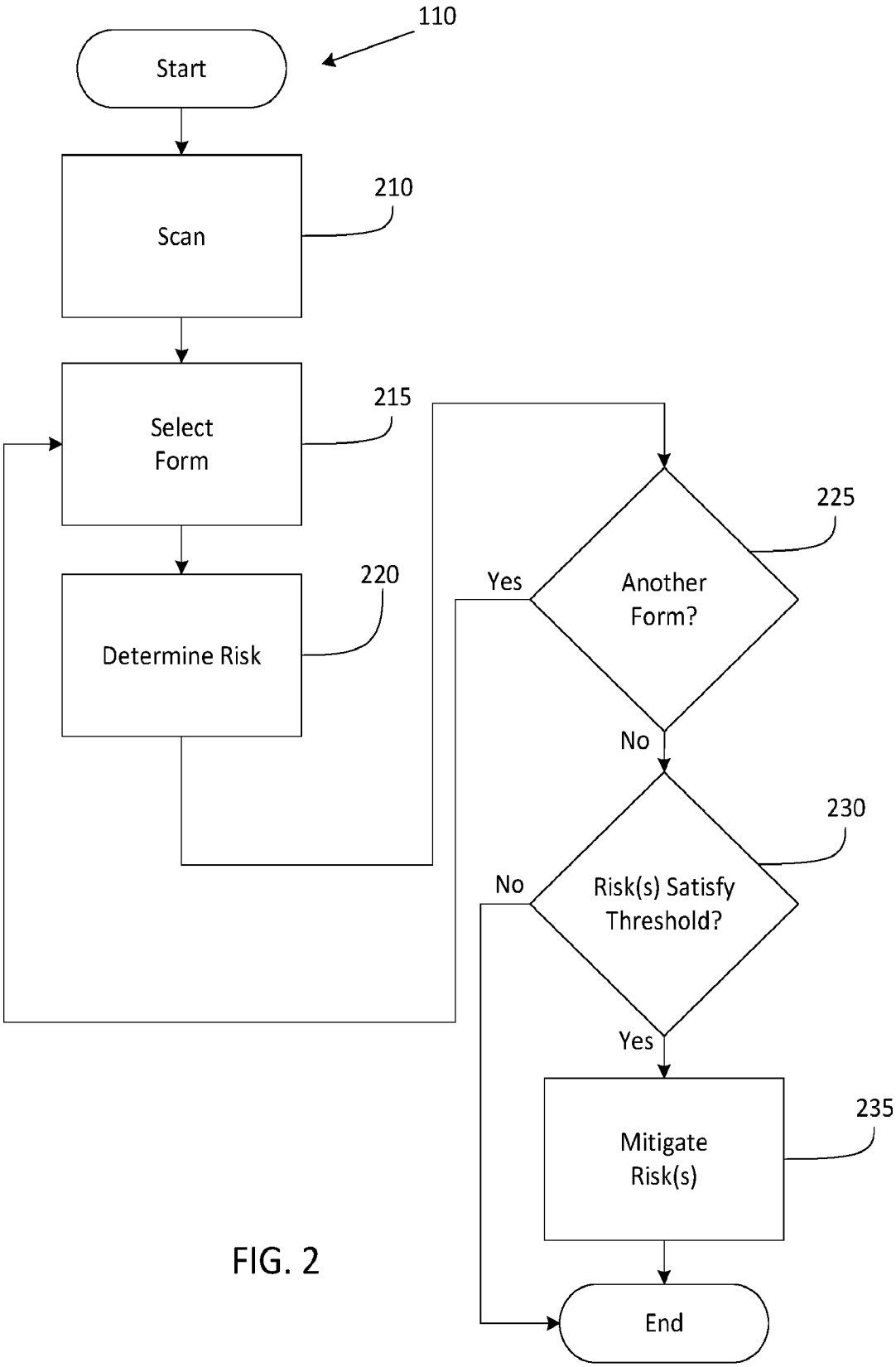
110
Start
Scan
210
Select Form
215
Determine Risk
220
Another Form?
225
Yes
No
Risk(s) Satisfy Threshold?
230
No
Yes
Mitigate Risk(s)
235
End

FIG. 2

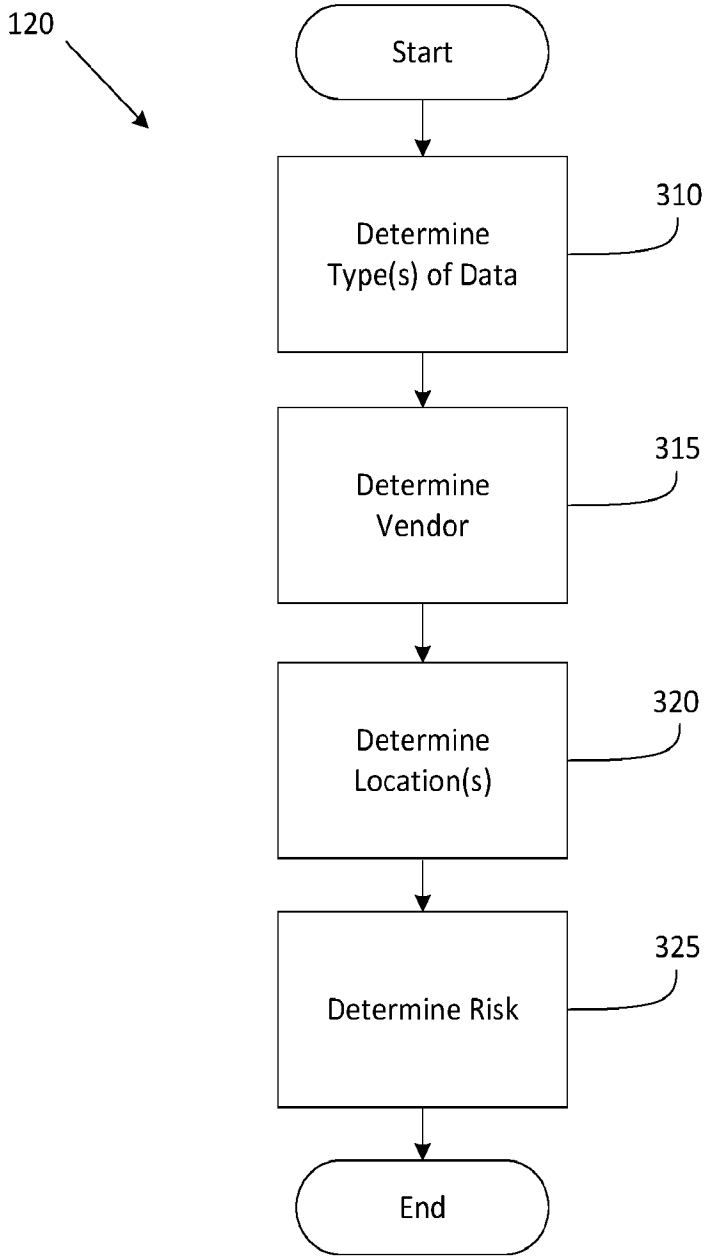
120
Start
Determine Type(s) of Data
310
Determine Vendor
315
Determine Location(s)
320
Determine Risk
325
End

FIG. 3

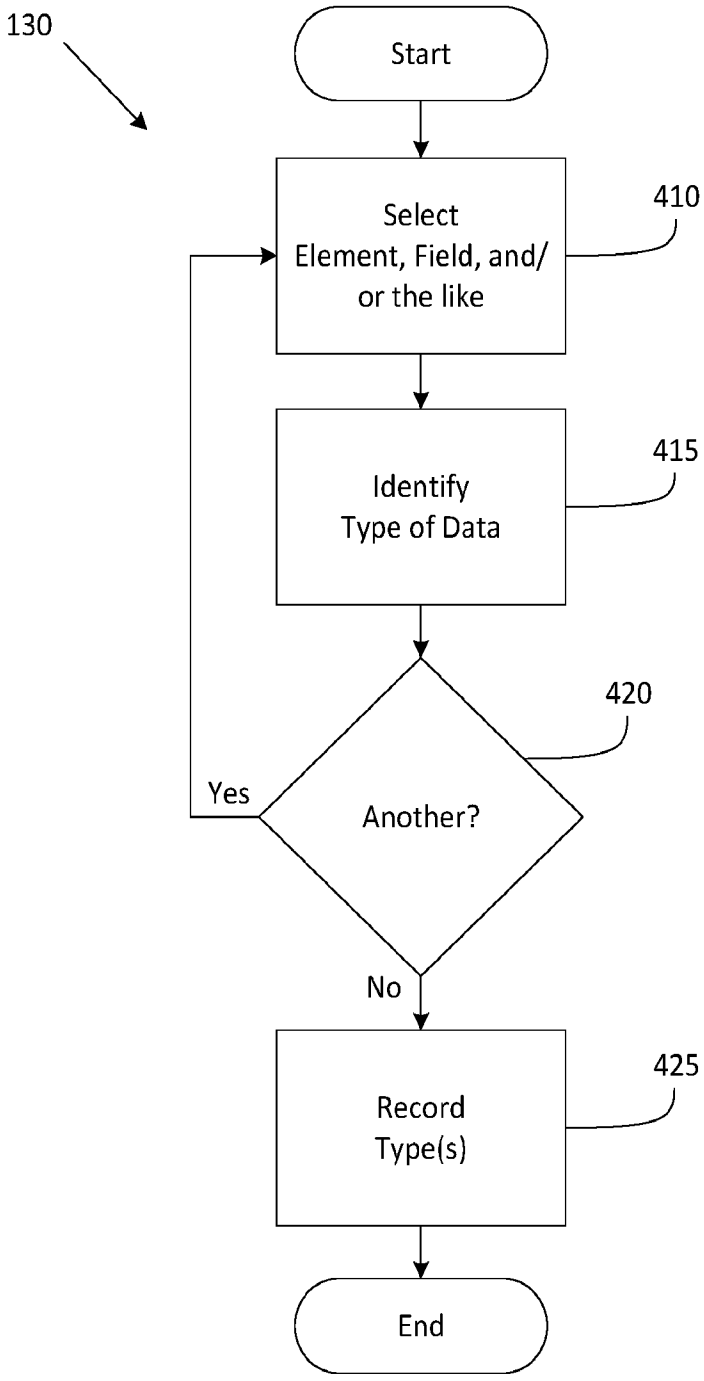
130
Start
Select Element, Field, and/ or the like
410
Identify Type of Data
415
Another?
420
Yes
No
Record Type(s)
425
End

FIG. 4

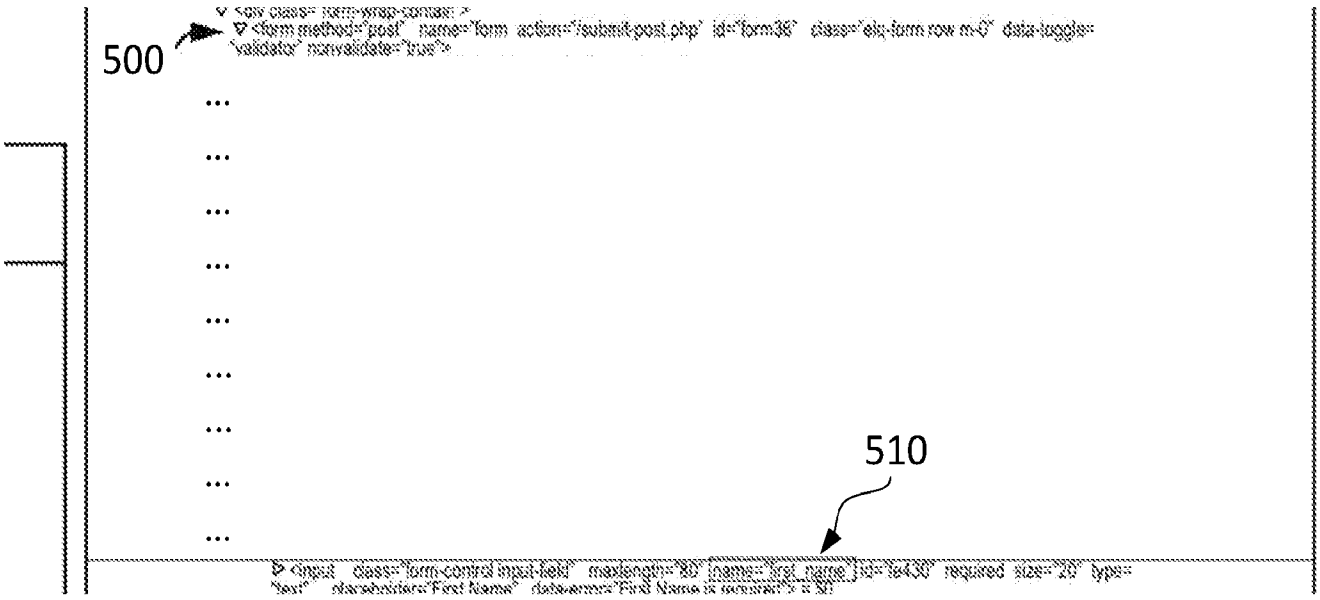


FIG. 5

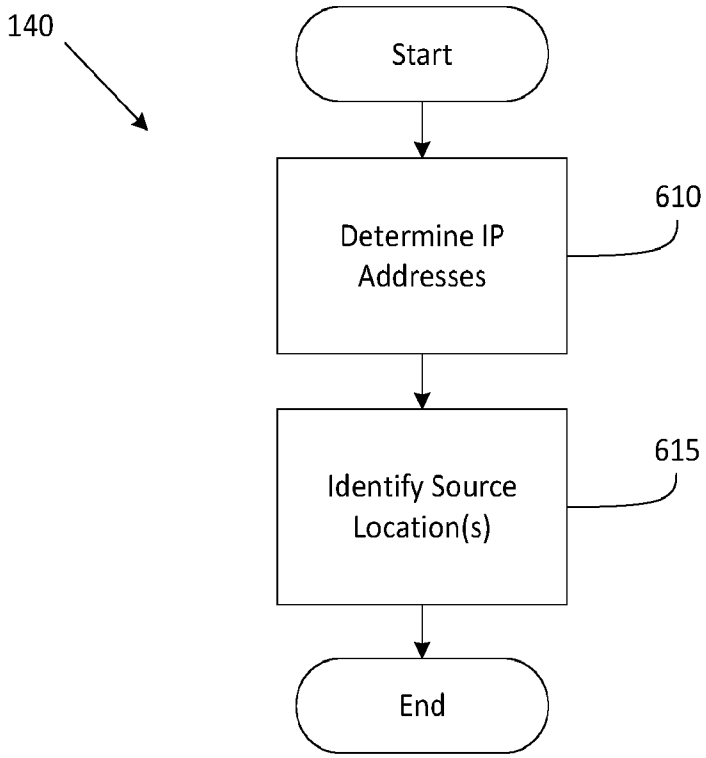
140
Start
Determine IP Addresses
610
Identify Source Location(s)
615
End

FIG. 6

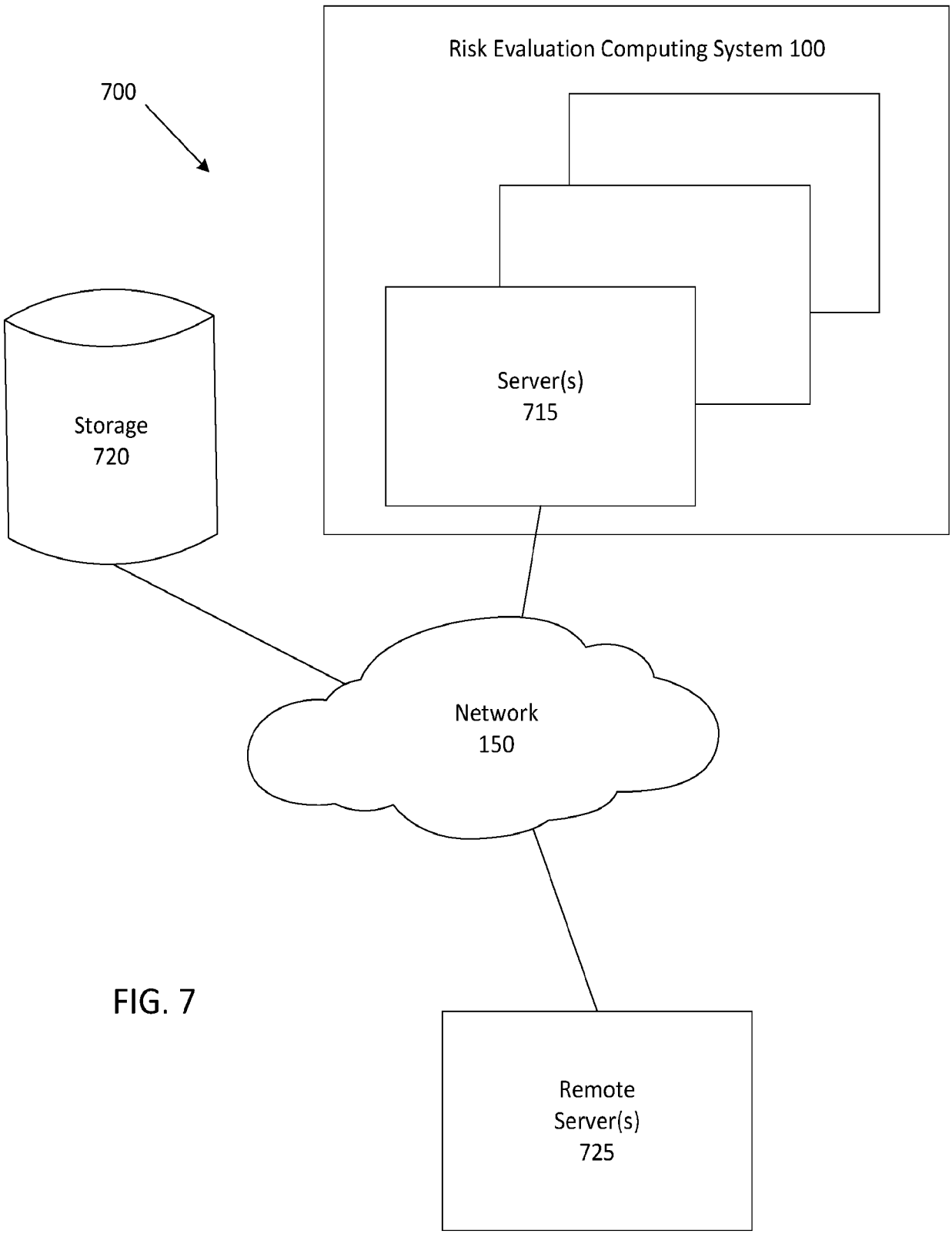
Risk Evaluation Computing System 100
700
Server(s)
715
Storage
720
Network
150
Remote
Server(s)
725

FIG. 7

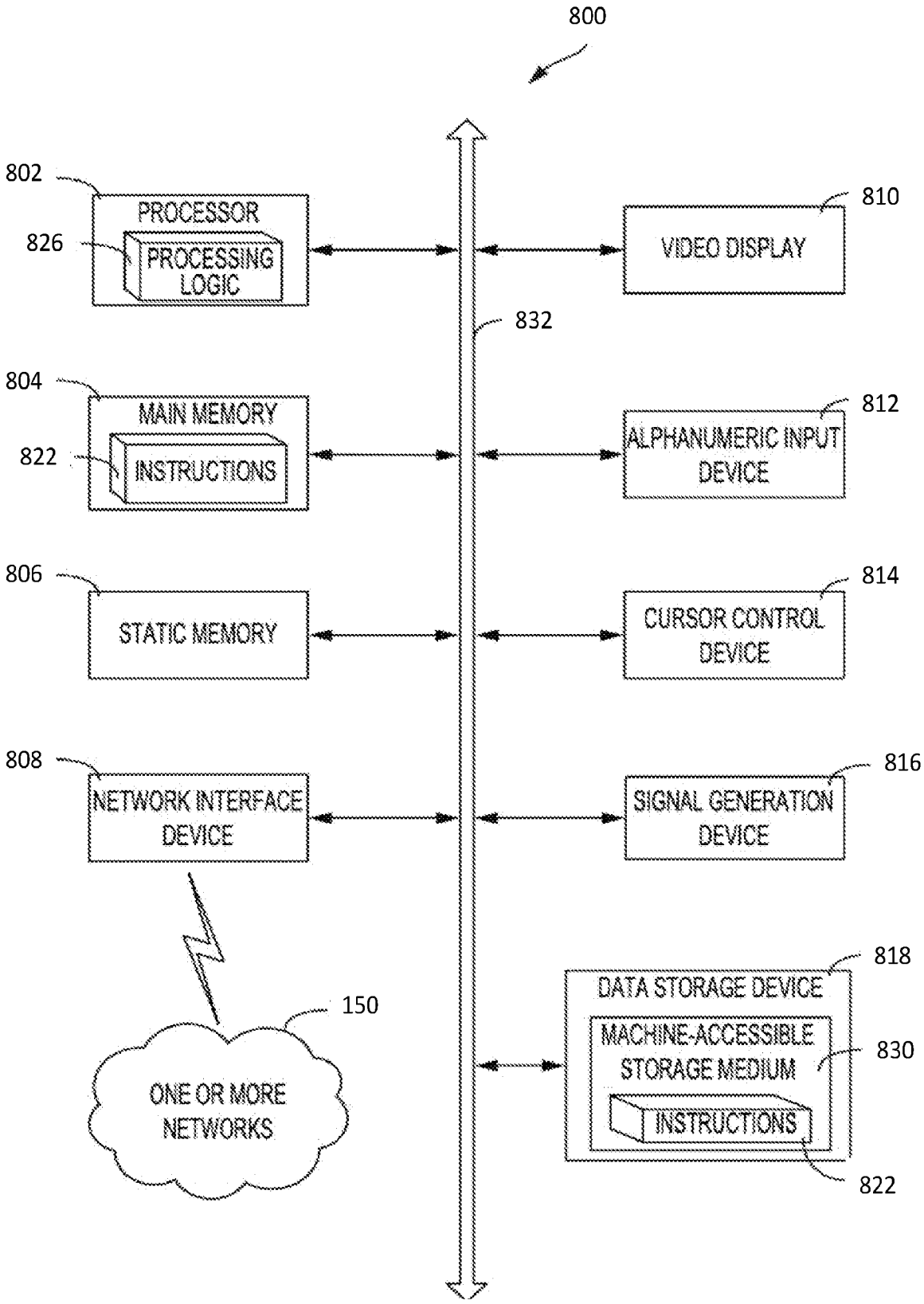
800
802
PROCESSOR
826
PROCESSING LOGIC
804
MAIN MEMORY
822
INSTRUCTIONS
806
STATIC MEMORY
808
NETWORK INTERFACE DEVICE
150
ONE OR MORE NETWORKS
832
810
VIDEO DISPLAY
812
ALPHANUMERIC INPUT DEVICE
814
CURSOR CONTROL DEVICE
816
SIGNAL GENERATION DEVICE
818
DATA STORAGE DEVICE
MACHINE-ACCESSIBLE STORAGE MEDIUM
830
INSTRUCTIONS
822

FIG. 8

# INTERNATIONAL SEARCH REPORT

International application No
PCT/US2021/051217

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F21/55 G06F21/57 G06F21/62 H04L29/06 G06N20/00
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06F H04L G06N

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | US 2019/050596 A1 (BARDAY KABIR A [US] ET AL) 14 February 2019 (2019-02-14) paragraph [0005] - paragraph [0235] | 1-20 |
| X | US 2019/384899 A1 (BRANNON JONATHAN BLAKE [US]) 19 December 2019 (2019-12-19) paragraph [0003] - paragraph [0137] | 1-20 |
| X | US 2020/257782 A1 (BRANNON JONATHAN BLAKE [US] ET AL) 13 August 2020 (2020-08-13) paragraph [0003] - paragraph [0405] | 1-20 |
| X | US 2020/220901 A1 (BARDAY KABIR A [US] ET AL) 9 July 2020 (2020-07-09) paragraph [0003] - paragraph [0261] | 1-20 |

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 3 December 2021 | 22/12/2021 |

| Name and mailing address of the ISA/ | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040,<br>Fax: (+31-70) 340-3016 | Veshi, Erzim |

1

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2021/051217

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| US 2019050596 A1 | 14-02-2019 | US 2019050596 A1 | 14-02-2019 |
| | | US 2019258823 A1 | 22-08-2019 |
| | | US 2020218828 A1 | 09-07-2020 |
| | | US 2021294906 A1 | 23-09-2021 |
| US 2019384899 A1 | 19-12-2019 | US 2019384899 A1 | 19-12-2019 |
| | | US 2021081542 A1 | 18-03-2021 |
| US 2020257782 A1 | 13-08-2020 | NONE | |
| US 2020220901 A1 | 09-07-2020 | NONE | |