

NOTA: LA PRESENTE SOLICITUD FUE TRAMITADA BAJO EL CAPÍTULO II DEL PCT. SE PRESENTARON MODIFICACIONES CONFORME AL ART. 34.

- 5 A CONTINUACIÓN, SE INCLUYE LA TRADUCCIÓN AL ESPAÑOL DE LA DESCRIPCIÓN MODIFICADA CONFORME AL ART. 34 (SEGÚN SE PUEDE CORROBORAR EN LOS ANEXOS DEL INFORME DE CAPÍTULO II) Y LA DESCRIPCIÓN ORIGINAL, A FIN DE CUMPLIR CON TODOS LOS REQUERIMIENTOS.
- 10 AGRADECEMOS TENER EN CUENTA LA DESCRIPCIÓN MODIFICADA PARA EL EXAMEN DE PATENTABILIDAD.

SISTEMA Y MÉTODO IMPLEMENTADO POR COMPUTADOR PARA ESTUDIOS FEDERADOS SEGUROS

15

(MODIFICACIÓN BAJO ART. 34 – A SER TENIDA EN CUENTA PARA EL EXAMEN DE PATENTABILIDAD)

Solicitud relacionada

- 20 [0001] La presente solicitud está relacionada con la solicitud de patente provisional australiana No. 2023900994 titulada “SISTEMA Y MÉTODO IMPLEMENTADO POR COMPUTADOR PARA ESTUDIOS FEDERADOS SEGUROS” y presentada el 5 de abril de 2023 a nombre de Evidentli Pty Ltd, cuyo contenido completo se incorpora aquí por referencia como si se reprodujera íntegramente en este
- 25 documento.

Campo técnico

- [0002] La presente divulgación se refiere a un sistema y métodos asociados para estudios federados confiables y seguros. En particular, se refiere a un sistema informático y métodos relacionados para su uso en estudios federados confiables y
- 30 seguros en el contexto de estudios clínicos, estudios observacionales y otras actividades de investigación.

Antecedentes

[0003] Los estudios federados son ensayos clínicos, estudios observacionales y otros tipos de actividades de investigación que se llevan a cabo en múltiples ubicaciones físicas (centros), como hospitales, laboratorios, universidades, instituciones de investigación, etc. Los centros donde se realizan los estudios se denominan aquí sitios de origen, y el centro donde se analizan los datos se denomina aquí sitio de análisis. Una misma ubicación puede ser tanto sitio de origen como sitio de análisis.

[0004] Las ventajas de los estudios federados incluyen, entre otras:

- 10 • Los participantes en un estudio federado representan una muestra de la población mayor que la que podría representar cualquier centro individual.
- La logística de realizar cada estudio es más sencilla que la de realizar un único estudio de gran envergadura, y
- 15 • La heterogeneidad de los propios sitios representa una muestra más amplia de sitios.

[0005] En un estudio federado, cada sitio de origen realiza de forma independiente su propia versión del estudio, y los datos de todos los centros se analizan conjuntamente en un sitio de análisis para obtener un resultado combinado. Si bien existen varios métodos para realizar estudios federados, cada uno presenta limitaciones que impiden su adopción universal. Por ejemplo, las distintas jurisdicciones pueden definir los diagnósticos de manera diferente o utilizar parámetros o rangos distintos en relación con los diagnósticos o las afecciones.

[0006] Independientemente de los métodos matemáticos empleados para combinar los resultados de múltiples centros en un estudio federado, este tipo de estudio puede presentar problemas de coherencia entre los centros participantes. Las variaciones en la producción y recopilación de datos, así como en los materiales utilizados, pueden generar imprecisiones inexplicables en los resultados. Además,

las restricciones no técnicas, como los requisitos legales jurisdiccionales en materia de privacidad y transmisión de datos, pueden afectar la recopilación de datos.

[0007] Por lo tanto, existe la necesidad de proporcionar un sistema mejorado para la realización de estudios federados.

5 **Resumen**

[0008] La presente divulgación se refiere a un sistema implementado por computador y a los métodos asociados para realizar estudios federados.

[0009] Un primer aspecto de la presente divulgación proporciona un sistema de estudios federados implementado por computador que comprende:

- 10 un modelo de datos común predefinido asociado a un estudio;
 un conjunto de sitios de origen, estando cada sitio de origen asociado a un conjunto de dispositivos de entrada para la adquisición de datos de estudio de los sujetos y un dispositivo de datos de estudio, donde cada dispositivo de datos de estudio incluye:
 - 15 una base de datos nativa para almacenar datos de estudio asociados con el sitio de origen respectivo y derivados de dicho conjunto de dispositivos de entrada;
 - un traductor de modelos de datos para traducir los datos de estudio almacenados a un formato que corresponda a dicho modelo de datos común;
 - 20 una base de datos de modelo común para almacenar, en dicho modelo de datos común, los valores individuales de los participantes derivados de los datos del estudio;
 - un módulo de cálculo de estudio para determinar los valores agregados del sitio a partir de los valores individuales de los participantes; y
 - 25 una base de datos agregada para almacenar los valores agregados de dicho sitio; y

un sitio de análisis conectado a cada sitio de origen mediante una red de comunicaciones, donde dicho sitio de análisis está asociado a un dispositivo de análisis de estudio que incluye:

un módulo de cálculo de análisis para procesar, de acuerdo con un perfil de análisis predefinido, los valores agregados del sitio recibidos de dichos sitios de origen; y

un módulo de informes para generar informes de estudio basados en los valores agregados del sitio procesados;

un repositorio central para almacenar y recuperar un conjunto de elementos de traducción de datos para gestionar extensiones y mapeos dentro de los conjuntos de datos de estudio asociados con cada sitio de origen; y

un Servicio de Emisión de Identificadores configurado para servir de interfaz entre cada sitio de origen y dicho repositorio central para garantizar que los elementos dentro de los conjuntos de datos que tengan el mismo significado semántico tengan un identificador único común.

[0010] Un segundo aspecto de la presente divulgación proporciona un método para llevar a cabo un estudio federado en varios sitios de origen y utilizando un sitio de análisis, dicho método comprende los siguientes pasos:

definir un modelo de datos común para el estudio;

definir un conjunto de instrucciones de análisis para el estudio;

en cada sitio de origen:

adquirir, utilizando al menos un dispositivo de entrada, datos de estudio de un conjunto de sujetos;

traducir dichos datos de estudio para generar valores individuales de los participantes almacenados en dicho modelo de datos común;

procesar dichos valores individuales de los participantes para generar un conjunto de valores agregados del sitio;

transmitir dichos valores agregados del sitio al sitio de análisis; y

procesar los valores agregados del estudio recibidos del sitio de análisis de acuerdo con dicho conjunto definido de instrucciones de análisis; y en dicho sitio de análisis:

5 procesar los valores agregados del sitio recibidos de al menos uno de dichos sitios de origen de acuerdo con dicho conjunto definido de instrucciones de análisis; y

generar un informe basado en los valores agregados del sitio procesados;

10 en donde el método unifica las traducciones de datos en los diversos Sitios de Origen mediante:

almacenar en un repositorio central de un conjunto de elementos de traducción de datos para gestionar extensiones y mapeos dentro del estudio federado;

15 al recibir una solicitud de una nueva extensión, un Servicio Emisor de Identificadores se comunica con dicho repositorio central para determinar si la extensión solicitada existe;

cuando la extensión solicitada existe, el Servicio Emisor de Identificadores la recupera del repositorio central y la envía al cliente de traducción solicitante; y

cuando la extensión solicitada no existe, el Servicio Emisor de Identificadores:

20 emite un nuevo identificador único para la extensión solicitada, envía el nuevo identificador al cliente de traducción solicitante y lo envía al repositorio central para su almacenamiento.

[0011] Según otro aspecto, la presente divulgación proporciona un aparato para implementar cualquiera de los métodos anteriormente mencionados.

25 [0012] Según otro aspecto, la presente divulgación proporciona un producto de programa informático que incluye un medio legible por computador que tiene grabado un programa informático que, cuando se ejecuta en un procesador de computador, implementa cualquiera de los métodos descritos anteriormente.

[0013] También se proporcionan otros aspectos de la presente divulgación.

Breve descripción de las figuras

[0014] A continuación se describirán una o más modalidades de la presente divulgación mediante ejemplos específicos con referencia a las figuras adjuntas, en las que:

5 [0015] La figura 1 es una representación esquemática mediante diagrama de bloques de un sistema de estudio federado implementado por computador, de acuerdo con una modalidad de la presente divulgación;

[0016] La figura 2 es una representación esquemática mediante diagrama de bloques de un sitio de análisis de acuerdo con una modalidad de la presente
10 divulgación;

[0017] La figura 3 es una representación esquemática en diagrama de bloques de un Sitio de Origen de acuerdo con una modalidad de la presente divulgación;

[0018] La figura 4 es una representación esquemática mediante diagrama de bloques de un sistema de estudios federados en el que se pueden poner en práctica
15 una o más modalidades de la presente descripción;

[0019] La figura 5 es una representación esquemática mediante diagrama de bloques de un sistema que incluye un computador de propósito general en el cual se pueden poner en práctica una o más modalidades de la presente descripción;

[0020] La figura 6 es una representación esquemática mediante diagrama de
20 bloques de un sistema que incluye un teléfono inteligente general en el que se pueden poner en práctica una o más modalidades de la presente descripción;

[0021] La figura 7 es un diagrama de secuencia que ilustra un flujo de trabajo de práctica ejemplar de un estudio federado realizado por una modalidad del sistema de la presente divulgación;

25 [0022] La figura 8 es una representación esquemática mediante diagrama de bloques de un sitio de origen implementado utilizando una arquitectura informática basada en la nube;

[0023] La figura 9 es una representación esquemática mediante diagrama de bloques de un sistema de estudios federados implementado utilizando una arquitectura informática basada en la nube;

[0024] La figura 10 es una representación esquemática mediante diagrama de bloques de un sistema de estudios federados implementado utilizando una arquitectura informática basada en la nube;

[0025] La figura 11 es una representación esquemática mediante diagrama de bloques de un sistema de estudios federados implementado utilizando una arquitectura informática basada en la nube; y

[0026] La figura 12 es una representación esquemática en diagrama de bloques que ilustra una porción de una red de estudios federados que cuenta con un sitio de origen, un servicio emisor de identificadores y un repositorio central.

[0027] Los pasos o características del método en las figuras adjuntas que tengan los mismos números de referencia se considerarán que tienen las mismas funciones u operaciones, a menos que se exprese o se dé a entender la intención contraria.

Glosario

Sitio de Origen

[0028] Un sitio que almacena datos sensibles sobre personas que se utilizarán en un proyecto de investigación sin transmitirlos fuera del sitio. Esta descripción se refiere a múltiples sitios de origen, denominados aquí *Sitio₁*, *Sitio₂*, etc., hasta el *Sitio_s*. En general, cada sitio se denomina *Sitio_i*.

Sitio de Análisis

[0029] El sitio de análisis se denomina en el presente documento como *Sitio_A*. El Sitio de Análisis puede coincidir (estar ubicado en el mismo lugar o ser parte integral de) uno de los sitios de origen o puede ser una entidad separada e independiente.

Valor para el Participante Individual: Y_j

[0030] Uno o más valores numéricos sobre una persona (persona j), que potencialmente podrían usarse para identificar o exponer información sensible sobre la persona j .

5 *Valor Agregado: Y*

[0031] Un elemento de datos que contiene información sobre un grupo de personas y que no puede utilizarse para identificar ni revelar información confidencial sobre una o varias personas. Los valores agregados se pueden transmitir de forma segura entre sitios. Ejemplos de valores agregados son el recuento de participantes y los promedios de los valores individuales de los participantes. Según el escenario y la implementación específicos, los valores agregados pueden intercambiarse entre sitios varias veces durante un mismo estudio federado.

Agregado del Sitio: $\hat{Y}_i$

15 [0032] Un valor agregado que se refiere a la información disponible para un sitio de origen. El símbolo de un agregado de sitio va seguido del subíndice i , por ejemplo, x_i . Un ejemplo de agregado de sitio es el número de participantes en el sitio (n_i).

Estudio Agregado: $\hat{Y}$

20 [0033] Valor agregado que engloba la información disponible para un estudio completo. El símbolo de un valor agregado del estudio no va seguido de un subíndice, por ejemplo, x_i . Un ejemplo de valor agregado del estudio es el número de participantes en el estudio completo (N).

Tarea

[0034] Un agregado de estudio que $Sitio_A$ tiene la tarea de calcular (es decir, una fórmula que necesita datos del sitio para ser calculada).

25 *Número de Participantes en el Estudio: N*

[0035] Un agregado de estudio especial que representa el número de participantes en todo el estudio.

Número de Participantes en el Sitio: n_i

[0036] Un agregado especial del sitio que es el número de participantes en *el sitio*_{*i*}.

Número de Sitios: s

[0037] El número total de sitios de origen que participan en un estudio.

5 Descripción detallada

[0038] La presente descripción proporciona un sistema implementado por computador y los métodos asociados para realizar estudios federados. El sistema incluye un conjunto de sitios de origen y un conjunto de sitios de análisis. En algunas implementaciones, los sitios de análisis están separados e independientes de los sitios de origen, estando cada sitio de origen conectado a un sitio de análisis mediante un enlace de comunicaciones que permite la transmisión de datos desde cada sitio de origen al sitio de análisis correspondiente. En otras implementaciones, uno o más sitios de análisis están ubicados en el mismo lugar o integrados en un sitio de origen, y uno o más enlaces de comunicaciones permiten la transmisión de datos entre los sitios de origen y un sitio de análisis en los casos en que un sitio de análisis no esté ubicado en el mismo lugar o integrado en un sitio de origen.

[0039] Una configuración básica utiliza un único sitio de análisis. Otras configuraciones pueden utilizar varios sitios de análisis. Por ejemplo, una configuración se refiere a un escenario de red con subredes jerárquicas de sitios de origen y sitios de análisis, donde estos últimos envían datos a un sitio de análisis central. Esta disposición jerárquica proporciona un conjunto en cascada de sitios de origen y sitios de análisis.

[0040] Una posible implementación podría relacionarse, por ejemplo, con un sitio de análisis centralizado administrado por una empresa farmacéutica. Las subredes se relacionan con los distintos distritos sanitarios que administran sitios de análisis distritales, y cada sitio de análisis distrital recibe datos de los hospitales asociados a su distrito sanitario. En dicha implementación, los sitios de análisis distritales se convierten en sitios de origen para el sitio de análisis centralizado.

[0041] Para mayor claridad, las modalidades descritas aquí se referirán generalmente a un único nivel jerárquico en el que un conjunto de sitios de origen informa a un único sitio de análisis. No obstante, un experto en la materia comprenderá que el alcance de la invención aquí descrita abarca modalidades
5 alternativas con múltiples niveles de análisis.

[0042] Para mayor claridad, las modalidades descritas aquí presuponen una configuración de conexiones de red simples. Sin embargo, un experto en la materia comprenderá que el alcance de la invención aquí descrita abarca otras configuraciones de redes que pueden incluir, por ejemplo, transmisión (por ejemplo,
10 enrutadores), seguridad (por ejemplo, cortafuegos) y aceleración (por ejemplo, servidores caché), entre otras.

[0043] Los enfoques actuales para estudios federados presentan la dificultad técnica de recopilar datos de manera consistente en diferentes sitios de origen, de modo que estos puedan procesarse de forma uniforme. Además, existen
15 dificultades técnicas para garantizar el anonimato de los datos personales recopilados en dichos sitios.

[0044] El sistema descrito en la presente divulgación garantiza la compatibilidad de datos entre los sitios de origen para asegurar la fiabilidad de los resultados. En particular, el sistema incluye uno o más pasos que transfieren datos de múltiples
20 fuentes para su agregación y estandarización conforme a un modelo de datos común predefinido. En algunas implementaciones, la transferencia de datos puede incluir traducción, cuando sea necesario. En algunas implementaciones, el sistema utiliza una entrada de datos inteligente que permite agregar y estandarizar datos de múltiples fuentes conforme a un modelo de datos común predefinido. Un modelo de
25 datos se ajusta al modelo de datos común predefinido si dicho modelo puede ser utilizado por un sitio de origen para generar un agregado de datos comprensible para el sitio de análisis.

[0045] En algunas implementaciones, los sitios de origen y el sitio de análisis comparten un modelo de datos común coordinado. Este modelo puede ser predefinido por el administrador del estudio o bien coordinado de otra manera entre los distintos sitios de origen y el sitio de análisis. En dichas implementaciones, los datos transmitidos desde un sitio de origen al sitio de análisis han sido preparados por el sitio de origen correspondiente para ajustarse al modelo de datos común coordinado. En otras implementaciones, cada sitio de origen utiliza un modelo de datos, que puede o no ajustarse al modelo de datos común coordinado, y el sitio de análisis dispone de un servicio de traducción que puede convertir, cuando sea necesario, la información agregada de cada sitio de origen a un modelo de datos común.

[0046] En algunas implementaciones, el modelo de datos común es el Modelo de Datos Comunes de la Asociación de Resultados Médicos Observacionales (OMOP CDM u OMOP para abreviar). También se pueden utilizar Oracle Health Foundations, National Patient-Centered Clinical Research Network (PCORNet), Sentinel e Informatics for Integrating Biology & the Bedside (I2B2), o un modelo de datos definido por el usuario.

[0047] En algunas implementaciones, el sistema proporciona gobernanza de datos que aplica y audita los permisos de acceso a los datos. El sistema no transmite datos relativos a participantes individuales, lo que resuelve las preocupaciones legales y de privacidad.

[0048] En algunas implementaciones, el sistema proporciona una interfaz de usuario en forma de panel de control que ofrece a los usuarios análisis clínicos de autoservicio, lo que permite a usuarios con diferentes niveles de conocimientos técnicos consultar datos de investigación y generar informes.

Agrupación de Datos

[0049] La recopilación de datos de los centros donde se registran y su transmisión a un centro donde se analizarán se denomina “Agrupación de Datos”. La agrupación

de datos se considera el estándar actual para obtener los resultados más fiables de estudios federados. Sin embargo, la transmisión de datos individuales de participantes entre centros supone un riesgo para la privacidad de los pacientes y puede estar sujeta a controles legales.

- 5 [0050] Cabe destacar que la limitación no se restringe a la protección criptográfica de los datos en tránsito o en reposo, ni a la eliminación fiable de la información de identificación personal. El problema radica en que la mayoría de las jurisdicciones tienen restricciones legales que prohíben la transmisión y el almacenamiento de datos, y estas restricciones no técnicas son imposibles de superar únicamente con
- 10 tecnologías como el cifrado. En consecuencia, la agrupación de datos suele limitarse a un único sitio, lo que puede impedir obtener tamaños de muestra estadísticamente útiles.

Metaanálisis

- [0051] El metaanálisis es un método que evita el envío de datos individuales de los
- 15 participantes y reduce el riesgo para su privacidad al transmitir únicamente información agregada calculada en cada centro. Estos datos agregados se combinan en un centro común que genera el resultado del estudio. Sin embargo, para la mayoría de los valores, estos cálculos solo pueden aproximarse a los resultados que se obtendrían mediante la agrupación de datos.
- 20 [0052] En las técnicas de metaanálisis se utilizan diversas técnicas de aproximación para combinar datos de sitios web. Cada técnica de aproximación se basa en ciertas suposiciones, que a menudo no se especifican explícitamente. No incluir los supuestos tiene varias consecuencias negativas, entre ellas: (i) dificultar la comprensión del informe; (ii) no transmitir las intenciones del autor; y/o (iii) obligar
- 25 al lector a inferir dichos supuestos a partir de los métodos empleados, lo cual puede dar lugar a errores. No indicar los supuestos asociados a las técnicas de aproximación puede llevar al lector a interpretar las aproximaciones como valores exactos, lo cual es erróneo.

Agregados de Datos

[0053] Algunos métodos para realizar estudios federados utilizan agregados de datos de diferentes centros de estudio. Ejemplos de datos agregados son el número de participantes fallecidos durante el periodo de estudio, la mediana de la duración de la estancia de una cohorte de participantes y la reducción del riesgo absoluto entre dos grupos de estudio. Los datos agregados son fáciles de calcular en cada centro donde se almacenan y se transmiten desde los centros con los datos originales (sitios de origen), en lugar de los datos sensibles, al centro remoto donde se realiza el análisis (sitio de análisis).

[0054] Algunos datos agregados pueden calcularse con exactitud en el lugar del análisis, obteniéndose resultados idénticos a los que se habrían obtenido mediante la agrupación de datos. Por ejemplo, para conocer el número total de pacientes del estudio, se puede sumar el número de pacientes de cada centro.

[0055] Otros tipos de datos agregados no se pueden calcular con exactitud a partir de otros agregados de datos. Por ejemplo, para calcular con exactitud la reducción del riesgo absoluto entre dos grupos experimentales, se calcula el promedio de cada centro y se envía al sitio de análisis. Para calcular la reducción del riesgo absoluto promedio en toda la población del estudio, es necesario realizar ciertas suposiciones. Por ejemplo, la reducción del riesgo absoluto para cada centro debe ponderarse según el número de participantes en cada centro o, más comúnmente, según la homogeneidad de los pacientes en cada centro. Con estas suposiciones, la reducción del riesgo absoluto calculada es solo una aproximación del valor que se habría obtenido mediante la agrupación de datos.

Sistema Implementado por Computador y Método Asociado para la Realización de Estudios Federados

[0056] Independientemente de los métodos matemáticos empleados para combinar los resultados de múltiples centros en un estudio federado, los métodos existentes para realizar dichos estudios pueden presentar problemas de

consistencia entre los sitios de estudio debido a las suposiciones generales que se realizan para correlacionar los datos. La presente descripción proporciona un sistema implementado por computador que permite realizar estudios federados en un conjunto de sitios de estudio sin necesidad de recurrir a suposiciones generales.

5 [0057] La figura 1 es un diagrama de bloques esquemático de un sistema implementado por computador para estudios federados 100, de acuerdo con una modalidad de la presente descripción. El sistema 100 incluye un conjunto 105 de Sitios de Origen 110a a 110s. Cada Sitio de Origen 110a a 110s es una ubicación donde se adquieren datos para un estudio e incluye un dispositivo informático
10 asociado conectado a una red de comunicaciones que permite la comunicación entre cada Sitio de Origen 110a a 110s y un Sitio de Análisis 150. La red de comunicaciones puede constar de uno o más enlaces de comunicaciones cableados, enlaces de comunicaciones inalámbricos o cualquier combinación de estos. En particular, la red de comunicaciones puede incluir una red de área local
15 (LAN), una red de área extensa (WAN), una red de telecomunicaciones o cualquier combinación de estas. Una red de telecomunicaciones puede incluir, entre otras cosas, una red de telefonía, como una red telefónica pública conmutada (RTPC) o una red de telefonía móvil celular, Internet o cualquier combinación de las mismas.
[0058] El Sitio de Análisis 150 es una ubicación central para analizar los datos
20 derivados de estudios realizados en uno o más de los Sitios de Origen 110a a 110s. El Sitio de Análisis 150 está conectado a un Repositorio de Análisis 120, el cual es una base de datos que almacena diversos análisis que pueden ser realizados y una forma de análisis se selecciona con base en el estudio que se esté realizando. Cada análisis consiste en un conjunto de instrucciones informáticas que especifican cómo
25 realizar un análisis de datos en particular, y cada estudio se asocia con uno de estos análisis. Estas instrucciones informáticas para realizar análisis de datos pueden representarse en un programa informático, código fuente u otro documento o código electrónico que un computador puede interpretar como una serie de operaciones

que, al ejecutarse sobre un conjunto de datos del estudio, generan uno o más Valores Agregados significativos.

[0059] El Sitio de Análisis 150 puede obtener instrucciones de análisis de datos para un conjunto de datos determinado de diversas maneras. En algunas implementaciones, el Repositorio de Análisis 120 mantiene un enlace entre el Sitio de Análisis 150 y los Sitios de Origen 110a a 110s. En otras implementaciones, las instrucciones para el Sitio de Análisis 150 incluyen una referencia a las instrucciones para los Sitios de Origen 110a a 110s.

[0060] En otras modalidades, el sistema 100 incluye un administrador de análisis (no mostrado) que coordina las instrucciones de análisis entre los Sitios de Origen 110a ... 110s y el Sitio de Análisis 150. El administrador de análisis puede implementarse utilizando un dispositivo informático programado o un operador humano o una combinación de ambos.

[0061] En otras modalidades, se utilizan otros métodos de sincronización, como el Acuerdo Bizantino, para garantizar que el Sitio de Análisis 150 utilice un conjunto correcto de instrucciones de análisis del Repositorio de Análisis 120.

[0062] La figura 2 muestra una vista ampliada de los módulos funcionales de una modalidad del Sitio de Análisis 150. El Sitio de Análisis 150 incluye un dispositivo de análisis de datos 151 implementado por computador que está configurado para:

- Comunicarse con cada base de datos de valores agregados del sitio en cada Sitio de Origen
- Interpretar las instrucciones representadas en un análisis del Repositorio de Análisis en un algoritmo
- Ejecutar un algoritmo interpretado a partir de un análisis del Repositorio de Análisis
- Reportar los resultados del estudio

[0063] En el contexto de la figura 2, el dispositivo de análisis de datos 151 recibe la comunicación 205 de uno o más de los Sitios de Origen 110a a 110s o del

Repositorio de Análisis 120. Las comunicaciones entre el dispositivo de análisis de datos 151 y el Repositorio de Análisis 120 pueden estar dirigidas a transmitir un conjunto de instrucciones de análisis desde el Repositorio de Análisis 120 al dispositivo de análisis de datos 151. Dichas instrucciones de análisis pueden
5 entonces transmitirse desde el dispositivo de análisis de datos 151 a uno o más de los Sitios de Origen 110a a 110s.

[0064] En algunas modalidades, la comunicación 205 se refiere a un enlace proporcionado por, o recuperado de, el Repositorio de Análisis 120 y enviado al dispositivo de análisis de datos 151 para reenviarlo a uno o más de los Sitios de
10 Origen 110a ... 110s, de modo que el Sitio de Origen 110a ... 110s respectivo pueda utilizar el enlace para recuperar un conjunto apropiado de instrucciones de análisis directamente del Repositorio de Análisis 120.

[0065] La comunicación 205 también puede referirse al intercambio de información entre el dispositivo de análisis de datos 151 y uno o más de los Sitios de Origen
15 110a ... 110s, en cuyo caso la información puede incluir, por ejemplo, solicitudes de datos agregados o la transmisión de datos agregados relacionados con un estudio.

[0066] La comunicación recibida 205 es procesada por un módulo de cálculo 152 que interpreta las instrucciones recibidas del Repositorio de Análisis y aplica esas instrucciones para analizar los datos recibidos de uno o más Sitios de Origen 110a
20 ... 110s. Los datos analizados se presentan luego a un módulo de informes 154 que informa los resultados del estudio.

[0067] Según la implementación, el módulo de informes 154 presenta los resultados del estudio a través de una interfaz de usuario (como un panel de control), imprimiéndolos, enviándolos por correo electrónico, mostrándolos en un
25 sitio web, generando un documento electrónico o mediante cualquier combinación de estos métodos. El sistema es configurable para seleccionar un conjunto de usuarios a quienes se envían los resultados del estudio o que están autorizados a acceder a ellos. En algunas implementaciones, el nivel de autorización asignado a

un usuario determina el nivel de detalle de los resultados del estudio a los que tiene acceso. Por ejemplo, los usuarios con alto nivel de autorización pueden tener acceso a todos los niveles de los resultados del estudio, mientras que los usuarios con bajo nivel de autorización pueden tener acceso solo a un nivel general de los resultados o a un subconjunto de los mismos.

[0068] La figura 3 es un diagrama de bloques esquemático del Sitio de Origen 110a de la figura 1 y representa cualquiera de los Sitios de Origen 110a a 110s. El Sitio de Origen 110a incluye un dispositivo de datos de estudio computarizado 115a conectado a un conjunto de dispositivos de entrada 111a a 111d para la adquisición de datos de estudio de uno o más sujetos. Los dispositivos de entrada 111a a 111d pueden implementarse utilizando cualquier dispositivo adecuado para la captura digital de datos que se puedan usar en un estudio federado. Los dispositivos de entrada 111a a 111d pueden incluir, por ejemplo, sistemas de registros médicos electrónicos, sistemas de administración de pacientes, sistemas computarizados de prescripción médica, sistemas de encuestas, dispositivos portátiles, sensores de laboratorio, equipos de imagen, termómetros electrónicos, etc. Los dispositivos portátiles pueden incluir, por ejemplo, pero no se limitan a, monitores de frecuencia cardíaca, monitores de presión arterial, monitores de electrocardiograma (ECG), auriculares de electroencefalografía (EEG), oxímetros de pulso y similares.

[0069] Para los estudios federados relacionados con la atención médica, los dispositivos de entrada pueden incluir cualquier cosa que genere o capture datos en un entorno sanitario, y dichos dispositivos no se limitan a la captura de datos clínicos, sino que también pueden incluir dispositivos que capturen, adquieran, almacenen o transmitan datos no clínicos, como detalles de seguros, datos bancarios, datos meteorológicos relativos a un sitio de origen (por ejemplo, temperatura, presión barométrica, humedad), datos geográficos relativos a un sitio de origen (por ejemplo, altitud), etc.

[0070] Según la naturaleza de los dispositivos de entrada 111a a 111d, estos pueden requerir la introducción manual de datos por parte del usuario, estar conectados al dispositivo de datos de estudio 115a (implementado por computador) o utilizar una combinación de introducción y/o transferencia de datos manual y automatizada. Por ejemplo, los monitores de EEG o ECG pueden configurarse para transmitir datos directamente al dispositivo de datos de estudio 115a (implementado por computador) a medida que se obtienen datos del paciente, mientras que los sistemas de administración de pacientes y las historias clínicas electrónicas probablemente requieran al menos cierta introducción manual de datos antes de su transmisión al dispositivo de datos de estudio 115a.

[0071] El dispositivo de datos de estudio 115a incluye una base de datos nativa 112 que almacena los valores individuales de los participantes, obtenidos de uno o más de los dispositivos de entrada 111a a 111d asociados al sitio de origen correspondiente, que en el ejemplo de la figura 3 es el sitio de origen 110a. En dicho ejemplo, la base de datos nativa 112 está integrada en el dispositivo de datos de estudio 115a. En otras implementaciones, la base de datos nativa 112 puede estar integrada en el sitio de origen correspondiente, ubicada en el mismo sitio o ubicada remotamente y conectada al sitio de origen, por ejemplo, mediante una plataforma en la nube. Los valores individuales de los participantes para cada sitio de origen pueden almacenarse en cualquier modelo de datos apropiado, según la implementación. Así mismo, los valores individuales de los participantes para diferentes estudios realizados en un mismo sitio de origen pueden almacenarse en el mismo modelo de datos o en modelos diferentes.

[0072] Un único estudio puede involucrar uno o más conjuntos de datos diferentes. Cada conjunto de datos puede estar asociado a un modelo de datos distinto. Alternativamente, uno o más conjuntos de datos pueden compartir un modelo de datos. En este sentido, se puede proporcionar un traductor para convertir los diferentes conjuntos de datos a un modelo de datos común. Volviendo a la Fig. 3,

los Valores Individuales de los Participantes se capturan desde uno o más de los dispositivos de entrada 111a ... 111b y luego se almacenan en la Base de Datos Nativa 112. Un traductor 114 convierte los Valores Individuales de los Participantes desde el/los modelo(s) de datos utilizados en la(s) Base(s) de Datos Nativa(s) a un

5 Modelo de Datos Común, donde el Modelo de Datos Común es un modelo de datos predefinido común a todos los Sitios de Origen 110a ... 110s y al Sitio de Análisis 150. Los valores traducidos se almacenan, en el formato del Modelo de Datos Común, en una Base de Datos de Valores Individuales de los Participantes en un Modelo de Datos Común 116.

10 [0073] Un módulo de cálculo de origen 118 interpreta las instrucciones representadas en un análisis del Repositorio de Análisis y las convierte en un algoritmo. A continuación, el módulo de cálculo de origen 118 ejecuta el algoritmo interpretado en relación con la Base de Datos de Valores Individuales de Participantes en un Modelo de Datos Común 116 y escribe el resultado del algoritmo

15 ejecutado en una Base de Datos de Valores Agregados del Sitio 119. Esta Base de Datos de Valores Agregados del Sitio se comunica, a través de una red de comunicaciones, con el dispositivo de análisis de datos 151 del Sitio de Análisis 150.

[0074] Durante el transcurso de un estudio, los valores agregados pueden intercambiarse repetidamente entre uno o más de los sitios de origen 110a...s y el

20 sitio de análisis 150, a medida que se calculan. La naturaleza de los valores agregados que se envían y el momento de su envío se definen y coordinan mediante conjuntos de instrucciones de envío y recepción incluidas en las instrucciones de análisis. Dichas instrucciones de análisis pueden estar contenidas en un perfil de análisis predefinido que forma parte de las instrucciones de análisis definidas para

25 un estudio en particular.

Modelo de Datos Común

[0075] Se supone que cada sitio de origen tiene los valores de participantes individuales necesarios para calcular los agregados del sitio según un algoritmo,

denominado aquí Análisis_O. Sin embargo, los valores de participantes individuales se pueden representar de varias maneras (modelos de datos), incluso de una manera que sea única para ese sitio en particular.

[0076] El sitio de análisis debe acordar con cada sitio de origen un modelo de datos común para su uso en las comunicaciones. Este modelo puede ser predefinido o específico para el estudio.

[0077] El sistema de traducción del modelo de datos en cada sitio de origen, a veces conocido como extracción, transformación y carga (ETL), se configura específicamente para el sitio de origen y sus bases de datos nativas. Sin embargo, la salida del traductor del modelo de datos en cada sitio de origen representa los mismos valores individuales de los participantes en el modelo de datos común.

Precondiciones

- Se identifica un sitio de análisis. Para conveniencia, en adelante se hará referencia a dicho sitio como Sitio_A.

- Se identifican uno o más sitios de origen. A cada sitio de origen se le asigna un nombre único para que el sitio de análisis pueda referirse a él. Para mayor claridad, en este documento los sitios de origen se denominan Sitio₁, Sitio₂, etc., donde el último sitio es Sitio_S, y a cada sitio de origen se le denomina genéricamente Sitio_i.

- El modelo de datos común es utilizado tanto por el Sitio_A como por el Sitio_i.
- Se configura uno o más traductores de modelos de datos en cada Sitio_i.
- Es posible que uno de los sitios de origen coincida con el Sitio_A, pero no se supone que este sea el caso.
- Los permisos y el acceso a la red están configurados de la siguiente manera:

- El sitio de análisis puede acceder a los datos agregados de cada sitio de origen.
- El sitio de análisis puede acceder y descargar análisis del Repositorio de Análisis.

- Cada sitio de origen puede:
 - Acceder y descargar análisis del Repositorio de Análisis según las indicaciones del sitio de análisis, o
 - se enviarán análisis desde el sitio de análisis
- 5 ● El Repositorio de Análisis contiene un análisis que indica al sitio de origen cómo calcular uno o más Agregados del Sitio. Para mayor comodidad, en adelante nos referiremos a este análisis como Análisis_O, independientemente del número de Valores Agregados que calcule.
- 10 ● Análisis_O puede utilizar valores individuales de participantes, agregados de sitios, agregados de estudios o cualquier combinación de los mismos.
- El Repositorio de Análisis tiene por lo menos un análisis que instruye un sitio de análisis sobre cómo calcular uno o más Agregados de Estudio. Para mayor comodidad, nos referiremos a este análisis en adelante como Análisis_A.
- 15 ● Análisis_A está marcado en el Repositorio de Análisis como vinculado al Análisis_O. Es decir, el Análisis_O puede ser utilizado por más de un Análisis_A. Cada estudio único tiene su propio Análisis_A junto con uno o más Análisis_O. Cuando se repite un estudio (por ejemplo, posteriormente o en otro conjunto de sitios), se puede volver a utilizar el mismo conjunto de análisis.
- 20 ● Análisis_A especifica un conjunto de Análisis_O en los que se basa, pero ningún Análisis_O está restringido a ser utilizado por un solo Análisis_A. Por ejemplo, el Análisis_O puede estar relacionado con “contar el número de pacientes asmáticos” y puede ser utilizado por cualquier estudio de pacientes asmáticos.
- 25 ● Análisis_A no utiliza agregados del sitio que no sean calculados por un Análisis_O.
- Análisis_A no utiliza valores individuales de los participantes.

Instigación

[0078] El cálculo de la tarea puede iniciarse de tres maneras:

1. El estudio es iniciado por el Sitio_A en un momento programado, o como resultado de la información proporcionada por el usuario:
 - 5 a. El Sitio_A notifica a cada Sitio_i su intención de sincronizar inmediatamente la hora de inicio del estudio.
 - b. Cada Sitio_i acusa recibo de la solicitud con un mensaje al Sitio_A.
 - i. En algunas implementaciones, el Sitio_i puede responder con una solicitud para reprogramar y sugerir una hora alternativa.
 - 10 ii. En tales casos, el Sitio_A enviará una cancelación a cada Sitio_i y repetirá este Paso 1 más adelante.
 - iii. 1. bi y 1.b.ii se pueden repetir hasta que todos los Sitios_i estén listos.
2. El estudio se inicia a una hora determinada.
 - 15 a. Cada Sitio_i notifica al Sitio_A su disposición para realizar el estudio.
3. El estudio se inicia cuando
 - a. Cada Sitio_i notifica al Sitio_A que se cumple una condición que lo prepara para realizar el estudio.
 - b. El Sitio_A espera hasta recibir dichos mensajes de cada Sitio_i.
 - 20 c. El Sitio_A inicia el estudio inmediatamente según el paso 1 anterior.

[0079] Independientemente del método de inicio utilizado, al finalizar el proceso, todos los Sitios cuentan con la información necesaria para llevar a cabo el estudio y están listos para continuar.

Sincronización de análisis

- 25 1. El Sitio_A descarga el último Análisis_A, es decir, una representación computacional del procedimiento utilizado para calcular los Agregados del Estudio a partir de los Agregados del Sitio.
2. El Sitio_A podría:

- a. Indicar a cada Sitio_i que descargue el mismo Análisis_O al
 - i. Identificar el Análisis_O asociado con el Análisis_A,
 - ii. Enviar la identidad del Análisis_O a cada Sitio_i de manera que
 - iii. Cada Sitio_i descarga el mismo análisis del Repositorio de Análisis.
- b. O bien, descarga el Análisis_O asociado con el Análisis_A y lo sube a cada Sitio_i.

[0080] Al final de esta etapa, el Sitio_A tiene una copia del Análisis_A y cada Sitio_i tiene una copia idéntica del Análisis_O asociado.

10 Estructura de análisis

[0081] Cada análisis es una representación de uno o más algoritmos que realizan parte del estudio.

- El Análisis_O utiliza los valores individuales de los participantes, los agregados del sitio y los agregados del estudio para calcular los agregados del sitio.
- 15 • El Análisis_A utiliza los agregados del sitio para calcular el agregado del estudio.
- El Análisis_O puede incluir instrucciones implícitas o explícitas para pausar temporalmente la ejecución del algoritmo hasta que estén presentes los agregados de estudio necesarios.
- 20 • El Análisis_A puede incluir instrucciones para pausar temporalmente la ejecución del algoritmo hasta que estén presentes los agregados del sitio necesarios.
- El Análisis_A puede incluir instrucciones implícitas o explícitas para enviar agregados del estudio a cada Sitio_i.
- 25 • El resultado del Análisis_O son todos los agregados del sitio necesarios para completar el Análisis_A
- El resultado del Análisis_A son todos los Agregados del Estudio, por ejemplo, en la forma de:

- un informe de texto,
 - un informe en un formato que pueda enviarse a un sistema de notificación automatizado, como el exigido por una autoridad de salud pública u otra autoridad, y/o
- 5 ○ valores en un formato interpretable por otro sistema, como un panel de control o un sistema de visualización de datos.

Implementación por Computador

[0082] La figura 4 es un diagrama de bloques esquemático de un sistema de estudios federados 400 en el que se pueden implementar una o más modalidades de la presente descripción. El sistema 400 incluye un conjunto 410 de sitios de origen, entre los que se encuentran el Sitio₁ 412, el Sitio₂ 414 y el Sitio₃ 416. Los Sitios de Origen _{1...3} (412, 414 y 416) pueden estar ubicados en el mismo lugar, en lugares distintos o en una combinación de ambos.

[0083] Cada Sitio de Origen 412, 414 y 416 está asociado con al menos un dispositivo de procesamiento de datos 413, 415 y 417, respectivamente, conectado a una red de comunicaciones 490. Los dispositivos de procesamiento de datos 413, 415 y 417 transmiten datos agregados a través de la red de comunicaciones 490. Esta red puede estar compuesta por uno o más enlaces de comunicaciones cableados, enlaces de comunicaciones inalámbricos o cualquier combinación de ambos. En particular, la red de comunicaciones 490 puede incluir una red de área local (LAN), una red de área extensa (WAN), una red de telecomunicaciones o cualquier combinación de estas. Una red de telecomunicaciones puede incluir, entre otras, una red de telefonía, como una red telefónica pública conmutada (RTPC) o una red de telefonía móvil celular, Internet o cualquier combinación de estas.

[0084] El sistema 400 también incluye un Sitio de Análisis 450. El Sitio de Análisis 450 incluye un Dispositivo de Datos de Estudio 451 que incluye: una Base de Datos Nativa 452, un Traductor de Modelo de Datos 454, una Base de Datos de Valores de Participantes Individuales en Modelo de Datos Común 456, un módulo de

Computación 458 y una Base de Datos de Valores Agregados del Sitio 460, cada uno de los cuales está conectado a un bus de comunicaciones 459 que permite el intercambio de información entre todos los componentes del Sitio de Análisis 450.

[0085] El traductor de modelos de datos 454 se utiliza cuando el Sitio de Análisis 450 necesita traducir los datos recibidos a un formato de modelo de datos común. En los casos en que los sitios de origen solo transmiten datos agregados según un modelo de datos común predefinido, el traductor de modelos de datos 454 no es necesario.

[0086] El dispositivo de datos de estudio 451 está conectado a la red de comunicaciones 490 para permitir la comunicación entre el Sitio de Análisis 450 y cada uno de los Sitios de Origen 412, 414 y 416. En particular, los Sitios de Origen 412, 414 y 416 utilizan los respectivos dispositivos de computación de datos 413, 415 y 417 para transmitir conjuntos de datos agregados correspondientes a uno o más estudios, a través de la red de comunicaciones 490, al dispositivo de datos de estudio 451.

[0087] En el ejemplo de la figura 4, el sistema 400 incluye opcionalmente un primer dispositivo informático de observación 470 y un primer observador 475 asociado. El observador 475 puede acceder al dispositivo informático de observación 470 para consultar una interfaz de usuario proporcionada por el sitio de análisis 450 y visualizar los datos asociados a uno o más estudios federados. El dispositivo informático de observación 470 puede estar integrado en el dispositivo de datos de estudio 451 o conectado a este mediante uno o más enlaces de comunicación cableados o inalámbricos.

[0088] En el ejemplo de la figura 4, el sistema 400 incluye opcionalmente un segundo dispositivo informático de observación remota 480 y un segundo observador asociado 485. El dispositivo informático de observación remota 480 está conectado al Dispositivo de Datos de Estudio 451 mediante uno o más enlaces de comunicación cableados o inalámbricos, incluida la red de comunicaciones 490. El

segundo observador 485 puede acceder al dispositivo informático de observación remota 480 para acceder a una interfaz de usuario proporcionada por el Sitio de Análisis 450 y así visualizar los datos asociados a uno o más estudios federados.

[0089] A continuación se describirá la funcionalidad del sistema de estudios federados 400 en relación con un estudio de ejemplo relativo al peso corporal de pacientes con diabetes tipo 2.

Ejemplo: Media ($\hat{Y}$) y Desviación Estándar (σ) del peso corporal de pacientes con diagnóstico de diabetes mellitus tipo 2 en 3 centros desde 2010.

Dispositivos de entrada y bases de datos nativas de los sitios de origen

- Los sitios de origen (Sitio₁, Sitio₂ y Sitio₃) son centros de atención médica.
- Cada sitio de origen cuenta con un sistema de registro electrónico de salud asociado.
- El Sitio₁ cuenta con básculas electrónicas que registran automáticamente el peso corporal en kilogramos (kg) en el registro electrónico de salud del paciente.
- En el Sitio₂, un médico pesa a los pacientes e introduce el peso corporal en kg en el historial clínico electrónico del paciente.
- En el Sitio₃, una enfermera pesa a los pacientes e ingresa el peso corporal del paciente en libras (lb) en el registro electrónico de salud del paciente.
- En cada Sitio de Origen, un médico introduce un diagnóstico de diabetes mellitus tipo 2 en el historial clínico electrónico del paciente diagnosticado.

Modelo de datos común

[0090] En este ejemplo, el modelo de datos común para el estudio incluye una tabla de condiciones y una tabla de observaciones. En otras implementaciones relacionadas con diferentes estudios, el modelo de datos común puede incluir una o más tablas u otros formatos de datos adecuados. En este ejemplo, la información de la tabla de condiciones y la tabla de observaciones se obtiene de la base de datos nativa en el sitio de origen correspondiente.

[0091] Parte de la información en la Tabla de Condiciones y la Tabla de Observaciones del modelo de datos común utilizado en este ejemplo emplea códigos de la terminología clínica SNOMED-CT (snomed.org), que proporciona una traducción de estos códigos a una descripción textual estandarizada e inequívoca de la condición, observación, etc. En este ejemplo, las descripciones textuales de los códigos relevantes no se incluyen en el modelo de datos común, pero se proporcionan en el cuerpo del texto para mayor comodidad.

[0092] Los diagnósticos de diabetes mellitus tipo 2 se registran en una tabla de “afecciones”, con el número de identificación único del paciente (UID), la fecha y hora de la medición y el código SNOMED-CT 44054006 (“Diabetes mellitus tipo 2”) como se ilustra en la Tabla 1 a continuación:

Condiciones		
UID	Fecha y hora	código de condición
1	21 de abril de 2010, 14:35	44054006

Tabla 1

[0093] El peso corporal es registrado en una tabla de “Observaciones” con el UID del paciente, con el código SNOMED 27113001 (“Peso corporal”) y el valor correspondiente en kg y la unidad como código SNOMED 258683005 (“kg”) como se ilustra en la Tabla 2 a continuación.

Observaciones				
UID	Fecha y hora	código de observación	valor de observación	código de unidad
1	21 de abril de 2010, 14:30	27113001	98.3	258683005

Tabla 2

Traductores del modelo de datos de los sitios de origen

[0094] Cada Sitio de Origen cuenta con su propio traductor, como el Traductor de
5 Modelo de Datos 114 de la figura 1, para traducir los datos de su propio sistema de historia clínica electrónica al modelo de datos común. En el Sitio₃, el traductor incluye la conversión de pesos corporales de libras (lb) a kilogramos (kg).

Ejemplo: cálculo de la media y la desviación estándar del peso corporal de los participantes en un estudio federado.

10 **Fórmulas para calcular la media de la fecha y hora ($\hat{Y}$) y la desviación estándar (σ)**

$$\hat{Y} = \frac{1}{N} \sum_{j=1}^N Y_j \quad \text{Ecuación (1)}$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{j=1}^N (Y_j - \hat{Y})^2} \quad \text{Ecuación (2)}$$

Pseudocódigo de Análisis_A

- 15 1. Solicitar y esperar el número de participantes incluidos (n_1 , n_2 y n_3) de cada Sitio de Origen correspondiente.
2. Calcular el agregado del estudio “número de pacientes incluidos en el estudio” N como $N = \sum_{i=1}^3 n_i$
3. Solicitar y esperar el valor agregado del sitio “peso corporal medio del sitio” (
20 $\hat{Y}_1$, $\hat{Y}_2$ y $\hat{Y}_3$) de cada sitio de origen correspondiente.
4. Calcular el agregado del estudio “peso corporal medio del estudio” $\hat{Y}$ como $\hat{Y} = \frac{1}{N} \sum_{i=1}^3 \hat{Y}_i \cdot n_i$
5. Solicitar y esperar el agregado del sitio “suma de diferencias al cuadrado” (
 $\hat{Z}_1$, $\hat{Z}_2$ y $\hat{Z}_3$) de cada sitio de origen correspondiente.
- 25 6. Calcular la “desviación estándar” del agregado del estudio σ a partir de la “suma de las diferencias al cuadrado” de los agregados del sitio ($\hat{Z}_1$, $\hat{Z}_2$, $\hat{Z}_3$)

y el “número de pacientes incluidos en el estudio” del agregado del estudio

$$N \text{ como } \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^3 \hat{Z}_i \cdot n_i}$$

7. Informe del estudio Agregado “peso corporal medio del estudio” $\hat{Y}$ y “desviación estándar” σ

5 Pseudocódigo de Análisis

1. Establecer el número de sitio como i
2. Contar el número de pacientes únicos incluidos con el código de condición 44054006 con fecha y hora de condición entre el 1 de enero de 2010 a las 12:00 a. m. y el 31 de diciembre de 2010 a las 11:59 p. m. como n_i
- 10 3. Calcular el “peso corporal medio del sitio” $\hat{Y}_i$ a partir del peso corporal del valor individual del participante (Y) para cada paciente incluido j y el número de pacientes incluidos en el sitio n como $\hat{Y}_i = \frac{1}{n} \sum_{j=1}^n Y_j$
4. Solicitar y esperar el dato agregado del estudio “peso corporal medio del estudio” $\hat{Y}$ del Sitio_A.
- 15 5. Calcular el agregado del sitio “suma de las diferencias al cuadrado” $\hat{Z}_i$ del peso corporal de cada paciente incluido j y el agregado del estudio “peso corporal medio del estudio” $\hat{Y}$ como $\hat{Z}_i = \sum_{j=1}^n (Y_j - \hat{Y})^2$

[0095] El sistema de estudios federados de la presente descripción puede implementarse utilizando uno o más dispositivos informáticos, tales como un

20 computador de propósito general, un servidor, una arquitectura distribuida en la nube, una combinación de estos o cualquier arquitectura similar programada para realizar una o más de las funciones mostradas y descritas en las figuras 1 a 4, dando lugar así a un dispositivo informático nuevo y mejorado. En particular, el dispositivo de datos de estudio 115a y el dispositivo de análisis de datos 451 representan

25 avances en tecnología informática implementados mediante uno o más dispositivos informáticos.

[0096] La figura 5 es un diagrama de bloques esquemático de un sistema 500 que incluye un computador de propósito general 510. El computador de propósito general 510 incluye varios componentes, entre ellos: un procesador 512, una memoria 514, un medio de almacenamiento 516, interfaces de entrada/salida (E/S) 520 y puertos de entrada/salida (E/S) 522. Los componentes del computador de propósito general 510 generalmente se comunican mediante uno o más buses 548.

[0097] La memoria 514 puede implementarse mediante memoria de acceso aleatorio (RAM), memoria de solo lectura (ROM) o una combinación de ambas. El medio de almacenamiento 516 puede implementarse como un disco duro, una unidad de estado sólido (sólido), una unidad de disco óptico u otro medio de almacenamiento. El medio de almacenamiento 516 puede utilizarse para almacenar uno o más programas informáticos, incluyendo un sistema operativo, aplicaciones de software y datos. En un modo de funcionamiento, las instrucciones de uno o más programas informáticos almacenados en el medio de almacenamiento 516 se cargan en la memoria 514 a través del bus 548. Las instrucciones cargadas en la memoria 514 se ponen a disposición del procesador 512 a través del bus 548 u otros medios para su ejecución, implementando así un modo de funcionamiento acorde con las instrucciones ejecutadas.

[0098] Uno o más dispositivos periféricos pueden conectarse al computador de propósito general 510 a través de los puertos de E/S 522. En el ejemplo de la figura 5, el computador de propósito general 510 está conectado a un altavoz 524, una cámara 526, una pantalla 530, un dispositivo de entrada 532, una impresora 534 y un dispositivo de almacenamiento externo 536. El altavoz 524 puede implementarse mediante uno o más altavoces, como en un sistema de sonido estéreo o envolvente. En el ejemplo en el que el computador de propósito general 510 se utiliza para implementar una o más funciones de un sistema de estudios federados, como un dispositivo de datos de estudio, uno o más dispositivos periféricos pueden

corresponder a los dispositivos de entrada 111a...d de la figura 3, conectados a los puertos de E/S 522 de forma inalámbrica o mediante conexión por cable.

[0099] La cámara 526 puede ser una webcam u otra cámara digital de fotos o vídeo, y puede descargar y cargar información desde y hacia el computador de uso general 510 a través de los puertos de E/S 522, según la implementación específica.

Por ejemplo, las imágenes grabadas por la cámara 526 pueden cargarse en el dispositivo de almacenamiento 516 del computador de uso general 510. Del mismo modo, las imágenes almacenadas en el dispositivo de almacenamiento 516 pueden descargarse a la memoria o al dispositivo de almacenamiento de la cámara 526. La cámara 526 puede incluir un sistema de lentes, una unidad de sensor y un dispositivo de grabación.

[00100] El dispositivo de visualización 530 puede ser un monitor de computador, como una pantalla de tubo de rayos catódicos, una pantalla de plasma o una pantalla de cristal líquido (LCD). La pantalla 530 recibe información del computador 510 de forma convencional, y esta se presenta en el dispositivo 530 para que el usuario la visualice. Opcionalmente, el dispositivo de visualización 530 puede incorporar una pantalla táctil para que el usuario pueda interactuar con el computador 510. La pantalla táctil puede ser, por ejemplo, capacitiva, resistiva, de ondas acústicas superficiales, etc.

[00101] El dispositivo de entrada 532 puede ser un teclado, un ratón, un lápiz óptico, una tableta gráfica o cualquier combinación de estos, para recibir datos del usuario. El medio de almacenamiento externo 536 puede incluir un disco duro externo (HDD), una unidad óptica, una unidad de disquete, una unidad flash, una unidad de estado sólido (SSD) o cualquier combinación de estos, y puede implementarse como una o varias instancias de cualquiera de estos dispositivos. Por ejemplo, el medio de almacenamiento externo 536 puede implementarse como una matriz de discos duros.

[00102] Las interfaces de E/S 520 facilitan el intercambio de información entre el dispositivo informático de propósito general 510 y otros dispositivos informáticos. Estas interfaces pueden implementarse mediante un módem interno o externo, una conexión Ethernet, etc., para permitir la conexión a un medio de transmisión. En el ejemplo de la figura 5, las interfaces de E/S 522 están conectadas a una red de comunicaciones 538 y directamente a un dispositivo informático 542. Este dispositivo 542 se muestra como un computador personal, pero también podría ser un teléfono inteligente, un computador portátil o una tableta. La comunicación directa entre el computador de propósito general 510 y el dispositivo informático 542 puede implementarse mediante un enlace de transmisión inalámbrico o cableado.

[00103] La red de comunicaciones 538 puede implementarse mediante uno o más enlaces de transmisión alámbricos o inalámbricos, e incluir, por ejemplo, un enlace de comunicaciones dedicado, una red de área local (LAN), una red de área extensa (WAN), Internet, una red de telecomunicaciones o cualquier combinación de estas.

Una red de telecomunicaciones puede incluir, entre otras, una red de telefonía, como una red telefónica pública conmutada (RTPC), una red celular de telefonía móvil, una red de mensajes cortos (SMS) o cualquier combinación de estas. El computador de uso general 510 puede comunicarse a través de la red de comunicaciones 538 con otros dispositivos informáticos conectados a dicha red, como el teléfono móvil 544, el teléfono inteligente con pantalla táctil 546, el computador personal 540 y el dispositivo informático 542.

[00104] Se pueden utilizar una o más unidades del computador de propósito general 510 para implementar una o más funciones del dispositivo de datos de estudio 115a de la figura 3, con el fin de implementar un Sitio de Origen de un sistema de estudios federados, de acuerdo con la presente descripción. En dicha implementación, la memoria 514 y el almacenamiento 516 se utilizan para almacenar datos relativos a datos de pacientes, datos de análisis, etc. El software para implementar el sistema de estudios federados se almacena en la memoria 514 o en el almacenamiento 516,

o en ambos, para su ejecución en el procesador 512. El software incluye código de programa informático para implementar los pasos del método de acuerdo con los módulos funcionales descritos en este documento.

[00105] La figura 6 es un diagrama de bloques esquemático de un sistema 600 en el que se puede implementar uno o más aspectos de un método y sistema federado de la presente descripción. El sistema 600 incluye un dispositivo informático portátil, un teléfono inteligente 610, que puede ser utilizado por un usuario registrado del sistema de estudios federados de la figura 1. El teléfono inteligente 610 incluye varios componentes, entre ellos: un procesador 612, una memoria 614, un medio de almacenamiento 616, una batería 618, una antena 620, un transmisor y receptor de radiofrecuencia (RF) 622, una tarjeta SIM 624, un altavoz 626, un dispositivo de entrada 628, una cámara 630, una pantalla 632 y un transmisor y receptor inalámbrico 634. Los componentes del teléfono inteligente 610 se comunican generalmente mediante una o más conexiones de bus 648 u otras conexiones entre ellos. El teléfono inteligente 610 también incluye una conexión por cable 645 para acoplarse a una toma de corriente para recargar la batería 618 o para conectarse a un dispositivo informático, como el computador de uso general 510 de la figura 5. La conexión por cable 645 puede incluir uno o más conectores y puede estar adaptada para permitir la carga y descarga de contenido desde y hacia la memoria 614 y la tarjeta SIM 624.

[00106] El teléfono inteligente 610 puede incluir muchos otros componentes funcionales, como un convertidor digital-analógico y analógico-digital de audio y un amplificador, pero se omiten para mayor claridad. Sin embargo, un experto en la materia los conocería y comprendería fácilmente.

[00107] La memoria 614 puede incluir memoria de acceso aleatorio (RAM), memoria de solo lectura (ROM) o una combinación de ambas. El medio de almacenamiento 616 puede ser una unidad de estado sólido (sólido), un medio de almacenamiento extraíble, como una tarjeta SD o microSD, u otro tipo de almacenamiento. El medio

de almacenamiento 616 puede utilizarse para almacenar uno o más programas informáticos, incluyendo un sistema operativo, aplicaciones de software y datos. En un modo de funcionamiento, las instrucciones de uno o más programas informáticos almacenados en el medio de almacenamiento 616 se cargan en la memoria 614 a través del bus 648. Las instrucciones cargadas en la memoria 614 se ponen a disposición del procesador 612 a través del bus 648 u otros medios para su ejecución, implementando así un modo de funcionamiento acorde con las instrucciones ejecutadas.

[00108] El teléfono inteligente 610 también incluye un módulo de interfaz de programación de aplicaciones (API) 636, que permite a los programadores escribir aplicaciones de software para ejecutar en el procesador 612. Dichas aplicaciones incluyen una pluralidad de instrucciones que pueden estar preinstaladas en la memoria 614 o descargadas a la memoria 614 desde una fuente externa, a través del transmisor y receptor de RF 622 que operan en asociación con la antena 620 o a través de la conexión cableada 645.

[00109] El teléfono inteligente 610 incluye además un módulo de localización GPS 638. Este módulo se utiliza para determinar la posición geográfica del dispositivo mediante satélites GPS, triangulación de antenas de telefonía móvil o una combinación de ambos métodos. Dicha posición geográfica se puede poner a disposición de uno o más programas o aplicaciones que se ejecutan en el procesador 612.

[00110] El transmisor y receptor inalámbricos 634 pueden utilizarse para comunicarse de forma inalámbrica con dispositivos periféricos externos mediante Bluetooth, infrarrojos u otro protocolo inalámbrico. En el ejemplo de la figura 6, el teléfono inteligente 610 está conectado a una impresora 640, un dispositivo de almacenamiento externo 644 y un computador 642. El computador 642 puede implementarse, por ejemplo, utilizando el computador de propósito general 510 de la figura 5.

[00111] La cámara 626 puede incluir una o más cámaras digitales, tanto de fotografía como de vídeo, adaptadas para capturar y grabar en la memoria 614 o en la tarjeta SIM 624 imágenes fijas o de vídeo, o una combinación de ambas. La cámara 626 puede incluir un sistema de lentes, una unidad de sensor y un medio de grabación. El usuario del teléfono inteligente 610 puede cargar las imágenes grabadas a otro computador o dispositivo periférico mediante el transmisor y receptor inalámbricos 634, el transmisor y receptor de radiofrecuencia 622 o la conexión por cable 645.

[00112] En un ejemplo, el dispositivo de visualización 632 utiliza una pantalla de cristal líquido (LCD). La pantalla 632 se emplea para mostrar contenido al usuario del teléfono inteligente 610. Opcionalmente, la pantalla 632 puede implementarse mediante una pantalla táctil, como una pantalla táctil capacitiva o resistiva, para que el usuario pueda interactuar con el teléfono inteligente 610.

[00113] El dispositivo de entrada 628 puede ser, por ejemplo, un teclado, un lápiz óptico o un micrófono, para recibir datos del usuario. Si se trata de un teclado, este puede implementarse como un conjunto de teclas físicas en el teléfono inteligente 610. Alternativamente, puede ser un teclado virtual que se muestra en la pantalla 632.

[00114] La tarjeta SIM 624 se utiliza para almacenar la Identidad Internacional del Abonado Móvil (IMSI) y una clave asociada que permite identificar y autenticar al usuario en la red celular a la que está suscrito. La tarjeta SIM 624 suele ser extraíble y se puede usar indistintamente en diferentes teléfonos inteligentes o celulares. En ella se pueden almacenar los contactos del usuario, incluyendo nombres y números de teléfono. También puede almacenar fotos y videos. Como alternativa, los contactos se pueden guardar en la memoria 614.

[00115] El transmisor y receptor de radiofrecuencia 622, junto con la antena 620, permiten el intercambio de información entre el teléfono inteligente 610 y otros dispositivos informáticos a través de una red de comunicaciones 690. En el ejemplo

de la figura 6, el transmisor y receptor de radiofrecuencia 622 permiten que el teléfono inteligente 610 se comuniquen a través de la red de comunicaciones 690 con un teléfono celular 650, un teléfono inteligente o tableta 652, un dispositivo informático 654 y otro dispositivo informático 642. Los dispositivos informáticos 654 y 642 se muestran como computadores personales, pero también podrían ser teléfonos inteligentes, computadores portátiles o tabletas.

[00116] La red de comunicaciones 690 puede implementarse mediante uno o más enlaces de transmisión alámbricos o inalámbricos, e incluir, por ejemplo, una red de telefonía celular, un enlace de comunicaciones dedicado, una red de área local (LAN), una red de área extensa (WAN), Internet, una red de telecomunicaciones o cualquier combinación de estas. Una red de telecomunicaciones puede incluir, entre otras, una red de telefonía, como una red telefónica pública conmutada (RTPC), una red celular de telefonía móvil, una red de mensajes cortos (SMS) o cualquier combinación de estas.

[00117] Cuando se implementan una o más funciones del sistema de estudios federados descrito en este documento utilizando el teléfono inteligente 610 de la figura 6, se puede utilizar una aplicación de programa ("app") que se ejecuta en el procesador 612 para implementar cualquiera de las funciones descritas y mostradas en relación con las figuras 1 a 4. En algunas implementaciones, la app es una aplicación nativa que se ejecuta en el teléfono inteligente 600. En otras implementaciones, la app es una aplicación web que se muestra en un navegador que se ejecuta en el procesador 612, con el teléfono inteligente 610 conectado a un servidor remoto, como el dispositivo informático 642, en el que se ejecuta la app.

[00118] La figura 7 es un diagrama de secuencia que ilustra el flujo de trabajo 700 de un estudio federado realizado por una implementación del sistema descrito en la presente divulgación. El flujo de trabajo 700 se refiere a un sistema que cuenta con un repositorio 702, un usuario 704, un sitio de análisis 706 y tres sitios de origen: un

primer sitio de origen 708, un segundo sitio de origen 710 y un tercer sitio de origen 712.

[00119] En un primer paso, el usuario 704 inicia un estudio federado en el sitio de análisis 706, cuyo inicio se produce cuando dicho sitio recupera el AnálisisA del repositorio 702. Como se describió anteriormente, cada AnálisisA está asociado a un conjunto de uno o más AnálisisO. El sitio de análisis 706 identifica el AnálisisO asociado al AnálisisA recuperado y, a continuación, lo recupera del repositorio 702.

[00120] El sitio de análisis 706 proporciona entonces el AnálisisO recuperado a cada uno de los sitios de origen 708, 710 y 712. De este modo, cada sitio de origen 708, 710 y 712 dispone del AnálisisO para llevar a cabo un estudio.

[00121] Cada sitio de origen (708, 710, 712) agrega los datos del estudio correspondientes. En la figura 7, esto se muestra mediante el cálculo de n_1 por análisis (708) para el primer sitio de origen, n_2 por análisis (710) para el segundo y n_3 por análisis (712) para el tercero.

[00122] El sitio de análisis 706 recupera los valores calculados de n_1 , n_2 y n_3 de los respectivos sitios de origen 708, 710 y 712, y luego calcula N de acuerdo con AnalysisO. Posteriormente, el sitio de análisis 706 proporciona un informe al usuario 704, basado en el valor calculado de N .

[00123] La figura 8 es un diagrama de bloques esquemático de la arquitectura de una máquina virtual en la nube para implementar un sitio de origen 800 en un sistema descrito en la presente divulgación. El ejemplo de la figura 8 utiliza la plataforma de computación en la nube AWS, proporcionada por Amazon Web Services, Inc. En concreto, se utiliza Amazon Virtual Private Cloud (VPC) para implementar el sitio de origen, y Amazon Elastic Block Store (EBS) como almacenamiento a nivel de bloque para almacenar datos. Dichos datos pueden incluir, por ejemplo, datos de pacientes, datos de estudios, análisis, etc.

[00124] El ejemplo de la figura 8 también utiliza Amazon Elastic Cloud Compute (EC2) como máquina virtual 810, que aloja una aplicación de software para

implementar la funcionalidad del sitio de origen. La máquina virtual 810 incluye un módulo de cómputo 818, equivalente al módulo de cómputo 118 de la figura 3; una base de datos de participantes 816, equivalente a la base de datos 116 de la figura 3; y una base de datos de datos agregados 820, equivalente a la base de datos de valores agregados 120 de la figura 3.

[00125] El sitio de origen 800 es accesible a través de Internet por usuarios finales que acceden mediante navegadores web en dispositivos informáticos. El sitio de origen 800 también está conectado, mediante una puerta de enlace a Internet, a una base de datos 850 que funciona como repositorio de análisis.

10 [00126] La figura 9 es un diagrama de bloques esquemático de un sistema de Sitio de Origen 900 implementado mediante una arquitectura informática "sin servidor" basada en la nube 910, proporcionada en este ejemplo por Amazon Web Services. Un sitio de administración 920 se implementa utilizando una máquina virtual que incluye el sistema de distribución de software Elastic Container Registry (ECR) para
15 distribuir software al sitio del cliente 930.

[00127] El sitio del cliente 930 se implementa utilizando una máquina virtual 932 que implementa el sitio de origen 110a de las figuras 1 y 3. El sitio del cliente 930 incluye una serie de componentes funcionales, entre ellos un módulo de computación, una base de datos de datos agregados y una base de datos de datos de participantes,
20 tal como se establece y describe en la figura 3. El sitio del cliente 930 está conectado, mediante uno o más enlaces de comunicación, a un computador de usuario final 940 y a un repositorio de análisis 950.

[00128] La figura 10 es un diagrama de bloques esquemático de un sistema de estudios federados 1000 implementado mediante una arquitectura informática
25 basada en la nube. El sistema 1000 incluye un módulo de análisis federado 1005 que representa conceptualmente cómo un conjunto de instrucciones del sitio de análisis 1006 y un conjunto de instrucciones del sitio de análisis 1006 y un conjunto

de instrucciones del sitio de origen 1007 podría organizarse de manera que parezca un único módulo funcional.

[00129] Se implementa un sitio de origen 1010 mediante un sistema basado en la nube, como el descrito en la figura 8. El sitio de origen 1010 incluye un conjunto de instrucciones 1012 y un modelo de datos común 1014 definido para un estudio federado específico. El sitio de origen 1010 está conectado a una puerta de enlace de red 1060 que permite el intercambio de datos entre el sitio de origen 1010, un sitio de análisis 1050 y un repositorio de análisis 1020. El sitio de análisis 1050 incluye un conjunto de instrucciones 1052.

[00130] La figura 11 es un diagrama de bloques esquemático de un sistema de estudios federados 1100 implementado mediante una arquitectura informática basada en la nube. El sistema 1100 del ejemplo de la figura 11 incluye un primer sitio de origen 1115 alojado en un primer centro de datos, un segundo sitio de origen 1125 alojado en un segundo centro de datos, un sitio de análisis 1150 alojado en un tercer centro de datos y un repositorio de análisis 1160; cada uno de estos elementos se implementa en este ejemplo en un entorno de computación en la nube.

[00131] El primer sitio de origen 1115 está conectado a una primera puerta de enlace de red 1118, la cual está conectada a una red de comunicaciones 1190. El segundo sitio de origen 1125 está conectado a una segunda puerta de enlace de red 1128, que también está conectada a la red de comunicaciones 1190. El sitio de análisis 1150 y el repositorio de análisis 1160 se encuentran en un tercer centro de datos común y están conectados, a través de una tercera puerta de enlace de red 1125, a la red de comunicaciones 1190.

[00132] La red de comunicaciones 1190 puede implementarse utilizando uno o más enlaces de comunicaciones alámbricos o inalámbricos, y puede incluir Internet, que permiten la transmisión de datos entre el primer sitio de origen 1115, el segundo sitio de origen 1125, el sitio de análisis 1150 y el repositorio de análisis 1160.

[00133] El sistema de estudios federados 1100 de la figura 11 permite que dos sitios de origen 1115, 1125 recopilen datos de participantes de sitios separados y luego transmitan esos datos, de acuerdo con un modelo de datos común predefinido, para su procesamiento por el sitio de análisis 1150. Dicho sistema de estudios federados 1100 proporciona una arquitectura informática mejorada para realizar estudios federados.

Unificación de dialectos del modelo de datos

[00134] La pluralidad de representaciones de datos en distintos sitios puede generar variabilidad en el “modelo” o “formato” de los datos en un repositorio, como una base de datos relacional o un lago de datos. Esta variabilidad puede introducir errores en el análisis, la recuperación y la actualización de datos en dichos repositorios, especialmente cuando se combinan varios repositorios mediante agregación (copiando contenido de uno o más repositorios a otro) o federación (combinando la información resumida obtenida en cada repositorio).

[00135] Para abordar este problema, se han propuesto varios estándares de datos comunes, como FHIR, OMOP e I2B2 en medicina. Sin embargo, la traducción de datos a estos modelos comunes suele requerir conocimientos especializados, subjetividad y criterio, lo que conlleva diferencias entre el modelo de datos traducido y el estandarizado (a veces denominadas “dialectos”). Estas diferencias entre dialectos limitan la capacidad de combinar datos, incluso de repositorios que utilizan el mismo modelo de datos común.

[00136] Existen dos razones principales por las que se producen dialectos: la interpretación y la omisión. La interpretación se produce cuando las instrucciones para modelar datos según un estándar son ambiguas o poco claras, lo que permite que quienes transforman los datos en diferentes lugares o en diferentes momentos las interpreten de forma distinta. Como resultado, aunque los conjuntos de datos traducidos sean nominalmente compatibles con el mismo estándar, es posible que

no sean interoperables entre sí. La solución consiste en utilizar las mismas asignaciones en cada lugar.

[00137] Se produce una omisión cuando los desarrolladores del estándar no previeron un caso de uso al publicarlo, antes de conocer algunos de los datos que debía representar el modelo de datos. La mayoría de los estándares proporcionan directrices para la extensión del estándar del modelo de datos original. Por ejemplo, FHIR proporciona un recurso de extensión y OMOP reserva un rango de identificadores únicos para extensiones de código de concepto (2.000.000.000 y superiores). Sin embargo, incluso cuando las extensiones se crean según las directrices, la flexibilidad de la extensión implica que los conjuntos de datos de cada sitio suelen ser incompatibles entre sí.

[00138] Por ejemplo, si dos sitios, el Sitio 1 y el Sitio 2, necesitan representar un nuevo virus (X) y un nuevo procedimiento quirúrgico (Y), cada uno podría asignarles códigos de concepto diferentes, como se muestra en la Tabla 3. La solución consiste en utilizar los mismos códigos de extensión para las mismas extensiones en ambos sitios.

	ID de concepto asignado en el sitio 1	ID de concepto asignado en el sitio 2
Nuevo virus X	2.000.000.001	2.000.000.002
Nuevo procedimiento quirúrgico Y	2.000.000.002	2.000.000.003

Tabla 3

[00139] Dado que ambos sitios son compatibles con el estándar pero no entre sí, el modelo de datos de cada sitio se denomina “dialecto” del estándar. El estándar es análogo a un idioma, y los datos procedentes de distintos sitios son incompatibles porque cada sitio utiliza un dialecto diferente de ese idioma.

[00140] La ambigüedad y la omisión se cuelan en la documentación del estándar, a menudo cuando sus autores no prevén todos los conjuntos de datos que podrían necesitar transformación ni las decisiones de traducción específicas que estos requieren. Estudios recientes demuestran que dos sistemas que utilizan el mismo estándar pueden ser compatibles solo entre un 30 % y un 60 % debido a una combinación de estos factores. Ver, por ejemplo, Elmer V. Bernstam et al., *“Quantitating and assessing interoperability between electronic health records”*, Journal of the American Medical Informatics Association, volumen 29, número 5, mayo de 2022, páginas 753-760, <https://doi.org/10.1093/jamia/ocab289>.

[00141] Independientemente de la causa, el resultado es que cada sitio desarrolla su propio dialecto específico. Este fenómeno es tan común que la formación de dialectos se produce casi siempre que se estandariza un nuevo conjunto de datos a un modelo de datos común, generalmente sin que nadie se dé cuenta hasta después de que los datos ya se han traducido. Los transformadores deben tener especial cuidado y, como medida de precaución, es necesario acordar un protocolo.

[00142] Algunas implementaciones de la presente invención proporcionan un método para automatizar la unificación de traducciones de datos entre múltiples sitios que colaboran, de modo que se genere el mismo dialecto en todos ellos. Esto aumenta la interoperabilidad entre dichos sitios. El método utiliza un repositorio central al que puede acceder un sistema de traducción en cada sitio. En algunas implementaciones, el repositorio central se implementa mediante un medio de almacenamiento legible por computador e instrucciones ejecutables por computador que se ejecutan en el procesador de un dispositivo informático conectado a una red de comunicaciones para comunicarse con uno o más sitios de origen.

[00143] El repositorio central almacena cada extensión y mapeo, y permite que las mismas extensiones y mapeos se compartan entre todos los sitios. Cada extensión y mapeo está asociado a un identificador único en el repositorio. El método también

utiliza un cliente de traducción presente en cada sitio. El cliente de traducción es un software que se ejecuta en uno o más procesadores del sitio correspondiente.

[00144] Este método utiliza además un Servicio de Emisión de Identificadores, que proporciona una conexión entre cada sitio de origen y el repositorio central. El

5 Servicio de Emisión de Identificadores se encarga de emitir nuevos identificadores únicos para las nuevas extensiones y asignaciones, y de reutilizar los identificadores existentes para las extensiones o asignaciones que ya se encuentran en el repositorio.

[00145] En algunas implementaciones, el Servicio de Emisión de Identificadores se
10 implementa utilizando un medio de almacenamiento legible por computador e instrucciones ejecutables por computador que se ejecutan en un procesador de un dispositivo informático conectado a una red de comunicaciones para la comunicación con uno o más sitios de origen y el repositorio central.

[00146] En algunas implementaciones, el repositorio central se incluye dentro del
15 Repositorio de Análisis 120, 702, 850, 950, 1020, 1160 descrito anteriormente. En otras implementaciones, el repositorio central se implementa como un nuevo nodo de red conectado a una red de comunicaciones para que sea accesible desde todos los sitios de origen dentro de una red de estudios federada.

[00147] En algunas implementaciones, el Servicio de Emisión de Identificadores se
20 encuentra ubicado junto al repositorio central o forma parte integral del mismo, como por ejemplo, al integrarse en el Repositorio de Análisis 120, 702, 850, 950, 1020 y 1160 descritos anteriormente. En otras implementaciones, el Servicio de Emisión de Identificadores se implementa como un nuevo nodo de red conectado a una red de comunicaciones, de modo que sea accesible para todos los sitios de origen y el
25 repositorio central dentro de una red de estudios federada.

[00148] La figura 12 es un diagrama de bloques esquemático que ilustra una porción de una red de estudios federados con un Sitio de Origen 1240, representativo de cualquier sitio de origen dentro de dicha red. El Sitio de Origen 1240 incluye el

software de traducción 1250 y un cliente de traducción 1260. El cliente de traducción 1260 está configurado para comunicarse, a través de una primera red de comunicaciones 1270, con un Servicio de Emisión de Identificadores 1210. Este Servicio de Emisión de Identificadores 1210 está conectado, mediante una segunda
5 red de comunicaciones 1230, a un repositorio central 1220.

[00149] Según la implementación, la primera red de comunicación 1270 y la segunda red de comunicación 1230 pueden ser la misma red, como una red de área local (LAN), una red de área extensa (WAN), una red de telecomunicaciones o cualquier combinación de estas. Una red de telecomunicaciones puede incluir, entre
10 otras, una red de telefonía, como una red telefónica pública conmutada (RTPC) o una red de telefonía móvil celular, Internet o cualquier combinación de estas. En otras implementaciones, la primera red de comunicación 1270 y la segunda red de comunicación 1230 pueden ser enlaces físicos directos, como un bus de computador, en circunstancias donde dos o más de los siguientes elementos (el
15 sitio de origen 1240, el servicio emisor de identificadores 1210 y el repositorio central 1220) se encuentran ubicados en el mismo lugar o integrados entre sí.

[00150] En el ejemplo de la Fig. 12, el Servicio de Emisión de Identificadores 1210 y el repositorio central 1220 se muestran como parte de una Ubicación Central 1200. Sin embargo, se debe tener en cuenta que el Servicio de Emisión de Identificadores
20 1210 y el repositorio central 1220 pueden ser nodos separados ubicados remotamente entre sí, ubicados cerca uno del otro o incluso integrados entre sí en diferentes modalidades de la presente descripción.

[00151] El repositorio central en línea almacena y recupera elementos de traducción de datos (como se describe a continuación) para un sistema, así como
25 identificadores únicos para las extensiones definidas. El repositorio central utiliza estos identificadores únicos para recuperar las extensiones solicitadas. Puede ser específico para estos elementos o, de forma más general, admitir también otros, como flujos de trabajo de análisis federado. Se pueden usar repositorios separados

para cada elemento, siempre que, en conjunto, sean lógicamente equivalentes a un único repositorio.

[00152] El repositorio central actúa como un sistema de almacenamiento que guarda y recupera elementos de traducción de datos, pero no los ejecuta. Los sistemas de traducción, ubicados en cada sitio de origen, utilizan estos elementos durante el proceso de traducción. Según la naturaleza de los elementos de traducción de datos y la aplicación específica, algunos se pueden ejecutar, otros usarse como referencias y otros como recursos del sistema de traducción.

[00153] El servicio de emisión de identificadores garantiza que todos los sitios web hagan referencia al mismo elemento de traducción utilizando el mismo identificador. El repositorio permite que todos los traductores de cada sitio de origen descarguen, mediante clientes de traducción, el mismo elemento de traducción utilizando el identificador único asociado a dicho elemento.

[00154] El resultado es que todos los sistemas de traducción en los respectivos sitios de origen hacen referencia al mismo elemento con el mismo identificador, además de utilizar el mismo elemento que cada sistema de traducción recupera del repositorio central. Los elementos de traducción solo pueden ser utilizados por los traductores en los sitios de origen, ya que la traducción se realiza sobre datos de pacientes, y estos datos no salen de su sitio de origen.

[00155] El cliente de traducción 1260 es utilizado por cada sistema de traducción en cada sitio de origen, como el traductor de modelos de datos 114 de la figura 3, para compartir y recuperar conjuntos de elementos de traducción de datos. Cuando el cliente de traducción 1260 detecta la necesidad de una extensión, la registra en el Servicio de Emisión de Identificadores 1210, solicitando un identificador único para la nueva extensión. El Servicio de Emisión de Identificadores 1210 envía una consulta al Repositorio Central 1220 para verificar si la extensión solicitada ya existe.

[00156] Cuando existe la extensión solicitada, el Repositorio Central devuelve el identificador solicitado asociado con la extensión existente al Servicio de Emisión de Identificadores 1210, que a su vez reenvía el identificador asociado al cliente de traducción 1260.

- 5 [00157] Cuando la extensión solicitada no existe en el Repositorio Central, el Servicio Emisor de Identificadores 1210 emite un nuevo identificador único para la extensión solicitada al cliente de traducción 1260.

[00158] En algunas modalidades, el conjunto de elementos de traducción de datos incluye, pero no se limita a, uno o más de los siguientes:

- 10
- Conceptos novedosos
 - Conjuntos de valores
 - Relaciones de conceptos novedosos
 - Recursos novedosos
 - Pruebas de calidad de datos
- 15
- Guiones de caracterización de poblaciones
 - Asignaciones de campos
 - Mapas conceptuales

[00159] Los conceptos novedosos son frases cortas que se utilizan en los datos que se están transformando para atribuirles significado cuando dichas frases no existen previamente. Los conceptos se utilizan habitualmente en los datos sanitarios para proporcionar representaciones comunes de diagnósticos, medicamentos, procedimientos y otros conceptos clínicos, así como metadatos.

20

[00160] Existen varias terminologías, algunas con millones de frases, pero durante el mapeo de un conjunto de datos aún pueden surgir nuevas. Por ejemplo, el descubrimiento de un nuevo patógeno puede requerir un nombre único para poder utilizarlo en un diagnóstico o un resultado de laboratorio. Las nuevas terapias requieren descripciones novedosas para referirse a ellas con precisión en los datos.

25

Si los registros de datos en múltiples sitios contienen referencias a un nuevo virus

o terapia, se necesita un concepto novedoso para garantizar la coherencia de dichas referencias y que se representen de la misma manera en todos los sitios y, por lo tanto, en todos los conjuntos de datos.

[00161] Los conjuntos de valores (también conocidos como conjuntos de conceptos)

5 son grupos de conceptos con nombre propio. Si bien normalmente no forman parte de un modelo de datos ni de un conjunto de datos, suelen utilizarse como referencia en el proceso de traducción y en otros elementos, como las pruebas de control de calidad de datos y los guiones de caracterización de poblaciones (véase más adelante). Por lo tanto, la variación en los conjuntos de valores puede dar lugar a la
10 formación de diferentes dialectos. En consecuencia, garantizar que todos los sitios utilicen los mismos conjuntos de valores es fundamental para asegurar la coherencia entre los conjuntos de datos.

[00162] Las relaciones entre conceptos nuevos son aquellas que se establecen entre dos o más elementos descritos por conceptos existentes o nuevos. Estas

15 relaciones se utilizan frecuentemente en traducción para encontrar la mejor opción cuando no existe una traducción perfecta, como por ejemplo, de “Fractura de cuello femoral” a “Estructura del cuello femoral”. Por ejemplo, una terminología estándar podría no diferenciar entre distintos tipos de fracturas, por lo que una frase como “fractura de tibia” podría traducirse como “fractura” por ser la alternativa más
20 cercana y genérica. En este caso, la relación entre “fractura de tibia” en una terminología y “fractura” en otra puede utilizarse para lograr la mejor traducción posible. También es posible emplear múltiples relaciones para pasar de un concepto a otro cuando no existe una relación directa entre conceptos adyacentes. Las relaciones entre conceptos nuevos son aquellas que se establecen entre un
25 concepto nuevo y uno existente, entre dos conceptos nuevos o entre dos conceptos existentes.

[00163] Recursos novedosos: cuando un estándar de datos no cuenta con una forma específica de representar diferentes tipos de datos importantes para múltiples

sitios, se puede crear una columna (o recurso) para almacenarlos. Por ejemplo, OMOP utiliza una tabla de "personas" para registrar la información de los pacientes, pero esta tabla no incluye información de identificación como el nombre y el número de teléfono. Se pueden agregar recursos novedosos, en forma de columnas adicionales, a la tabla de personas para almacenar este tipo de datos adicionales. En otro ejemplo, agregar datos de inventario de farmacia al estándar OMOP requeriría un recurso novedoso en forma de tabla. Los recursos novedosos pueden adoptar otras formas además de los dos ejemplos aquí presentados. Por ejemplo, en FHIR, los recursos novedosos tienen una estructura de subárbol.

[00164] Las pruebas de calidad de datos se utilizan antes, durante y después del proceso de traducción para verificar y comparar los datos antes y después de la traducción. Estas pruebas pueden adoptar diversas formas, como consultas SQL o scripts de Python, y generan un resultado de aprobado o suspenso. En algunos casos, se utilizan cientos de pruebas simultáneamente para validar la calidad de los datos. En otros casos, la calidad de los datos se determina mediante las pruebas y los aspectos de los datos que estas abarcan; por lo tanto, es fundamental que los distintos sitios que colaboran utilicen exactamente las mismas pruebas de calidad de datos. Los sistemas de traducción interpretan y utilizan estas pruebas en sus respectivos sitios de origen.

[00165] Los scripts de caracterización de población se diferencian de las pruebas de calidad de datos en que no generan una calificación de aprobado o suspenso. Los scripts de caracterización de población producen un valor, una distribución u otra indicación sobre un aspecto particular de la población, como la proporción de pacientes masculinos y femeninos en un conjunto de datos específico. Los resultados de los scripts de caracterización de población se utilizan para distinguir la variación en la población de pacientes de la variación dialectal, pero solo si todos los centros caracterizan a la población de la misma manera. Los sistemas de

traducción 1250 interpretan y utilizan los scripts de caracterización de población en sus respectivos sitios de origen.

[00166] Las asignaciones de campos registran cómo se tradujeron los datos de origen. El formato exacto de las asignaciones depende tanto del origen como del destino. Por ejemplo, una asignación de campos para datos de un sistema de historia clínica electrónica a OMOP puede incluir la siguiente asignación: “Los datos del campo “paciente” de la tabla “Encuentro” en el conjunto de datos de origen se asignaron al campo “person_id” de la tabla “visit_occurrence” del estándar. Las asignaciones de campos también pueden ser más complejas e incluir varios campos, condiciones IF/ELSE u otra lógica. Se pueden utilizar diferentes lenguajes de programación para representar las asignaciones de campos. Un ejemplo de lenguaje de programación específico para asignaciones de campos es LinkML (<https://linkml.io/>).

[00167] Los mapas conceptuales relacionan una frase o un concepto codificado de un sistema con un concepto codificado de otro. Al igual que los mapas de campos, los mapas conceptuales pueden representarse formalmente en diferentes lenguajes de programación. Un lenguaje específico para mapas de campos es el Estándar Simple para Compartir Mapeos de Ontologías (SSSOM) (<https://mapping-commons.github.io/sssom/>).

[00168] La representación formal de cada uno de estos elementos puede variar, desde el uso de fragmentos de lenguajes informáticos como Python o SQL, hasta lenguajes específicos para cada tipo de elemento. Sin embargo, todos los sistemas de traducción de datos deben ser capaces de analizar el elemento, ya sea directamente utilizando el mismo lenguaje o mediante otro servicio de traducción que convierta una representación formal a otra.

[00169] El Servicio de Emisión de Identificadores 1210 de la figura 12 es un componente del sistema que conecta cada sitio de origen 1240 con el repositorio central 1220. Su función principal es garantizar que los elementos de traducción de

datos con el mismo significado semántico tengan el mismo identificador único. El Servicio de Emisión de Identificadores puede buscar en el repositorio central 1220 y emitir nuevos identificadores únicos si no existe un elemento equivalente.

[00170] Como se mencionó anteriormente, cuando el software de traducción 1250, que se ejecuta en un sitio de origen, detecta la necesidad de una extensión del estándar, utiliza el cliente de traducción 1260 para enviar una solicitud al Servicio de Emisión de Identificadores 1210, solicitando un identificador único para la nueva extensión. El Servicio de Emisión de Identificadores 1210 verifica si ya existe una extensión equivalente en el repositorio central 1220. Si existe un elemento equivalente, el Servicio de Emisión de Identificadores 1210 recupera el identificador existente del repositorio central 1220 y lo devuelve al cliente de traducción 1260 solicitante, en lugar de emitir uno nuevo. El cliente de traducción 1260 reenvía el identificador único al software de traducción 1250.

[00171] Alternativamente, si no existe una extensión equivalente en el repositorio central 1220, el Servicio de Emisión de Identificadores 1210 genera un nuevo identificador único para la nueva extensión, por ejemplo, mediante un generador de números aleatorios o similar, y devuelve el nuevo identificador único al cliente de traducción solicitante 1260. El Servicio de Emisión de Identificadores 1210 también transmite el identificador único recién generado al repositorio central 1220 para su almacenamiento asociado a la nueva extensión.

[00172] Algunas implementaciones del Servicio de Emisión de Identificadores también pueden sugerir elementos similares, aunque no idénticos al elemento buscado, mediante coincidencia aproximada, inteligencia artificial u otros métodos. Algunas implementaciones del Servicio de Emisión de Identificadores pueden integrarse como parte del repositorio.

Aplicación industrial

[00173] Las disposiciones descritas son aplicables a las industrias de la investigación, la medicina y la salud.

5 [00174] Lo anterior describe únicamente algunas modalidades de la presente invención, y se pueden realizar modificaciones y/o cambios sin apartarse del alcance y el espíritu de la invención, siendo las modalidades ilustrativas y no restrictivas.

[00175] En esta descripción, las referencias a “una modalidad”, “una modalidad”, “algunas modalidades” o “modalidades” significan que una característica, estructura
10 o rasgo particular descrito en relación con la modalidad se incluye en al menos una modalidad de la presente invención. Por lo tanto, las apariciones de las frases “en una modalidad” en distintos lugares de esta memoria descriptiva no se refieren necesariamente a la misma modalidad, sino que pueden hacerlo. Además, las características, estructuras o rasgos particulares pueden combinarse de cualquier
15 manera adecuada, como resultaría evidente para un experto en la materia a partir de esta descripción, en una o más modalidades.

[00176] Si bien algunas modalidades descritas aquí incluyen algunas características presentes en otras modalidades, pero no otras, se entiende que las combinaciones de características de diferentes modalidades están comprendidas dentro del
20 alcance de la invención y constituyen modalidades distintas, tal como lo comprenderían los expertos en la materia. Por ejemplo, en las siguientes reivindicaciones, cualquiera de las modalidades reivindicadas puede utilizarse en cualquier combinación.

[00177] Además, algunas de las modalidades se describen aquí como un método o
25 una combinación de elementos de un método que puede ser implementado por un procesador de un sistema informático o por otros medios para llevar a cabo la función. Así, un procesador con las instrucciones necesarias para llevar a cabo dicho método o elemento constituye un medio para su ejecución. Así mismo, un

elemento descrito en este documento, perteneciente a una modalidad de aparato, es un ejemplo de un medio para llevar a cabo la función que realiza dicho elemento con el fin de implementar la invención.

[00178] En la presente descripción se detallan varios aspectos específicos. No obstante, se entiende que algunas modalidades de la invención pueden llevarse a cabo sin necesidad de estos detalles. En otros casos, no se han mostrado en detalle métodos, estructuras y técnicas bien conocidos para facilitar la comprensión de esta descripción.

[00179] Téngase en cuenta que cuando se describe un método que incluye varios elementos, por ejemplo, varios pasos, no se implica ningún orden de dichos elementos, por ejemplo, de dichos pasos, a menos que se indique específicamente.

[00180] En el contexto de esta descripción, la palabra “que comprende” y sus construcciones gramaticales asociadas significan “que incluye principalmente, pero no necesariamente de forma exclusiva”, “que tiene” o “que incluye”, y no “que consiste únicamente en”. Las variantes de la palabra “que comprende”, como “comprender” y “comprende”, tienen significados correspondientemente variados.

[00181] De igual modo, cabe señalar que el término “acoplado” no debe interpretarse como limitante a conexiones directas. Se pueden utilizar los términos “acoplado” y “conectado”, junto con sus derivados. Debe entenderse que estos términos no son sinónimos, aunque pueden serlo. Por lo tanto, el alcance de la expresión “un dispositivo A acoplado a un dispositivo B” no debe limitarse a dispositivos o sistemas donde una entrada o salida del dispositivo A esté conectada directamente a una entrada o salida del dispositivo B. Significa que existe una ruta entre el dispositivo A y el dispositivo B, la cual puede incluir otros dispositivos o medios intermedios. Además, “acoplado a” no implica dirección. Por consiguiente, la expresión “un dispositivo A está acoplado a un dispositivo B” puede ser sinónima de la expresión “un dispositivo B está acoplado a un dispositivo A”. “Acoplado” puede significar que dos o más elementos están en contacto físico o eléctrico

directo, o que dos o más elementos no están en contacto directo entre sí, pero aun así cooperan o interactúan.

[00182] En toda la presente descripción, a menos que se especifique lo contrario, el uso de adjetivos ordinales como “primero”, “segundo”, “tercero”, “cuarto”, etc., para describir objetos comunes o relacionados, indica que se hace referencia a diferentes instancias de esos objetos comunes o relacionados, y no pretende implicar que los objetos así descritos deban proporcionarse o colocarse en un orden o secuencia determinados, ya sea temporal, espacial, jerarquizado o de cualquier otra manera.

[00183] Aunque la invención ha sido descrita con referencia a ejemplos específicos, los expertos en la materia comprenderán que la invención puede materializarse en muchas otras formas.

**SISTEMA Y MÉTODO IMPLEMENTADO POR COMPUTADOR PARA
ESTUDIOS FEDERADOS SEGUROS**

(ORIGINAL)

Solicitud relacionada

[00184] La presente solicitud está relacionada con la solicitud de patente provisional australiana No. 2023900994 titulada “SISTEMA Y MÉTODO IMPLEMENTADO POR COMPUTADOR PARA ESTUDIOS FEDERADOS SEGUROS” y presentada el 5 de abril de 2023 a nombre de Evidentli Pty Ltd, cuyo contenido completo se incorpora aquí por referencia como si se reprodujera íntegramente en este documento.

Campo técnico

[00185] La presente divulgación se refiere a un sistema y métodos asociados para estudios federados confiables y seguros. En particular, se refiere a un sistema informático y métodos relacionados para su uso en estudios federados confiables y

seguros en el contexto de estudios clínicos, estudios observacionales y otras actividades de investigación.

Antecedentes

[00186] Los estudios federados son ensayos clínicos, estudios observacionales y otros tipos de actividades de investigación que se llevan a cabo en múltiples ubicaciones físicas (centros), como hospitales, laboratorios, universidades, instituciones de investigación, etc. Los centros donde se realizan los estudios se denominan aquí sitios de origen, y el centro donde se analizan los datos se denomina aquí sitio de análisis. Una misma ubicación puede ser tanto sitio de origen como sitio de análisis.

[00187] Las ventajas de los estudios federados incluyen, entre otras:

- Los participantes en un estudio federado representan una muestra de la población mayor que la que podría representar cualquier centro individual.
- La logística de realizar cada estudio es más sencilla que la de realizar un único estudio de gran envergadura, y
- La heterogeneidad de los propios sitios representa una muestra más amplia de sitios.

[00188] En un estudio federado, cada sitio de origen realiza de forma independiente su propia versión del estudio, y los datos de todos los centros se analizan conjuntamente en un sitio de análisis para obtener un resultado combinado. Si bien existen varios métodos para realizar estudios federados, cada uno presenta limitaciones que impiden su adopción universal. Por ejemplo, las distintas jurisdicciones pueden definir los diagnósticos de manera diferente o utilizar parámetros o rangos distintos en relación con los diagnósticos o las afecciones.

[00189] Independientemente de los métodos matemáticos empleados para combinar los resultados de múltiples centros en un estudio federado, este tipo de estudio puede presentar problemas de coherencia entre los centros participantes.

Las variaciones en la producción y recopilación de datos, así como en los materiales utilizados, pueden generar imprecisiones inexplicables en los resultados. Además, las restricciones no técnicas, como los requisitos legales jurisdiccionales en materia de privacidad y transmisión de datos, pueden afectar la recopilación de datos.

- 5 [00190] Por lo tanto, existe la necesidad de proporcionar un sistema mejorado para la realización de estudios federados.

Resumen

[00191] La presente divulgación se refiere a un sistema implementado por computador y a los métodos asociados para realizar estudios federados.

- 10 [00192] Un primer aspecto de la presente divulgación proporciona un sistema de estudios federados implementado por computador que comprende:

un modelo de datos común predefinido asociado a un estudio;

- un conjunto de sitios de origen, estando cada sitio de origen asociado a un conjunto de dispositivos de entrada para la adquisición de datos de estudio de los sujetos y un dispositivo de datos de estudio, donde cada dispositivo de datos de estudio incluye:
- 15

una base de datos nativa para almacenar datos de estudio asociados con el sitio de origen respectivo y derivados de dicho conjunto de dispositivos de entrada;

- 20 un traductor de modelos de datos para traducir los datos de estudio almacenados a un formato que corresponda a dicho modelo de datos común;

una base de datos de modelo común para almacenar, en dicho modelo de datos común, los valores individuales de los participantes derivados de los datos del estudio;

- 25 un módulo de cálculo de estudio para determinar los valores agregados del sitio a partir de los valores individuales de los participantes; y

una base de datos agregada para almacenar los valores agregados de dicho sitio; y

un sitio de análisis conectado a cada sitio de origen mediante una red de comunicaciones, donde dicho sitio de análisis está asociado a un dispositivo de análisis de estudio que incluye:

5 un módulo de cálculo de análisis para procesar, de acuerdo con un perfil de análisis predefinido, los valores agregados del sitio recibidos de dichos sitios de origen; y

un módulo de informes para generar informes de estudio basados en los valores agregados del sitio procesados.

[00193] Un segundo aspecto de la presente divulgación proporciona un método para
10 llevar a cabo un estudio federado en varios sitios de origen y utilizando un sitio de análisis, dicho método comprende los siguientes pasos:

definir un modelo de datos común para el estudio;

definir un conjunto de instrucciones de análisis para el estudio;

en cada sitio de origen:

15 adquirir, utilizando al menos un dispositivo de entrada, datos de estudio de un conjunto de sujetos;

traducir dichos datos de estudio para generar valores individuales de los participantes almacenados en dicho modelo de datos común;

20 procesar dichos valores individuales de los participantes para generar un conjunto de valores agregados del sitio;

transmitir dichos valores agregados del sitio al sitio de análisis; y

procesar los valores agregados del estudio recibidos del sitio de análisis de acuerdo con las instrucciones de análisis traducidas; y

en dicho sitio de análisis:

25 traducir dichas instrucciones de análisis;

procesar los valores agregados del sitio recibidos de al menos uno de dichos sitios de origen de acuerdo con dichas instrucciones de análisis traducidas; y

generar un informe basado en los valores agregados del sitio procesados.

[00194] Según otro aspecto, la presente divulgación proporciona un aparato para implementar cualquiera de los métodos anteriormente mencionados.

- 5 [00195] Según otro aspecto, la presente divulgación proporciona un producto de programa informático que incluye un medio legible por computador que tiene grabado un programa informático que, cuando se ejecuta en un procesador de computador, implementa cualquiera de los métodos descritos anteriormente.

[00196] También se proporcionan otros aspectos de la presente divulgación.

10 **Breve descripción de las figuras**

[00197] A continuación se describirán una o más modalidades de la presente divulgación mediante ejemplos específicos con referencia a las figuras adjuntas, en las que:

- [00198] La figura 1 es una representación esquemática mediante diagrama de bloques de un sistema de estudio federado implementado por computador, de acuerdo con una modalidad de la presente divulgación;

[00199] La figura 2 es una representación esquemática mediante diagrama de bloques de un sitio de análisis de acuerdo con una modalidad de la presente divulgación;

- 20 [00200] La figura 3 es una representación esquemática en diagrama de bloques de un Sitio de Origen de acuerdo con una modalidad de la presente divulgación;

[00201] La figura 4 es una representación esquemática mediante diagrama de bloques de un sistema de estudios federados en el que se pueden poner en práctica una o más modalidades de la presente descripción;

- 25 [00202] La figura 5 es una representación esquemática mediante diagrama de bloques de un sistema que incluye un computador de propósito general en el cual se pueden poner en práctica una o más modalidades de la presente descripción;

[00203] La figura 6 es una representación esquemática mediante diagrama de bloques de un sistema que incluye un teléfono inteligente general en el que se pueden poner en práctica una o más modalidades de la presente descripción;

[00204] La figura 7 es un diagrama de secuencia que ilustra un flujo de trabajo de práctica ejemplar de un estudio federado realizado por una modalidad del sistema de la presente divulgación;

[00205] La figura 8 es una representación esquemática mediante diagrama de bloques de un sitio de origen implementado utilizando una arquitectura informática basada en la nube;

[00206] La figura 9 es una representación esquemática mediante diagrama de bloques de un sistema de estudios federados implementado utilizando una arquitectura informática basada en la nube;

[00207] La figura 10 es una representación esquemática mediante diagrama de bloques de un sistema de estudios federados implementado utilizando una arquitectura informática basada en la nube;

[00208] La figura 11 es una representación esquemática mediante diagrama de bloques de un sistema de estudios federados implementado utilizando una arquitectura informática basada en la nube; y

[00209] La figura 12 es una representación esquemática en diagrama de bloques que ilustra una porción de una red de estudios federados que cuenta con un sitio de origen, un servicio emisor de identificadores y un repositorio central.

[00210] Los pasos o características del método en las figuras adjuntas que tengan los mismos números de referencia se considerarán que tienen las mismas funciones u operaciones, a menos que se exprese o se dé a entender la intención contraria.

Glosario

Sitio de Origen

[00211] Un sitio que almacena datos sensibles sobre personas que se utilizarán en un proyecto de investigación sin transmitirlos fuera del sitio. Esta

descripción se refiere a múltiples sitios de origen, denominados aquí $Sitio_1$, $Sitio_2$, etc., hasta el $Sitio_s$. En general, cada sitio se denomina $Sitio_i$.

Sitio de Análisis

[00212] El sitio de análisis se denomina en el presente documento como $Sitio_A$.

- 5 El Sitio de Análisis puede coincidir (estar ubicado en el mismo lugar o ser parte integral de) uno de los sitios de origen o puede ser una entidad separada e independiente.

Valor para el Participante Individual: Y_j

- [00213] Uno o más valores numéricos sobre una persona (persona j), que
10 potencialmente podrían usarse para identificar o exponer información sensible sobre la persona j .

Valor Agregado: Y

- [00214] Un elemento de datos que contiene información sobre un grupo de personas y que no puede utilizarse para identificar ni revelar información
15 confidencial sobre una o varias personas. Los valores agregados se pueden transmitir de forma segura entre sitios. Ejemplos de valores agregados son el recuento de participantes y los promedios de los valores individuales de los participantes. Según el escenario y la implementación específicos, los valores agregados pueden intercambiarse entre sitios varias veces durante un mismo
20 estudio federado.

Agregado del Sitio: $\hat{Y}_i$

- [00215] Un valor agregado que se refiere a la información disponible para un sitio de origen. El símbolo de un agregado de sitio va seguido del subíndice i , por ejemplo, x_i . Un ejemplo de agregado de sitio es el número de participantes en el
25 sitio (n_i).

Estudio Agregado: $\hat{Y}$

- [00216] Valor agregado que engloba la información disponible para un estudio completo. El símbolo de un valor agregado del estudio no va seguido de un

subíndice, por ejemplo, x_i . Un ejemplo de valor agregado del estudio es el número de participantes en el estudio completo (N).

Tarea

[00217] Un agregado de estudio que $Sitio_A$ tiene la tarea de calcular (es decir, una fórmula que necesita datos del sitio para ser calculada).

Número de Participantes en el Estudio: N

[00218] Un agregado de estudio especial que representa el número de participantes en todo el estudio.

Número de Participantes en el Sitio: n_i

[00219] Un agregado especial del sitio que es el número de participantes en el sitio.

Número de Sitios: s

[00220] El número total de sitios de origen que participan en un estudio.

Descripción detallada

[00221] La presente descripción proporciona un sistema implementado por computador y los métodos asociados para realizar estudios federados. El sistema incluye un conjunto de sitios de origen y un conjunto de sitios de análisis. En algunas implementaciones, los sitios de análisis están separados e independientes de los sitios de origen, estando cada sitio de origen conectado a un sitio de análisis mediante un enlace de comunicaciones que permite la transmisión de datos desde cada sitio de origen al sitio de análisis correspondiente. En otras implementaciones, uno o más sitios de análisis están ubicados en el mismo lugar o integrados en un sitio de origen, y uno o más enlaces de comunicaciones permiten la transmisión de datos entre los sitios de origen y un sitio de análisis en los casos en que un sitio de análisis no esté ubicado en el mismo lugar o integrado en un sitio de origen.

[00222] Una configuración básica utiliza un único sitio de análisis. Otras configuraciones pueden utilizar varios sitios de análisis. Por ejemplo, una configuración se refiere a un escenario de red con subredes jerárquicas de sitios de

origen y sitios de análisis, donde estos últimos envían datos a un sitio de análisis central. Esta disposición jerárquica proporciona un conjunto en cascada de sitios de origen y sitios de análisis.

[00223] Una posible implementación podría relacionarse, por ejemplo, con un sitio de análisis centralizado administrado por una empresa farmacéutica. Las subredes se relacionan con los distintos distritos sanitarios que administran sitios de análisis distritales, y cada sitio de análisis distrital recibe datos de los hospitales asociados a su distrito sanitario. En dicha implementación, los sitios de análisis distritales se convierten en sitios de origen para el sitio de análisis centralizado.

[00224] Para mayor claridad, las modalidades descritas aquí se referirán generalmente a un único nivel jerárquico en el que un conjunto de sitios de origen informa a un único sitio de análisis. No obstante, un experto en la materia comprenderá que el alcance de la invención aquí descrita abarca modalidades alternativas con múltiples niveles de análisis.

[00225] Para mayor claridad, las modalidades descritas aquí presuponen una configuración de conexiones de red simples. Sin embargo, un experto en la materia comprenderá que el alcance de la invención aquí descrita abarca otras configuraciones de redes que pueden incluir, por ejemplo, transmisión (por ejemplo, enrutadores), seguridad (por ejemplo, cortafuegos) y aceleración (por ejemplo, servidores caché), entre otras.

[00226] Los enfoques actuales para estudios federados presentan la dificultad técnica de recopilar datos de manera consistente en diferentes sitios de origen, de modo que estos puedan procesarse de forma uniforme. Además, existen dificultades técnicas para garantizar el anonimato de los datos personales recopilados en dichos sitios.

[00227] El sistema descrito en la presente divulgación garantiza la compatibilidad de datos entre los sitios de origen para asegurar la fiabilidad de los resultados. En particular, el sistema incluye uno o más pasos que transfieren datos de múltiples

fuentes para su agregación y estandarización conforme a un modelo de datos común predefinido. En algunas implementaciones, la transferencia de datos puede incluir traducción, cuando sea necesario. En algunas implementaciones, el sistema utiliza una entrada de datos inteligente que permite agregar y estandarizar datos de múltiples fuentes conforme a un modelo de datos común predefinido. Un modelo de datos se ajusta al modelo de datos común predefinido si dicho modelo puede ser utilizado por un sitio de origen para generar un agregado de datos comprensible para el sitio de análisis.

[00228] En algunas implementaciones, los sitios de origen y el sitio de análisis comparten un modelo de datos común coordinado. Este modelo puede ser predefinido por el administrador del estudio o bien coordinado de otra manera entre los distintos sitios de origen y el sitio de análisis. En dichas implementaciones, los datos transmitidos desde un sitio de origen al sitio de análisis han sido preparados por el sitio de origen correspondiente para ajustarse al modelo de datos común coordinado. En otras implementaciones, cada sitio de origen utiliza un modelo de datos, que puede o no ajustarse al modelo de datos común coordinado, y el sitio de análisis dispone de un servicio de traducción que puede convertir, cuando sea necesario, la información agregada de cada sitio de origen a un modelo de datos común.

[00229] En algunas implementaciones, el modelo de datos común es el Modelo de Datos Comunes de la Asociación de Resultados Médicos Observacionales (OMOP CDM u OMOP para abreviar). También se pueden utilizar Oracle Health Foundations, National Patient-Centered Clinical Research Network (PCORNet), Sentinel e Informatics for Integrating Biology & the Bedside (I2B2), o un modelo de datos definido por el usuario.

[00230] En algunas implementaciones, el sistema proporciona gobernanza de datos que aplica y audita los permisos de acceso a los datos. El sistema no transmite

datos relativos a participantes individuales, lo que resuelve las preocupaciones legales y de privacidad.

[00231] En algunas implementaciones, el sistema proporciona una interfaz de usuario en forma de panel de control que ofrece a los usuarios análisis clínicos de autoservicio, lo que permite a usuarios con diferentes niveles de conocimientos técnicos consultar datos de investigación y generar informes.

Agrupación de Datos

[00232] La recopilación de datos de los centros donde se registran y su transmisión a un centro donde se analizarán se denomina “Agrupación de Datos”. La agrupación de datos se considera el estándar actual para obtener los resultados más fiables de estudios federados. Sin embargo, la transmisión de datos individuales de participantes entre centros supone un riesgo para la privacidad de los pacientes y puede estar sujeta a controles legales.

[00233] Cabe destacar que la limitación no se restringe a la protección criptográfica de los datos en tránsito o en reposo, ni a la eliminación fiable de la información de identificación personal. El problema radica en que la mayoría de las jurisdicciones tienen restricciones legales que prohíben la transmisión y el almacenamiento de datos, y estas restricciones no técnicas son imposibles de superar únicamente con tecnologías como el cifrado. En consecuencia, la agrupación de datos suele limitarse a un único sitio, lo que puede impedir obtener tamaños de muestra estadísticamente útiles.

Metaanálisis

[00234] El metaanálisis es un método que evita el envío de datos individuales de los participantes y reduce el riesgo para su privacidad al transmitir únicamente información agregada calculada en cada centro. Estos datos agregados se combinan en un centro común que genera el resultado del estudio. Sin embargo, para la mayoría de los valores, estos cálculos solo pueden aproximarse a los resultados que se obtendrían mediante la agrupación de datos.

[00235] En las técnicas de metaanálisis se utilizan diversas técnicas de aproximación para combinar datos de sitios web. Cada técnica de aproximación se basa en ciertas suposiciones, que a menudo no se especifican explícitamente. No incluir los supuestos tiene varias consecuencias negativas, entre ellas: (i) dificultar la comprensión del informe; (ii) no transmitir las intenciones del autor; y/o (iii) obligar al lector a inferir dichos supuestos a partir de los métodos empleados, lo cual puede dar lugar a errores. No indicar los supuestos asociados a las técnicas de aproximación puede llevar al lector a interpretar las aproximaciones como valores exactos, lo cual es erróneo.

10 Agregados de Datos

[00236] Algunos métodos para realizar estudios federados utilizan agregados de datos de diferentes centros de estudio. Ejemplos de datos agregados son el número de participantes fallecidos durante el periodo de estudio, la mediana de la duración de la estancia de una cohorte de participantes y la reducción del riesgo absoluto entre dos grupos de estudio. Los datos agregados son fáciles de calcular en cada centro donde se almacenan y se transmiten desde los centros con los datos originales (sitios de origen), en lugar de los datos sensibles, al centro remoto donde se realiza el análisis (sitio de análisis).

[00237] Algunos datos agregados pueden calcularse con exactitud en el lugar del análisis, obteniéndose resultados idénticos a los que se habrían obtenido mediante la agrupación de datos. Por ejemplo, para conocer el número total de pacientes del estudio, se puede sumar el número de pacientes de cada centro.

[00238] Otros tipos de datos agregados no se pueden calcular con exactitud a partir de otros agregados de datos. Por ejemplo, para calcular con exactitud la reducción del riesgo absoluto entre dos grupos experimentales, se calcula el promedio de cada centro y se envía al sitio de análisis. Para calcular la reducción del riesgo absoluto promedio en toda la población del estudio, es necesario realizar ciertas suposiciones. Por ejemplo, la reducción del riesgo absoluto para cada centro debe

ponderarse según el número de participantes en cada centro o, más comúnmente, según la homogeneidad de los pacientes en cada centro. Con estas suposiciones, la reducción del riesgo absoluto calculada es solo una aproximación del valor que se habría obtenido mediante la agrupación de datos.

5 *Sistema Implementado por Computador y Método Asociado para la Realización de Estudios Federados*

[00239] Independientemente de los métodos matemáticos empleados para combinar los resultados de múltiples centros en un estudio federado, los métodos existentes para realizar dichos estudios pueden presentar problemas de
10 consistencia entre los sitios de estudio debido a las suposiciones generales que se realizan para correlacionar los datos. La presente descripción proporciona un sistema implementado por computador que permite realizar estudios federados en un conjunto de sitios de estudio sin necesidad de recurrir a suposiciones generales.

[00240] La figura 1 es un diagrama de bloques esquemático de un sistema
15 implementado por computador para estudios federados 100, de acuerdo con una modalidad de la presente descripción. El sistema 100 incluye un conjunto 105 de Sitios de Origen 110a a 110s. Cada Sitio de Origen 110a a 110s es una ubicación donde se adquieren datos para un estudio e incluye un dispositivo informático asociado conectado a una red de comunicaciones que permite la comunicación
20 entre cada Sitio de Origen 110a a 110s y un Sitio de Análisis 150. La red de comunicaciones puede constar de uno o más enlaces de comunicaciones cableados, enlaces de comunicaciones inalámbricos o cualquier combinación de estos. En particular, la red de comunicaciones puede incluir una red de área local (LAN), una red de área extensa (WAN), una red de telecomunicaciones o cualquier
25 combinación de estas. Una red de telecomunicaciones puede incluir, entre otras cosas, una red de telefonía, como una red telefónica pública conmutada (RTPC) o una red de telefonía móvil celular, Internet o cualquier combinación de las mismas.

[00241] El Sitio de Análisis 150 es una ubicación central para analizar los datos derivados de estudios realizados en uno o más de los Sitios de Origen 110a a 110s. El Sitio de Análisis 150 está conectado a un Repositorio de Análisis 120, el cual es una base de datos que almacena diversos análisis que pueden ser realizados y una forma de análisis se selecciona con base en el estudio que se esté realizando. Cada análisis consiste en un conjunto de instrucciones informáticas que especifican cómo realizar un análisis de datos en particular, y cada estudio se asocia con uno de estos análisis. Estas instrucciones informáticas para realizar análisis de datos pueden representarse en un programa informático, código fuente u otro documento o código electrónico que un computador puede interpretar como una serie de operaciones que, al ejecutarse sobre un conjunto de datos del estudio, generan uno o más Valores Agregados significativos.

[00242] El Sitio de Análisis 150 puede obtener instrucciones de análisis de datos para un conjunto de datos determinado de diversas maneras. En algunas implementaciones, el Repositorio de Análisis 120 mantiene un enlace entre el Sitio de Análisis 150 y los Sitios de Origen 110a a 110s. En otras implementaciones, las instrucciones para el Sitio de Análisis 150 incluyen una referencia a las instrucciones para los Sitios de Origen 110a a 110s.

[00243] En otras modalidades, el sistema 100 incluye un administrador de análisis (no mostrado) que coordina las instrucciones de análisis entre los Sitios de Origen 110a ... 110s y el Sitio de Análisis 150. El administrador de análisis puede implementarse utilizando un dispositivo informático programado o un operador humano o una combinación de ambos.

[00244] En otras modalidades, se utilizan otros métodos de sincronización, como el Acuerdo Bizantino, para garantizar que el Sitio de Análisis 150 utilice un conjunto correcto de instrucciones de análisis del Repositorio de Análisis 120.

[00245] La figura 2 muestra una vista ampliada de los módulos funcionales de una modalidad del Sitio de Análisis 150. El Sitio de Análisis 150 incluye un dispositivo de análisis de datos 151 implementado por computador que está configurado para:

- Comunicarse con cada base de datos de valores agregados del sitio en cada Sitio de Origen
- Interpretar las instrucciones representadas en un análisis del Repositorio de Análisis en un algoritmo
- Ejecutar un algoritmo interpretado a partir de un análisis del Repositorio de Análisis
- Reportar los resultados del estudio

[00246] En el contexto de la figura 2, el dispositivo de análisis de datos 151 recibe la comunicación 205 de uno o más de los Sitios de Origen 110a a 110s o del Repositorio de Análisis 120. Las comunicaciones entre el dispositivo de análisis de datos 151 y el Repositorio de Análisis 120 pueden estar dirigidas a transmitir un conjunto de instrucciones de análisis desde el Repositorio de Análisis 120 al dispositivo de análisis de datos 151. Dichas instrucciones de análisis pueden entonces transmitirse desde el dispositivo de análisis de datos 151 a uno o más de los Sitios de Origen 110a a 110s.

[00247] En algunas modalidades, la comunicación 205 se refiere a un enlace proporcionado por, o recuperado de, el Repositorio de Análisis 120 y enviado al dispositivo de análisis de datos 151 para reenviarlo a uno o más de los Sitios de Origen 110a ... 110s, de modo que el Sitio de Origen 110a ... 110s respectivo pueda utilizar el enlace para recuperar un conjunto apropiado de instrucciones de análisis directamente del Repositorio de Análisis 120.

[00248] La comunicación 205 también puede referirse al intercambio de información entre el dispositivo de análisis de datos 151 y uno o más de los Sitios de Origen 110a ... 110s, en cuyo caso la información puede incluir, por ejemplo, solicitudes de datos agregados o la transmisión de datos agregados relacionados con un estudio.

[00249] La comunicación recibida 205 es procesada por un módulo de cálculo 152 que interpreta las instrucciones recibidas del Repositorio de Análisis y aplica esas instrucciones para analizar los datos recibidos de uno o más Sitios de Origen 110a ... 110s. Los datos analizados se presentan luego a un módulo de informes 154 que
5 informa los resultados del estudio.

[00250] Según la implementación, el módulo de informes 154 presenta los resultados del estudio a través de una interfaz de usuario (como un panel de control), imprimiéndolos, enviándolos por correo electrónico, mostrándolos en un sitio web, generando un documento electrónico o mediante cualquier combinación
10 de estos métodos. El sistema es configurable para seleccionar un conjunto de usuarios a quienes se envían los resultados del estudio o que están autorizados a acceder a ellos. En algunas implementaciones, el nivel de autorización asignado a un usuario determina el nivel de detalle de los resultados del estudio a los que tiene acceso. Por ejemplo, los usuarios con alto nivel de autorización pueden tener
15 acceso a todos los niveles de los resultados del estudio, mientras que los usuarios con bajo nivel de autorización pueden tener acceso solo a un nivel general de los resultados o a un subconjunto de los mismos.

[00251] La figura 3 es un diagrama de bloques esquemático del Sitio de Origen 110a de la figura 1 y representa cualquiera de los Sitios de Origen 110a a 110s. El Sitio
20 de Origen 110a incluye un dispositivo de datos de estudio computarizado 115a conectado a un conjunto de dispositivos de entrada 111a a 111d para la adquisición de datos de estudio de uno o más sujetos. Los dispositivos de entrada 111a a 111d pueden implementarse utilizando cualquier dispositivo adecuado para la captura digital de datos que se puedan usar en un estudio federado. Los dispositivos de
25 entrada 111a a 111d pueden incluir, por ejemplo, sistemas de registros médicos electrónicos, sistemas de administración de pacientes, sistemas computarizados de prescripción médica, sistemas de encuestas, dispositivos portátiles, sensores de laboratorio, equipos de imagen, termómetros electrónicos, etc. Los dispositivos

portátiles pueden incluir, por ejemplo, pero no se limitan a, monitores de frecuencia cardíaca, monitores de presión arterial, monitores de electrocardiograma (ECG), auriculares de electroencefalografía (EEG), oxímetros de pulso y similares.

[00252] Para los estudios federados relacionados con la atención médica, los dispositivos de entrada pueden incluir cualquier cosa que genere o capture datos en un entorno sanitario, y dichos dispositivos no se limitan a la captura de datos clínicos, sino que también pueden incluir dispositivos que capturen, adquieran, almacenen o transmitan datos no clínicos, como detalles de seguros, datos bancarios, datos meteorológicos relativos a un sitio de origen (por ejemplo, temperatura, presión barométrica, humedad), datos geográficos relativos a un sitio de origen (por ejemplo, altitud), etc.

[00253] Según la naturaleza de los dispositivos de entrada 111a a 111d, estos pueden requerir la introducción manual de datos por parte del usuario, estar conectados al dispositivo de datos de estudio 115a (implementado por computador) o utilizar una combinación de introducción y/o transferencia de datos manual y automatizada. Por ejemplo, los monitores de EEG o ECG pueden configurarse para transmitir datos directamente al dispositivo de datos de estudio 115a (implementado por computador) a medida que se obtienen datos del paciente, mientras que los sistemas de administración de pacientes y las historias clínicas electrónicas probablemente requieran al menos cierta introducción manual de datos antes de su transmisión al dispositivo de datos de estudio 115a.

[00254] El dispositivo de datos de estudio 115a incluye una base de datos nativa 112 que almacena los valores individuales de los participantes, obtenidos de uno o más de los dispositivos de entrada 111a a 111d asociados al sitio de origen correspondiente, que en el ejemplo de la figura 3 es el sitio de origen 110a. En dicho ejemplo, la base de datos nativa 112 está integrada en el dispositivo de datos de estudio 115a. En otras implementaciones, la base de datos nativa 112 puede estar integrada en el sitio de origen correspondiente, ubicada en el mismo sitio o ubicada

remotamente y conectada al sitio de origen, por ejemplo, mediante una plataforma en la nube. Los valores individuales de los participantes para cada sitio de origen pueden almacenarse en cualquier modelo de datos apropiado, según la implementación. Así mismo, los valores individuales de los participantes para diferentes estudios realizados en un mismo sitio de origen pueden almacenarse en el mismo modelo de datos o en modelos diferentes.

[00255] Un único estudio puede involucrar uno o más conjuntos de datos diferentes.

Cada conjunto de datos puede estar asociado a un modelo de datos distinto.

Alternativamente, uno o más conjuntos de datos pueden compartir un modelo de

datos. En este sentido, se puede proporcionar un traductor para convertir los diferentes conjuntos de datos a un modelo de datos común. Volviendo a la Fig. 3,

los Valores Individuales de los Participantes se capturan desde uno o más de los dispositivos de entrada 111a ... 111b y luego se almacenan en la Base de Datos Nativa 112. Un traductor 114 convierte los Valores Individuales de los Participantes

desde el/los modelo(s) de datos utilizados en la(s) Base(s) de Datos Nativa(s) a un

Modelo de Datos Común, donde el Modelo de Datos Común es un modelo de datos predefinido común a todos los Sitios de Origen 110a ... 110s y al Sitio de Análisis

150. Los valores traducidos se almacenan, en el formato del Modelo de Datos Común, en una Base de Datos de Valores Individuales de los Participantes en un

Modelo de Datos Común 116.

[00256] Un módulo de cálculo de origen 118 interpreta las instrucciones

representadas en un análisis del Repositorio de Análisis y las convierte en un algoritmo. A continuación, el módulo de cálculo de origen 118 ejecuta el algoritmo

interpretado en relación con la Base de Datos de Valores Individuales de Participantes en un Modelo de Datos Común 116 y escribe el resultado del algoritmo

ejecutado en una Base de Datos de Valores Agregados del Sitio 119. Esta Base de Datos de Valores Agregados del Sitio se comunica, a través de una red de

comunicaciones, con el dispositivo de análisis de datos 151 del Sitio de Análisis 150.

[00257] Durante el transcurso de un estudio, los valores agregados pueden intercambiarse repetidamente entre uno o más de los sitios de origen 110a...s y el sitio de análisis 150, a medida que se calculan. La naturaleza de los valores agregados que se envían y el momento de su envío se definen y coordinan mediante conjuntos de instrucciones de envío y recepción incluidas en las instrucciones de análisis. Dichas instrucciones de análisis pueden estar contenidas en un perfil de análisis predefinido que forma parte de las instrucciones de análisis definidas para un estudio en particular.

Modelo de Datos Común

[00258] Se supone que cada sitio de origen tiene los valores de participantes individuales necesarios para calcular los agregados del sitio según un algoritmo, denominado aquí Análisis_O. Sin embargo, los valores de participantes individuales se pueden representar de varias maneras (modelos de datos), incluso de una manera que sea única para ese sitio en particular.

[00259] El sitio de análisis debe acordar con cada sitio de origen un modelo de datos común para su uso en las comunicaciones. Este modelo puede ser predefinido o específico para el estudio.

[00260] El sistema de traducción del modelo de datos en cada sitio de origen, a veces conocido como extracción, transformación y carga (ETL), se configura específicamente para el sitio de origen y sus bases de datos nativas. Sin embargo, la salida del traductor del modelo de datos en cada sitio de origen representa los mismos valores individuales de los participantes en el modelo de datos común.

Precondiciones

- Se identifica un sitio de análisis. Para conveniencia, en adelante se hará referencia a dicho sitio como Sitio_A.
- Se identifican uno o más sitios de origen. A cada sitio de origen se le asigna un nombre único para que el sitio de análisis pueda referirse a él. Para mayor claridad, en este documento los sitios de origen se denominan Sitio₁, Sitio₂,

etc., donde el último sitio es Sitio_S, y a cada sitio de origen se le denomina genéricamente Sitio_i.

- El modelo de datos común es utilizado tanto por el Sitio_A como por el Sitio_i.
- Se configura uno o más traductores de modelos de datos en cada Sitio_i.
- 5 • Es posible que uno de los sitios de origen coincida con el Sitio_A, pero no se supone que este sea el caso.
- Los permisos y el acceso a la red están configurados de la siguiente manera:
 - El sitio de análisis puede acceder a los datos agregados de cada sitio de origen.
 - 10 ○ El sitio de análisis puede acceder y descargar análisis del Repositorio de Análisis.
 - Cada sitio de origen puede:
 - Acceder y descargar análisis del Repositorio de Análisis según las indicaciones del sitio de análisis, o
 - 15 ■ se enviarán análisis desde el sitio de análisis
- El Repositorio de Análisis contiene un análisis que indica al sitio de origen cómo calcular uno o más Agregados del Sitio. Para mayor comodidad, en adelante nos referiremos a este análisis como Análisis_O, independientemente del número de Valores Agregados que calcule.
- 20 • Análisis_O puede utilizar valores individuales de participantes, agregados de sitios, agregados de estudios o cualquier combinación de los mismos.
- El Repositorio de Análisis tiene por lo menos un análisis que instruye un sitio de análisis sobre cómo calcular uno o más Agregados de Estudio. Para mayor comodidad, nos referiremos a este análisis en adelante como
- 25 Análisis_A.
- Análisis_A está marcado en el Repositorio de Análisis como vinculado al Análisis_O. Es decir, el Análisis_O puede ser utilizado por más de un Análisis_A. Cada estudio único tiene su propio Análisis_A junto con uno o más Análisis_O.

Cuando se repite un estudio (por ejemplo, posteriormente o en otro conjunto de sitios), se puede volver a utilizar el mismo conjunto de análisis.

- Análisis_A especifica un conjunto de Análisis_O en los que se basa, pero ningún Análisis_O está restringido a ser utilizado por un solo Análisis_A. Por ejemplo, el Análisis_O puede estar relacionado con “contar el número de pacientes asmáticos” y puede ser utilizado por cualquier estudio de pacientes asmáticos.
- Análisis_A no utiliza agregados del sitio que no sean calculados por un Análisis_O.
- Análisis_A no utiliza valores individuales de los participantes.

Instigación

[00261] El cálculo de la tarea puede iniciarse de tres maneras:

4. El estudio es iniciado por el Sitio_A en un momento programado, o como resultado de la información proporcionada por el usuario:

- a. El Sitio_A notifica a cada Sitio_i su intención de sincronizar inmediatamente la hora de inicio del estudio.
- b. Cada Sitio_i acusa recibo de la solicitud con un mensaje al Sitio_A.
 - i. En algunas implementaciones, el Sitio_i puede responder con una solicitud para reprogramar y sugerir una hora alternativa.
 - ii. En tales casos, el Sitio_A enviará una cancelación a cada Sitio_i y repetirá este Paso 1 más adelante.
 - iii. 1. bi y 1.b.ii se pueden repetir hasta que todos los Sitios_i estén listos.

5. El estudio se inicia a una hora determinada.

- a. Cada Sitio_i notifica al Sitio_A su disposición para realizar el estudio.

6. El estudio se inicia cuando

- a. Cada Sitio_i notifica al Sitio_A que se cumple una condición que lo prepara para realizar el estudio.

- b. El Sitio_A espera hasta recibir dichos mensajes de cada Sitio_i.
- c. El Sitio_A inicia el estudio inmediatamente según el paso 1 anterior.

[00262] Independientemente del método de inicio utilizado, al finalizar el proceso, todos los Sitios cuentan con la información necesaria para llevar a cabo el estudio y están listos para continuar.

Sincronización de análisis

- 3. El Sitio_A descarga el último Análisis_A, es decir, una representación computacional del procedimiento utilizado para calcular los Agregados del Estudio a partir de los Agregados del Sitio.

- 4. El Sitio_A podría:

- a. Indicar a cada Sitio_i que descargue el mismo Análisis_O al
 - i. Identificar el Análisis_O asociado con el Análisis_A,
 - ii. Enviar la identidad del Análisis_O a cada Sitio_i de manera que
 - iii. Cada Sitio_i descarga el mismo análisis del Repositorio de Análisis.
- b. O bien, descarga el Análisis_O asociado con el Análisis_A y lo sube a cada Sitio_i.

[00263] Al final de esta etapa, el Sitio_A tiene una copia del Análisis_A y cada Sitio_i tiene una copia idéntica del Análisis_O asociado.

Estructura de análisis

[00264] Cada análisis es una representación de uno o más algoritmos que realizan parte del estudio.

- El Análisis_O utiliza los valores individuales de los participantes, los agregados del sitio y los agregados del estudio para calcular los agregados del sitio.
- El Análisis_A utiliza los agregados del sitio para calcular el agregado del estudio.

- El Análisis_O puede incluir instrucciones implícitas o explícitas para pausar temporalmente la ejecución del algoritmo hasta que estén presentes los agregados de estudio necesarios.
- El Análisis_A puede incluir instrucciones para pausar temporalmente la ejecución del algoritmo hasta que estén presentes los agregados del sitio necesarios.
- El Análisis_A puede incluir instrucciones implícitas o explícitas para enviar agregados del estudio a cada Sitio_i.
- El resultado del Análisis_O son todos los agregados del sitio necesarios para completar el Análisis_A.
- El resultado del Análisis_A son todos los Agregados del Estudio, por ejemplo, en la forma de:
 - un informe de texto,
 - un informe en un formato que pueda enviarse a un sistema de notificación automatizado, como el exigido por una autoridad de salud pública u otra autoridad, y/o
 - valores en un formato interpretable por otro sistema, como un panel de control o un sistema de visualización de datos.

Implementación por Computador

[00265] La figura 4 es un diagrama de bloques esquemático de un sistema de estudios federados 400 en el que se pueden implementar una o más modalidades de la presente descripción. El sistema 400 incluye un conjunto 410 de sitios de origen, entre los que se encuentran el Sitio₁ 412, el Sitio₂ 414 y el Sitio₃ 416. Los Sitios de Origen _{1...3} (412, 414 y 416) pueden estar ubicados en el mismo lugar, en lugares distintos o en una combinación de ambos.

[00266] Cada Sitio de Origen 412, 414 y 416 está asociado con al menos un dispositivo de procesamiento de datos 413, 415 y 417, respectivamente, conectado a una red de comunicaciones 490. Los dispositivos de procesamiento de datos 413,

415 y 417 transmiten datos agregados a través de la red de comunicaciones 490. Esta red puede estar compuesta por uno o más enlaces de comunicaciones cableados, enlaces de comunicaciones inalámbricos o cualquier combinación de ambos. En particular, la red de comunicaciones 490 puede incluir una red de área local (LAN), una red de área extensa (WAN), una red de telecomunicaciones o cualquier combinación de estas. Una red de telecomunicaciones puede incluir, entre otras, una red de telefonía, como una red telefónica pública conmutada (RTPC) o una red de telefonía móvil celular, Internet o cualquier combinación de estas.

[00267] El sistema 400 también incluye un Sitio de Análisis 450. El Sitio de Análisis 450 incluye un Dispositivo de Datos de Estudio 451 que incluye: una Base de Datos Nativa 452, un Traductor de Modelo de Datos 454, una Base de Datos de Valores de Participantes Individuales en Modelo de Datos Común 456, un módulo de Computación 458 y una Base de Datos de Valores Agregados del Sitio 460, cada uno de los cuales está conectado a un bus de comunicaciones 459 que permite el intercambio de información entre todos los componentes del Sitio de Análisis 450.

[00268] El traductor de modelos de datos 454 se utiliza cuando el Sitio de Análisis 450 necesita traducir los datos recibidos a un formato de modelo de datos común. En los casos en que los sitios de origen solo transmiten datos agregados según un modelo de datos común predefinido, el traductor de modelos de datos 454 no es necesario.

[00269] El dispositivo de datos de estudio 451 está conectado a la red de comunicaciones 490 para permitir la comunicación entre el Sitio de Análisis 450 y cada uno de los Sitios de Origen 412, 414 y 416. En particular, los Sitios de Origen 412, 414 y 416 utilizan los respectivos dispositivos de computación de datos 413, 415 y 417 para transmitir conjuntos de datos agregados correspondientes a uno o más estudios, a través de la red de comunicaciones 490, al dispositivo de datos de estudio 451.

[00270] En el ejemplo de la figura 4, el sistema 400 incluye opcionalmente un primer dispositivo informático de observación 470 y un primer observador 475 asociado. El observador 475 puede acceder al dispositivo informático de observación 470 para consultar una interfaz de usuario proporcionada por el sitio de análisis 450 y
5 visualizar los datos asociados a uno o más estudios federados. El dispositivo informático de observación 470 puede estar integrado en el dispositivo de datos de estudio 451 o conectado a este mediante uno o más enlaces de comunicación cableados o inalámbricos.

[00271] En el ejemplo de la figura 4, el sistema 400 incluye opcionalmente un
10 segundo dispositivo informático de observación remota 480 y un segundo observador asociado 485. El dispositivo informático de observación remota 480 está conectado al Dispositivo de Datos de Estudio 451 mediante uno o más enlaces de comunicación cableados o inalámbricos, incluida la red de comunicaciones 490. El segundo observador 485 puede acceder al dispositivo informático de observación
15 remota 480 para acceder a una interfaz de usuario proporcionada por el Sitio de Análisis 450 y así visualizar los datos asociados a uno o más estudios federados.

[00272] A continuación se describirá la funcionalidad del sistema de estudios federados 400 en relación con un estudio de ejemplo relativo al peso corporal de pacientes con diabetes tipo 2.

20 Ejemplo: Media ($\bar{Y}$) y Desviación Estándar (σ) del peso corporal de pacientes con diagnóstico de diabetes mellitus tipo 2 en 3 centros desde 2010.

Dispositivos de entrada y bases de datos nativas de los sitios de origen

- Los sitios de origen (Sitio₁, Sitio₂ y Sitio₃) son centros de atención médica.
- Cada sitio de origen cuenta con un sistema de registro electrónico de salud
25 asociado.
- El Sitio₁ cuenta con básculas electrónicas que registran automáticamente el peso corporal en kilogramos (kg) en el registro electrónico de salud del paciente.

- En el Sitio₂, un médico pesa a los pacientes e introduce el peso corporal en kg en el historial clínico electrónico del paciente.
 - En el Sitio₃, una enfermera pesa a los pacientes e ingresa el peso corporal del paciente en libras (lb) en el registro electrónico de salud del paciente.
- 5 • En cada Sitio de Origen, un médico introduce un diagnóstico de diabetes mellitus tipo 2 en el historial clínico electrónico del paciente diagnosticado.

Modelo de datos común

[00273] En este ejemplo, el modelo de datos común para el estudio incluye una tabla de condiciones y una tabla de observaciones. En otras implementaciones
10 relacionadas con diferentes estudios, el modelo de datos común puede incluir una o más tablas u otros formatos de datos adecuados. En este ejemplo, la información de la tabla de condiciones y la tabla de observaciones se obtiene de la base de datos nativa en el sitio de origen correspondiente.

[00274] Parte de la información en la Tabla de Condiciones y la Tabla de
15 Observaciones del modelo de datos común utilizado en este ejemplo emplea códigos de la terminología clínica SNOMED-CT (snomed.org), que proporciona una traducción de estos códigos a una descripción textual estandarizada e inequívoca de la condición, observación, etc. En este ejemplo, las descripciones textuales de los códigos relevantes no se incluyen en el modelo de datos común, pero se
20 proporcionan en el cuerpo del texto para mayor comodidad.

[00275] Los diagnósticos de diabetes mellitus tipo 2 se registran en una tabla de “afecciones”, con el número de identificación único del paciente (UID), la fecha y hora de la medición y el código SNOMED-CT 44054006 (“Diabetes mellitus tipo 2”) como se ilustra en la Tabla 1 a continuación:

Condiciones		
UID	Fecha y hora	código de condición
1	21 de abril de 2010, 14:35	44054006

Tabla 1

[00276] El peso corporal es registrado en una tabla de “Observaciones” con el UID del paciente, con el código SNOMED 27113001 (“Peso corporal”) y el valor

5 correspondiente en kg y la unidad como código SNOMED 258683005 (“kg”) como se ilustra en la Tabla 2 a continuación.

Observaciones				
UID	Fecha y hora	código de observación	valor de observación	código de unidad
1	21 de abril de 2010, 14:30	27113001	98.3	258683005

Tabla 2

Traductores del modelo de datos de los sitios de origen

10 [00277] Cada Sitio de Origen cuenta con su propio traductor, como el Traductor de Modelo de Datos 114 de la figura 1, para traducir los datos de su propio sistema de historia clínica electrónica al modelo de datos común. En el Sitio₃, el traductor incluye la conversión de pesos corporales de libras (lb) a kilogramos (kg).

Ejemplo: cálculo de la media y la desviación estándar del peso corporal de los participantes en un estudio federado.

Fórmulas para calcular la media de la fecha y hora ($\hat{Y}$) y la desviación estándar

(σ)

$$\hat{Y} = \frac{1}{N} \sum_{j=1}^N Y_j$$

Ecuación (1)

$$\sigma = \sqrt{\frac{1}{N} \sum_{j=1}^N (Y_j - \hat{Y})^2}$$

Ecuación (2)

Pseudocódigo de Análisis_A

- 5
8. Solicitar y esperar el número de participantes incluidos (n_1 , n_2 y n_3) de cada Sitio de Origen correspondiente.
9. Calcular el agregado del estudio “número de pacientes incluidos en el estudio” N como $N = \sum_{i=1}^3 n_i$
10. Solicitar y esperar el valor agregado del sitio “peso corporal medio del sitio” ($\hat{Y}_1$, $\hat{Y}_2$ y $\hat{Y}_3$) de cada sitio de origen correspondiente.
- 10
11. Calcular el agregado del estudio “peso corporal medio del estudio” $\hat{Y}$ como $\hat{Y} = \frac{1}{N} \sum_{i=1}^3 \hat{Y}_i \cdot n_i$
12. Solicitar y esperar el agregado del sitio "suma de diferencias al cuadrado" ($\hat{Z}_1$, $\hat{Z}_2$ y $\hat{Z}_3$) de cada sitio de origen correspondiente.
- 15
13. Calcular la “desviación estándar” del agregado del estudio σ a partir de la “suma de las diferencias al cuadrado” de los agregados del sitio ($\hat{Z}_1$, $\hat{Z}_2$, $\hat{Z}_3$) y el “número de pacientes incluidos en el estudio” del agregado del estudio N como $\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^3 \hat{Z}_i \cdot n_i}$
14. Informe del estudio Agregado “peso corporal medio del estudio” $\hat{Y}$ y “desviación estándar” σ
- 20

Pseudocódigo de Análisis_O

6. Establecer el número de sitio como i
7. Contar el número de pacientes únicos incluidos con el código de condición 44054006 con fecha y hora de condición entre el 1 de enero de 2010 a las 12:00 a. m. y el 31 de diciembre de 2010 a las 11:59 p. m. como n_i
- 25

8. Calcular el “peso corporal medio del sitio” $\hat{Y}_i$ a partir del peso corporal del valor individual del participante (Y) para cada paciente incluido j y el número de pacientes incluidos en el sitio n como $\hat{Y}_i = \frac{1}{n} \sum_{j=1}^n Y_j$

9. Solicitar y esperar el dato agregado del estudio “peso corporal medio del estudio” $\hat{Y}$ del Sitio_A.

10. Calcular el agregado del sitio “suma de las diferencias al cuadrado” $\hat{Z}_i$ del peso corporal de cada paciente incluido j y el agregado del estudio “peso corporal medio del estudio” $\hat{Y}$ como $\hat{Z}_i = \sum_{j=1}^n (Y_j - \hat{Y})^2$

[00278] El sistema de estudios federados de la presente descripción puede implementarse utilizando uno o más dispositivos informáticos, tales como un computador de propósito general, un servidor, una arquitectura distribuida en la nube, una combinación de estos o cualquier arquitectura similar programada para realizar una o más de las funciones mostradas y descritas en las figuras 1 a 4, dando lugar así a un dispositivo informático nuevo y mejorado. En particular, el dispositivo de datos de estudio 115a y el dispositivo de análisis de datos 451 representan avances en tecnología informática implementados mediante uno o más dispositivos informáticos.

[00279] La figura 5 es un diagrama de bloques esquemático de un sistema que incluye un computador de propósito general 510. El computador de propósito general 510 incluye varios componentes, entre ellos: un procesador 512, una memoria 514, un medio de almacenamiento 516, interfaces de entrada/salida (E/S) 520 y puertos de entrada/salida (E/S) 522. Los componentes del computador de propósito general 510 generalmente se comunican mediante uno o más buses 548.

[00280] La memoria 514 puede implementarse mediante memoria de acceso aleatorio (RAM), memoria de solo lectura (ROM) o una combinación de ambas. El medio de almacenamiento 516 puede implementarse como un disco duro, una unidad de estado sólido (sólido), una unidad de disco óptico u otro medio de

almacenamiento. El medio de almacenamiento 516 puede utilizarse para almacenar uno o más programas informáticos, incluyendo un sistema operativo, aplicaciones de software y datos. En un modo de funcionamiento, las instrucciones de uno o más programas informáticos almacenados en el medio de almacenamiento 516 se cargan en la memoria 514 a través del bus 548. Las instrucciones cargadas en la memoria 514 se ponen a disposición del procesador 512 a través del bus 548 u otros medios para su ejecución, implementando así un modo de funcionamiento acorde con las instrucciones ejecutadas.

[00281] Uno o más dispositivos periféricos pueden conectarse al computador de propósito general 510 a través de los puertos de E/S 522. En el ejemplo de la figura 5, el computador de propósito general 510 está conectado a un altavoz 524, una cámara 526, una pantalla 530, un dispositivo de entrada 532, una impresora 534 y un dispositivo de almacenamiento externo 536. El altavoz 524 puede implementarse mediante uno o más altavoces, como en un sistema de sonido estéreo o envolvente.

En el ejemplo en el que el computador de propósito general 510 se utiliza para implementar una o más funciones de un sistema de estudios federados, como un dispositivo de datos de estudio, uno o más dispositivos periféricos pueden corresponder a los dispositivos de entrada 111a...d de la figura 3, conectados a los puertos de E/S 522 de forma inalámbrica o mediante conexión por cable.

[00282] La cámara 526 puede ser una webcam u otra cámara digital de fotos o vídeo, y puede descargar y cargar información desde y hacia el computador de uso general 510 a través de los puertos de E/S 522, según la implementación específica. Por ejemplo, las imágenes grabadas por la cámara 526 pueden cargarse en el dispositivo de almacenamiento 516 del computador de uso general 510. Del mismo modo, las imágenes almacenadas en el dispositivo de almacenamiento 516 pueden descargarse a la memoria o al dispositivo de almacenamiento de la cámara 526. La cámara 526 puede incluir un sistema de lentes, una unidad de sensor y un dispositivo de grabación.

[00283] El dispositivo de visualización 530 puede ser un monitor de computador, como una pantalla de tubo de rayos catódicos, una pantalla de plasma o una pantalla de cristal líquido (LCD). La pantalla 530 recibe información del computador 510 de forma convencional, y esta se presenta en el dispositivo 530 para que el usuario la visualice. Opcionalmente, el dispositivo de visualización 530 puede incorporar una pantalla táctil para que el usuario pueda interactuar con el computador 510. La pantalla táctil puede ser, por ejemplo, capacitiva, resistiva, de ondas acústicas superficiales, etc.

[00284] El dispositivo de entrada 532 puede ser un teclado, un ratón, un lápiz óptico, una tableta gráfica o cualquier combinación de estos, para recibir datos del usuario. El medio de almacenamiento externo 536 puede incluir un disco duro externo (HDD), una unidad óptica, una unidad de disquete, una unidad flash, una unidad de estado sólido (SSD) o cualquier combinación de estos, y puede implementarse como una o varias instancias de cualquiera de estos dispositivos. Por ejemplo, el medio de almacenamiento externo 536 puede implementarse como una matriz de discos duros.

[00285] Las interfaces de E/S 520 facilitan el intercambio de información entre el dispositivo informático de propósito general 510 y otros dispositivos informáticos. Estas interfaces pueden implementarse mediante un módem interno o externo, una conexión Ethernet, etc., para permitir la conexión a un medio de transmisión. En el ejemplo de la figura 5, las interfaces de E/S 522 están conectadas a una red de comunicaciones 538 y directamente a un dispositivo informático 542. Este dispositivo 542 se muestra como un computador personal, pero también podría ser un teléfono inteligente, un computador portátil o una tableta. La comunicación directa entre el computador de propósito general 510 y el dispositivo informático 542 puede implementarse mediante un enlace de transmisión inalámbrico o cableado.

[00286] La red de comunicaciones 538 puede implementarse mediante uno o más enlaces de transmisión alámbricos o inalámbricos, e incluir, por ejemplo, un enlace

de comunicaciones dedicado, una red de área local (LAN), una red de área extensa (WAN), Internet, una red de telecomunicaciones o cualquier combinación de estas. Una red de telecomunicaciones puede incluir, entre otras, una red de telefonía, como una red telefónica pública conmutada (RTPC), una red celular de telefonía móvil, una red de mensajes cortos (SMS) o cualquier combinación de estas. El computador de uso general 510 puede comunicarse a través de la red de comunicaciones 538 con otros dispositivos informáticos conectados a dicha red, como el teléfono móvil 544, el teléfono inteligente con pantalla táctil 546, el computador personal 540 y el dispositivo informático 542.

10 [00287] Se pueden utilizar una o más unidades del computador de propósito general 510 para implementar una o más funciones del dispositivo de datos de estudio 115a de la figura 3, con el fin de implementar un Sitio de Origen de un sistema de estudios federados, de acuerdo con la presente descripción. En dicha implementación, la memoria 514 y el almacenamiento 516 se utilizan para almacenar datos relativos a
15 datos de pacientes, datos de análisis, etc. El software para implementar el sistema de estudios federados se almacena en la memoria 514 o en el almacenamiento 516, o en ambos, para su ejecución en el procesador 512. El software incluye código de programa informático para implementar los pasos del método de acuerdo con los módulos funcionales descritos en este documento.

20 [00288] La figura 6 es un diagrama de bloques esquemático de un sistema 600 en el que se puede implementar uno o más aspectos de un método y sistema federado de la presente descripción. El sistema 600 incluye un dispositivo informático portátil, un teléfono inteligente 610, que puede ser utilizado por un usuario registrado del sistema de estudios federados de la figura 1. El teléfono inteligente 610 incluye
25 varios componentes, entre ellos: un procesador 612, una memoria 614, un medio de almacenamiento 616, una batería 618, una antena 620, un transmisor y receptor de radiofrecuencia (RF) 622, una tarjeta SIM 624, un altavoz 626, un dispositivo de entrada 628, una cámara 630, una pantalla 632 y un transmisor y receptor

inalámbrico 634. Los componentes del teléfono inteligente 610 se comunican generalmente mediante una o más conexiones de bus 648 u otras conexiones entre ellos. El teléfono inteligente 610 también incluye una conexión por cable 645 para acoplarse a una toma de corriente para recargar la batería 618 o para conectarse a un dispositivo informático, como el computador de uso general 510 de la figura 5. La conexión por cable 645 puede incluir uno o más conectores y puede estar adaptada para permitir la carga y descarga de contenido desde y hacia la memoria 614 y la tarjeta SIM 624.

[00289] El teléfono inteligente 610 puede incluir muchos otros componentes funcionales, como un convertidor digital-analógico y analógico-digital de audio y un amplificador, pero se omiten para mayor claridad. Sin embargo, un experto en la materia los conocería y comprendería fácilmente.

[00290] La memoria 614 puede incluir memoria de acceso aleatorio (RAM), memoria de solo lectura (ROM) o una combinación de ambas. El medio de almacenamiento 616 puede ser una unidad de estado sólido (sólido), un medio de almacenamiento extraíble, como una tarjeta SD o microSD, u otro tipo de almacenamiento. El medio de almacenamiento 616 puede utilizarse para almacenar uno o más programas informáticos, incluyendo un sistema operativo, aplicaciones de software y datos. En un modo de funcionamiento, las instrucciones de uno o más programas informáticos almacenados en el medio de almacenamiento 616 se cargan en la memoria 614 a través del bus 648. Las instrucciones cargadas en la memoria 614 se ponen a disposición del procesador 612 a través del bus 648 u otros medios para su ejecución, implementando así un modo de funcionamiento acorde con las instrucciones ejecutadas.

[00291] El teléfono inteligente 610 también incluye un módulo de interfaz de programación de aplicaciones (API) 636, que permite a los programadores escribir aplicaciones de software para ejecutar en el procesador 612. Dichas aplicaciones incluyen una pluralidad de instrucciones que pueden estar preinstaladas en la

memoria 614 o descargadas a la memoria 614 desde una fuente externa, a través del transmisor y receptor de RF 622 que operan en asociación con la antena 620 o a través de la conexión cableada 645.

[00292] El teléfono inteligente 610 incluye además un módulo de localización GPS 638. Este módulo se utiliza para determinar la posición geográfica del dispositivo mediante satélites GPS, triangulación de antenas de telefonía móvil o una combinación de ambos métodos. Dicha posición geográfica se puede poner a disposición de uno o más programas o aplicaciones que se ejecutan en el procesador 612.

10 [00293] El transmisor y receptor inalámbricos 634 pueden utilizarse para comunicarse de forma inalámbrica con dispositivos periféricos externos mediante Bluetooth, infrarrojos u otro protocolo inalámbrico. En el ejemplo de la figura 6, el teléfono inteligente 610 está conectado a una impresora 640, un dispositivo de almacenamiento externo 644 y un computador 642. El computador 642 puede
15 implementarse, por ejemplo, utilizando el computador de propósito general 510 de la figura 5.

[00294] La cámara 626 puede incluir una o más cámaras digitales, tanto de fotografía como de vídeo, adaptadas para capturar y grabar en la memoria 614 o en la tarjeta SIM 624 imágenes fijas o de vídeo, o una combinación de ambas. La
20 cámara 626 puede incluir un sistema de lentes, una unidad de sensor y un medio de grabación. El usuario del teléfono inteligente 610 puede cargar las imágenes grabadas a otro computador o dispositivo periférico mediante el transmisor y receptor inalámbricos 634, el transmisor y receptor de radiofrecuencia 622 o la conexión por cable 645.

25 [00295] En un ejemplo, el dispositivo de visualización 632 utiliza una pantalla de cristal líquido (LCD). La pantalla 632 se emplea para mostrar contenido al usuario del teléfono inteligente 610. Opcionalmente, la pantalla 632 puede implementarse

mediante una pantalla táctil, como una pantalla táctil capacitiva o resistiva, para que el usuario pueda interactuar con el teléfono inteligente 610.

[00296] El dispositivo de entrada 628 puede ser, por ejemplo, un teclado, un lápiz óptico o un micrófono, para recibir datos del usuario. Si se trata de un teclado, este puede implementarse como un conjunto de teclas físicas en el teléfono inteligente 610. Alternativamente, puede ser un teclado virtual que se muestra en la pantalla 632.

[00297] La tarjeta SIM 624 se utiliza para almacenar la Identidad Internacional del Abonado Móvil (IMSI) y una clave asociada que permite identificar y autenticar al usuario en la red celular a la que está suscrito. La tarjeta SIM 624 suele ser extraíble y se puede usar indistintamente en diferentes teléfonos inteligentes o celulares. En ella se pueden almacenar los contactos del usuario, incluyendo nombres y números de teléfono. También puede almacenar fotos y videos. Como alternativa, los contactos se pueden guardar en la memoria 614.

[00298] El transmisor y receptor de radiofrecuencia 622, junto con la antena 620, permiten el intercambio de información entre el teléfono inteligente 610 y otros dispositivos informáticos a través de una red de comunicaciones 690. En el ejemplo de la figura 6, el transmisor y receptor de radiofrecuencia 622 permiten que el teléfono inteligente 610 se comunique a través de la red de comunicaciones 690 con un teléfono celular 650, un teléfono inteligente o tableta 652, un dispositivo informático 654 y otro dispositivo informático 642. Los dispositivos informáticos 654 y 642 se muestran como computadores personales, pero también podrían ser teléfonos inteligentes, computadores portátiles o tabletas.

[00299] La red de comunicaciones 690 puede implementarse mediante uno o más enlaces de transmisión alámbricos o inalámbricos, e incluir, por ejemplo, una red de telefonía celular, un enlace de comunicaciones dedicado, una red de área local (LAN), una red de área extensa (WAN), Internet, una red de telecomunicaciones o cualquier combinación de estas. Una red de telecomunicaciones puede incluir, entre

otras, una red de telefonía, como una red telefónica pública conmutada (RTPC), una red celular de telefonía móvil, una red de mensajes cortos (SMS) o cualquier combinación de estas.

[00300] Cuando se implementan una o más funciones del sistema de estudios federados descrito en este documento utilizando el teléfono inteligente 610 de la figura 6, se puede utilizar una aplicación de programa (“app”) que se ejecuta en el procesador 612 para implementar cualquiera de las funciones descritas y mostradas en relación con las figuras 1 a 4. En algunas implementaciones, la app es una aplicación nativa que se ejecuta en el teléfono inteligente 600. En otras implementaciones, la app es una aplicación web que se muestra en un navegador que se ejecuta en el procesador 612, con el teléfono inteligente 610 conectado a un servidor remoto, como el dispositivo informático 642, en el que se ejecuta la app.

[00301] La figura 7 es un diagrama de secuencia que ilustra el flujo de trabajo 700 de un estudio federado realizado por una implementación del sistema descrito en la presente divulgación. El flujo de trabajo 700 se refiere a un sistema que cuenta con un repositorio 702, un usuario 704, un sitio de análisis 706 y tres sitios de origen: un primer sitio de origen 708, un segundo sitio de origen 710 y un tercer sitio de origen 712.

[00302] En un primer paso, el usuario 704 inicia un estudio federado en el sitio de análisis 706, cuyo inicio se produce cuando dicho sitio recupera el AnálisisA del repositorio 702. Como se describió anteriormente, cada AnálisisA está asociado a un conjunto de uno o más AnálisisO. El sitio de análisis 706 identifica el AnálisisO asociado al AnálisisA recuperado y, a continuación, lo recupera del repositorio 702.

[00303] El sitio de análisis 706 proporciona entonces el AnálisisO recuperado a cada uno de los sitios de origen 708, 710 y 712. De este modo, cada sitio de origen 708, 710 y 712 dispone del AnálisisO para llevar a cabo un estudio.

[00304] Cada sitio de origen (708, 710, 712) agrega los datos del estudio correspondientes. En la figura 7, esto se muestra mediante el cálculo de n_1 por

análisis (708) para el primer sitio de origen, n2 por análisis (710) para el segundo y n3 por análisis (712) para el tercero.

[00305] El sitio de análisis 706 recupera los valores calculados de n1, n2 y n3 de los respectivos sitios de origen 708, 710 y 712, y luego calcula N de acuerdo con AnalysisO. Posteriormente, el sitio de análisis 706 proporciona un informe al usuario 704, basado en el valor calculado de N.

[00306] La figura 8 es un diagrama de bloques esquemático de la arquitectura de una máquina virtual en la nube para implementar un sitio de origen 800 en un sistema descrito en la presente divulgación. El ejemplo de la figura 8 utiliza la plataforma de computación en la nube AWS, proporcionada por Amazon Web Services, Inc. En concreto, se utiliza Amazon Virtual Private Cloud (VPC) para implementar el sitio de origen, y Amazon Elastic Block Store (EBS) como almacenamiento a nivel de bloque para almacenar datos. Dichos datos pueden incluir, por ejemplo, datos de pacientes, datos de estudios, análisis, etc.

[00307] El ejemplo de la figura 8 también utiliza Amazon Elastic Cloud Compute (EC2) como máquina virtual 810, que aloja una aplicación de software para implementar la funcionalidad del sitio de origen. La máquina virtual 810 incluye un módulo de cómputo 818, equivalente al módulo de cómputo 118 de la figura 3; una base de datos de participantes 816, equivalente a la base de datos 116 de la figura 3; y una base de datos de datos agregados 820, equivalente a la base de datos de valores agregados 120 de la figura 3.

[00308] El sitio de origen 800 es accesible a través de Internet por usuarios finales que acceden mediante navegadores web en dispositivos informáticos. El sitio de origen 800 también está conectado, mediante una puerta de enlace a Internet, a una base de datos 850 que funciona como repositorio de análisis.

[00309] La figura 9 es un diagrama de bloques esquemático de un sistema de Sitio de Origen 900 implementado mediante una arquitectura informática "sin servidor" basada en la nube 910, proporcionada en este ejemplo por Amazon Web Services.

Un sitio de administración 920 se implementa utilizando una máquina virtual que incluye el sistema de distribución de software Elastic Container Registry (ECR) para distribuir software al sitio del cliente 930.

[00310] El sitio del cliente 930 se implementa utilizando una máquina virtual 932 que
5 implementa el sitio de origen 110a de las figuras 1 y 3. El sitio del cliente 930 incluye una serie de componentes funcionales, entre ellos un módulo de computación, una base de datos de datos agregados y una base de datos de datos de participantes, tal como se establece y describe en la figura 3. El sitio del cliente 930 está conectado, mediante uno o más enlaces de comunicación, a un computador de
10 usuario final 940 y a un repositorio de análisis 950.

[00311] La figura 10 es un diagrama de bloques esquemático de un sistema de estudios federados 1000 implementado mediante una arquitectura informática basada en la nube. El sistema 1000 incluye un módulo de análisis federado 1005 que representa conceptualmente cómo un conjunto de instrucciones del sitio de
15 análisis 1006 y un conjunto de instrucciones del sitio de análisis 1006 y un conjunto de instrucciones del sitio de origen 1007 podría organizarse de manera que parezca un único módulo funcional.

[00312] Se implementa un sitio de origen 1010 mediante un sistema basado en la nube, como el descrito en la figura 8. El sitio de origen 1010 incluye un conjunto de
20 instrucciones 1012 y un modelo de datos común 1014 definido para un estudio federado específico. El sitio de origen 1010 está conectado a una puerta de enlace de red 1060 que permite el intercambio de datos entre el sitio de origen 1010, un sitio de análisis 1050 y un repositorio de análisis 1020. El sitio de análisis 1050 incluye un conjunto de instrucciones 1052.

[00313] La figura 11 es un diagrama de bloques esquemático de un sistema de estudios federados 1100 implementado mediante una arquitectura informática basada en la nube. El sistema 1100 del ejemplo de la figura 11 incluye un primer sitio de origen 1115 alojado en un primer centro de datos, un segundo sitio de origen

1125 alojado en un segundo centro de datos, un sitio de análisis 1150 alojado en un tercer centro de datos y un repositorio de análisis 1160; cada uno de estos elementos se implementa en este ejemplo en un entorno de computación en la nube.

5 [00314] El primer sitio de origen 1115 está conectado a una primera puerta de enlace de red 1118, la cual está conectada a una red de comunicaciones 1190. El segundo sitio de origen 1125 está conectado a una segunda puerta de enlace de red 1128, que también está conectada a la red de comunicaciones 1190. El sitio de análisis 1150 y el repositorio de análisis 1160 se encuentran en un tercer centro de
10 datos común y están conectados, a través de una tercera puerta de enlace de red 1125, a la red de comunicaciones 1190.

[00315] La red de comunicaciones 1190 puede implementarse utilizando uno o más enlaces de comunicaciones alámbricos o inalámbricos, y puede incluir Internet, que permiten la transmisión de datos entre el primer sitio de origen 1115, el segundo
15 sitio de origen 1125, el sitio de análisis 1150 y el repositorio de análisis 1160.

[00316] El sistema de estudios federados 1100 de la figura 11 permite que dos sitios de origen 1115, 1125 recopilen datos de participantes de sitios separados y luego transmitan esos datos, de acuerdo con un modelo de datos común predefinido, para su procesamiento por el sitio de análisis 1150. Dicho sistema de estudios federados
20 1100 proporciona una arquitectura informática mejorada para realizar estudios federados.

Unificación de dialectos del modelo de datos

[00317] La pluralidad de representaciones de datos en distintos sitios puede generar variabilidad en el “modelo” o “formato” de los datos en un repositorio, como una base
25 de datos relacional o un lago de datos. Esta variabilidad puede introducir errores en el análisis, la recuperación y la actualización de datos en dichos repositorios, especialmente cuando se combinan varios repositorios mediante agregación

(copiando contenido de uno o más repositorios a otro) o federación (combinando la información resumida obtenida en cada repositorio).

[00318] Para abordar este problema, se han propuesto varios estándares de datos comunes, como FHIR, OMOP e I2B2 en medicina. Sin embargo, la traducción de datos a estos modelos comunes suele requerir conocimientos especializados, subjetividad y criterio, lo que conlleva diferencias entre el modelo de datos traducido y el estandarizado (a veces denominadas “dialectos”). Estas diferencias entre dialectos limitan la capacidad de combinar datos, incluso de repositorios que utilizan el mismo modelo de datos común.

[00319] Existen dos razones principales por las que se producen dialectos: la interpretación y la omisión. La interpretación se produce cuando las instrucciones para modelar datos según un estándar son ambiguas o poco claras, lo que permite que quienes transforman los datos en diferentes lugares o en diferentes momentos las interpreten de forma distinta. Como resultado, aunque los conjuntos de datos traducidos sean nominalmente compatibles con el mismo estándar, es posible que no sean interoperables entre sí. La solución consiste en utilizar las mismas asignaciones en cada lugar.

[00320] Se produce una omisión cuando los desarrolladores del estándar no previeron un caso de uso al publicarlo, antes de conocer algunos de los datos que debía representar el modelo de datos. La mayoría de los estándares proporcionan directrices para la extensión del estándar del modelo de datos original. Por ejemplo, FHIR proporciona un recurso de extensión y OMOP reserva un rango de identificadores únicos para extensiones de código de concepto (2.000.000.000 y superiores). Sin embargo, incluso cuando las extensiones se crean según las directrices, la flexibilidad de la extensión implica que los conjuntos de datos de cada sitio suelen ser incompatibles entre sí.

[00321] Por ejemplo, si dos sitios, el Sitio 1 y el Sitio 2, necesitan representar un nuevo virus (X) y un nuevo procedimiento quirúrgico (Y), cada uno podría asignarles

códigos de concepto diferentes, como se muestra en la Tabla 3. La solución consiste en utilizar los mismos códigos de extensión para las mismas extensiones en ambos sitios.

	ID de concepto asignado en el sitio 1	ID de concepto asignado en el sitio 2
Nuevo virus X	2.000.000.001	2.000.000.002
Nuevo procedimiento quirúrgico Y	2.000.000.002	2.000.000.003

Tabla 3

5

[00322] Dado que ambos sitios son compatibles con el estándar pero no entre sí, el modelo de datos de cada sitio se denomina “dialecto” del estándar. El estándar es análogo a un idioma, y los datos procedentes de distintos sitios son incompatibles porque cada sitio utiliza un dialecto diferente de ese idioma.

10

[00323] La ambigüedad y la omisión se cuelan en la documentación del estándar, a menudo cuando sus autores no prevén todos los conjuntos de datos que podrían necesitar transformación ni las decisiones de traducción específicas que estos requieren. Estudios recientes demuestran que dos sistemas que utilizan el mismo estándar pueden ser compatibles solo entre un 30 % y un 60 % debido a una combinación de estos factores. Ver, por ejemplo, Elmer V. Bernstam et al., “Quantitating and assessing interoperability between electronic health records”, Journal of the American Medical Informatics Association, volumen 29, número 5, mayo de 2022, páginas 753-760, <https://doi.org/10.1093/jamia/ocab289>.

15

20

[00324] Independientemente de la causa, el resultado es que cada sitio desarrolla su propio dialecto específico. Este fenómeno es tan común que la formación de dialectos se produce casi siempre que se estandariza un nuevo conjunto de datos a un modelo de datos común, generalmente sin que nadie se dé cuenta hasta

después de que los datos ya se han traducido. Los transformadores deben tener especial cuidado y, como medida de precaución, es necesario acordar un protocolo.

[00325] Algunas implementaciones de la presente invención proporcionan un método para automatizar la unificación de traducciones de datos entre múltiples sitios que colaboran, de modo que se genere el mismo dialecto en todos ellos. Esto aumenta la interoperabilidad entre dichos sitios. El método utiliza un repositorio central al que puede acceder un sistema de traducción en cada sitio. En algunas implementaciones, el repositorio central se implementa mediante un medio de almacenamiento legible por computador e instrucciones ejecutables por computador que se ejecutan en el procesador de un dispositivo informático conectado a una red de comunicaciones para comunicarse con uno o más sitios de origen.

[00326] El repositorio central almacena cada extensión y mapeo, y permite que las mismas extensiones y mapeos se compartan entre todos los sitios. Cada extensión y mapeo está asociado a un identificador único en el repositorio. El método también utiliza un cliente de traducción presente en cada sitio. El cliente de traducción es un software que se ejecuta en uno o más procesadores del sitio correspondiente.

[00327] Este método utiliza además un Servicio de Emisión de Identificadores, que proporciona una conexión entre cada sitio de origen y el repositorio central. El Servicio de Emisión de Identificadores se encarga de emitir nuevos identificadores únicos para las nuevas extensiones y asignaciones, y de reutilizar los identificadores existentes para las extensiones o asignaciones que ya se encuentran en el repositorio.

[00328] En algunas implementaciones, el Servicio de Emisión de Identificadores se implementa utilizando un medio de almacenamiento legible por computador e instrucciones ejecutables por computador que se ejecutan en un procesador de un dispositivo informático conectado a una red de comunicaciones para la comunicación con uno o más sitios de origen y el repositorio central.

[00329] En algunas implementaciones, el repositorio central se incluye dentro del Repositorio de Análisis 120, 702, 850, 950, 1020, 1160 descrito anteriormente. En otras implementaciones, el repositorio central se implementa como un nuevo nodo de red conectado a una red de comunicaciones para que sea accesible desde todos los sitios de origen dentro de una red de estudios federada.

[00330] En algunas implementaciones, el Servicio de Emisión de Identificadores se encuentra ubicado junto al repositorio central o forma parte integral del mismo, como por ejemplo, al integrarse en el Repositorio de Análisis 120, 702, 850, 950, 1020 y 1160 descritos anteriormente. En otras implementaciones, el Servicio de Emisión de Identificadores se implementa como un nuevo nodo de red conectado a una red de comunicaciones, de modo que sea accesible para todos los sitios de origen y el repositorio central dentro de una red de estudios federada.

[00331] La figura 12 es un diagrama de bloques esquemático que ilustra una porción de una red de estudios federados con un Sitio de Origen 1240, representativo de cualquier sitio de origen dentro de dicha red. El Sitio de Origen 1240 incluye el software de traducción 1250 y un cliente de traducción 1260. El cliente de traducción 1260 está configurado para comunicarse, a través de una primera red de comunicaciones 1270, con un Servicio de Emisión de Identificadores 1210. Este Servicio de Emisión de Identificadores 1210 está conectado, mediante una segunda red de comunicaciones 1230, a un repositorio central 1220.

[00332] Según la implementación, la primera red de comunicación 1270 y la segunda red de comunicación 1230 pueden ser la misma red, como una red de área local (LAN), una red de área extensa (WAN), una red de telecomunicaciones o cualquier combinación de estas. Una red de telecomunicaciones puede incluir, entre otras, una red de telefonía, como una red telefónica pública conmutada (RTPC) o una red de telefonía móvil celular, Internet o cualquier combinación de estas. En otras implementaciones, la primera red de comunicación 1270 y la segunda red de comunicación 1230 pueden ser enlaces físicos directos, como un bus de

computador, en circunstancias donde dos o más de los siguientes elementos (el sitio de origen 1240, el servicio emisor de identificadores 1210 y el repositorio central 1220) se encuentran ubicados en el mismo lugar o integrados entre sí.

[00333] En el ejemplo de la Fig. 12, el Servicio de Emisión de Identificadores 1210 y el repositorio central 1220 se muestran como parte de una Ubicación Central 1200. Sin embargo, se debe tener en cuenta que el Servicio de Emisión de Identificadores 1210 y el repositorio central 1220 pueden ser nodos separados ubicados remotamente entre sí, ubicados cerca uno del otro o incluso integrados entre sí en diferentes modalidades de la presente descripción.

[00334] El repositorio central en línea almacena y recupera elementos de traducción de datos (como se describe a continuación) para un sistema, así como identificadores únicos para las extensiones definidas. El repositorio central utiliza estos identificadores únicos para recuperar las extensiones solicitadas. Puede ser específico para estos elementos o, de forma más general, admitir también otros, como flujos de trabajo de análisis federado. Se pueden usar repositorios separados para cada elemento, siempre que, en conjunto, sean lógicamente equivalentes a un único repositorio.

[00335] El repositorio central actúa como un sistema de almacenamiento que guarda y recupera elementos de traducción de datos, pero no los ejecuta. Los sistemas de traducción, ubicados en cada sitio de origen, utilizan estos elementos durante el proceso de traducción. Según la naturaleza de los elementos de traducción de datos y la aplicación específica, algunos se pueden ejecutar, otros usarse como referencias y otros como recursos del sistema de traducción.

[00336] El servicio de emisión de identificadores garantiza que todos los sitios web hagan referencia al mismo elemento de traducción utilizando el mismo identificador. El repositorio permite que todos los traductores de cada sitio de origen descarguen, mediante clientes de traducción, el mismo elemento de traducción utilizando el identificador único asociado a dicho elemento.

[00337] El resultado es que todos los sistemas de traducción en los respectivos sitios de origen hacen referencia al mismo elemento con el mismo identificador, además de utilizar el mismo elemento que cada sistema de traducción recupera del repositorio central. Los elementos de traducción solo pueden ser utilizados por los traductores en los sitios de origen, ya que la traducción se realiza sobre datos de pacientes, y estos datos no salen de su sitio de origen.

[00338] El cliente de traducción 1260 es utilizado por cada sistema de traducción en cada sitio de origen, como el traductor de modelos de datos 114 de la figura 3, para compartir y recuperar conjuntos de elementos de traducción de datos. Cuando el cliente de traducción 1260 detecta la necesidad de una extensión, la registra en el Servicio de Emisión de Identificadores 1210, solicitando un identificador único para la nueva extensión. El Servicio de Emisión de Identificadores 1210 envía una consulta al Repositorio Central 1220 para verificar si la extensión solicitada ya existe.

[00339] Cuando existe la extensión solicitada, el Repositorio Central devuelve el identificador solicitado asociado con la extensión existente al Servicio de Emisión de Identificadores 1210, que a su vez reenvía el identificador asociado al cliente de traducción 1260.

[00340] Cuando la extensión solicitada no existe en el Repositorio Central, el Servicio Emisor de Identificadores 1210 emite un nuevo identificador único para la extensión solicitada al cliente de traducción 1260.

[00341] En algunas modalidades, el conjunto de elementos de traducción de datos incluye, pero no se limita a, uno o más de los siguientes:

- Conceptos novedosos
- Conjuntos de valores
- Relaciones de conceptos novedosos
- Recursos novedosos
- Pruebas de calidad de datos

- Guiones de caracterización de poblaciones
- Asignaciones de campos
- Mapas conceptuales

[00342] Los conceptos novedosos son frases cortas que se utilizan en los datos que se están transformando para atribuirles significado cuando dichas frases no existen previamente. Los conceptos se utilizan habitualmente en los datos sanitarios para proporcionar representaciones comunes de diagnósticos, medicamentos, procedimientos y otros conceptos clínicos, así como metadatos.

[00343] Existen varias terminologías, algunas con millones de frases, pero durante el mapeo de un conjunto de datos aún pueden surgir nuevas. Por ejemplo, el descubrimiento de un nuevo patógeno puede requerir un nombre único para poder utilizarlo en un diagnóstico o un resultado de laboratorio. Las nuevas terapias requieren descripciones novedosas para referirse a ellas con precisión en los datos. Si los registros de datos en múltiples sitios contienen referencias a un nuevo virus o terapia, se necesita un concepto novedoso para garantizar la coherencia de dichas referencias y que se representen de la misma manera en todos los sitios y, por lo tanto, en todos los conjuntos de datos.

[00344] Los conjuntos de valores (también conocidos como conjuntos de conceptos) son grupos de conceptos con nombre propio. Si bien normalmente no forman parte de un modelo de datos ni de un conjunto de datos, suelen utilizarse como referencia en el proceso de traducción y en otros elementos, como las pruebas de control de calidad de datos y los guiones de caracterización de poblaciones (véase más adelante). Por lo tanto, la variación en los conjuntos de valores puede dar lugar a la formación de diferentes dialectos. En consecuencia, garantizar que todos los sitios utilicen los mismos conjuntos de valores es fundamental para asegurar la coherencia entre los conjuntos de datos.

[00345] Las relaciones entre conceptos nuevos son aquellas que se establecen entre dos o más elementos descritos por conceptos existentes o nuevos. Estas

relaciones se utilizan frecuentemente en traducción para encontrar la mejor opción cuando no existe una traducción perfecta, como por ejemplo, de “Fractura de cuello femoral” a “Estructura del cuello femoral”. Por ejemplo, una terminología estándar podría no diferenciar entre distintos tipos de fracturas, por lo que una frase como

5 “fractura de tibia” podría traducirse como “fractura” por ser la alternativa más cercana y genérica. En este caso, la relación entre “fractura de tibia” en una terminología y “fractura” en otra puede utilizarse para lograr la mejor traducción posible. También es posible emplear múltiples relaciones para pasar de un concepto a otro cuando no existe una relación directa entre conceptos adyacentes. Las

10 relaciones entre conceptos nuevos son aquellas que se establecen entre un concepto nuevo y uno existente, entre dos conceptos nuevos o entre dos conceptos existentes.

[00346] Recursos novedosos: cuando un estándar de datos no cuenta con una forma específica de representar diferentes tipos de datos importantes para múltiples

15 sitios, se puede crear una columna (o recurso) para almacenarlos. Por ejemplo, OMOP utiliza una tabla de "personas" para registrar la información de los pacientes, pero esta tabla no incluye información de identificación como el nombre y el número de teléfono. Se pueden agregar recursos novedosos, en forma de columnas adicionales, a la tabla de personas para almacenar este tipo de datos adicionales.

20 En otro ejemplo, agregar datos de inventario de farmacia al estándar OMOP requeriría un recurso novedoso en forma de tabla. Los recursos novedosos pueden adoptar otras formas además de los dos ejemplos aquí presentados. Por ejemplo, en FHIR, los recursos novedosos tienen una estructura de subárbol.

[00347] Las pruebas de calidad de datos se utilizan antes, durante y después del

25 proceso de traducción para verificar y comparar los datos antes y después de la traducción. Estas pruebas pueden adoptar diversas formas, como consultas SQL o scripts de Python, y generan un resultado de aprobado o suspenso. En algunos casos, se utilizan cientos de pruebas simultáneamente para validar la calidad de los

datos. En otros casos, la calidad de los datos se determina mediante las pruebas y los aspectos de los datos que estas abarcan; por lo tanto, es fundamental que los distintos sitios que colaboran utilicen exactamente las mismas pruebas de calidad de datos. Los sistemas de traducción 1250 interpretan y utilizan estas pruebas en sus respectivos sitios de origen.

[00348] Los scripts de caracterización de población se diferencian de las pruebas de calidad de datos en que no generan una calificación de aprobado o suspenso. Los scripts de caracterización de población producen un valor, una distribución u otra indicación sobre un aspecto particular de la población, como la proporción de pacientes masculinos y femeninos en un conjunto de datos específico. Los resultados de los scripts de caracterización de población se utilizan para distinguir la variación en la población de pacientes de la variación dialectal, pero solo si todos los centros caracterizan a la población de la misma manera. Los sistemas de traducción 1250 interpretan y utilizan los scripts de caracterización de población en sus respectivos sitios de origen.

[00349] Las asignaciones de campos registran cómo se tradujeron los datos de origen. El formato exacto de las asignaciones depende tanto del origen como del destino. Por ejemplo, una asignación de campos para datos de un sistema de historia clínica electrónica a OMOP puede incluir la siguiente asignación: “Los datos del campo “paciente” de la tabla “Encuentro” en el conjunto de datos de origen se asignaron al campo “person_id” de la tabla “visit_occurrence” del estándar. Las asignaciones de campos también pueden ser más complejas e incluir varios campos, condiciones IF/ELSE u otra lógica. Se pueden utilizar diferentes lenguajes de programación para representar las asignaciones de campos. Un ejemplo de lenguaje de programación específico para asignaciones de campos es LinkML (<https://linkml.io/>).

[00350] Los mapas conceptuales relacionan una frase o un concepto codificado de un sistema con un concepto codificado de otro. Al igual que los mapas de campos,

los mapas conceptuales pueden representarse formalmente en diferentes lenguajes de programación. Un lenguaje específico para mapas de campos es el Estándar Simple para Compartir Mapeos de Ontologías (SSSOM) (<https://mapping-commons.github.io/sssom/>).

5 [00351] La representación formal de cada uno de estos elementos puede variar, desde el uso de fragmentos de lenguajes informáticos como Python o SQL, hasta lenguajes específicos para cada tipo de elemento. Sin embargo, todos los sistemas de traducción de datos deben ser capaces de analizar el elemento, ya sea directamente utilizando el mismo lenguaje o mediante otro servicio de traducción
10 que convierta una representación formal a otra.

[00352] El Servicio de Emisión de Identificadores 1210 de la figura 12 es un componente del sistema que conecta cada sitio de origen 1240 con el repositorio central 1220. Su función principal es garantizar que los elementos de traducción de datos con el mismo significado semántico tengan el mismo identificador único. El
15 Servicio de Emisión de Identificadores puede buscar en el repositorio central 1220 y emitir nuevos identificadores únicos si no existe un elemento equivalente.

[00353] Como se mencionó anteriormente, cuando el software de traducción 1250, que se ejecuta en un sitio de origen, detecta la necesidad de una extensión del estándar, utiliza el cliente de traducción 1260 para enviar una solicitud al Servicio
20 de Emisión de Identificadores 1210, solicitando un identificador único para la nueva extensión. El Servicio de Emisión de Identificadores 1210 verifica si ya existe una extensión equivalente en el repositorio central 1220. Si existe un elemento equivalente, el Servicio de Emisión de Identificadores 1210 recupera el identificador existente del repositorio central 1220 y lo devuelve al cliente de traducción 1260
25 solicitante, en lugar de emitir uno nuevo. El cliente de traducción 1260 reenvía el identificador único al software de traducción 1250.

[00354] Alternativamente, si no existe una extensión equivalente en el repositorio central 1220, el Servicio de Emisión de Identificadores 1210 genera un nuevo

identificador único para la nueva extensión, por ejemplo, mediante un generador de números aleatorios o similar, y devuelve el nuevo identificador único al cliente de traducción solicitante 1260. El Servicio de Emisión de Identificadores 1210 también transmite el identificador único recién generado al repositorio central 1220 para su almacenamiento asociado a la nueva extensión.

[00355] Algunas implementaciones del Servicio de Emisión de Identificadores también pueden sugerir elementos similares, aunque no idénticos al elemento buscado, mediante coincidencia aproximada, inteligencia artificial u otros métodos. Algunas implementaciones del Servicio de Emisión de Identificadores pueden integrarse como parte del repositorio.

Aplicación industrial

[00356] Las disposiciones descritas son aplicables a las industrias de la investigación, la medicina y la salud.

[00357] Lo anterior describe únicamente algunas modalidades de la presente invención, y se pueden realizar modificaciones y/o cambios sin apartarse del alcance y el espíritu de la invención, siendo las modalidades ilustrativas y no restrictivas.

[00358] En esta descripción, las referencias a “una modalidad”, “una modalidad”, “algunas modalidades” o “modalidades” significan que una característica, estructura o rasgo particular descrito en relación con la modalidad se incluye en al menos una modalidad de la presente invención. Por lo tanto, las apariciones de las frases “en una modalidad” en distintos lugares de esta memoria descriptiva no se refieren necesariamente a la misma modalidad, sino que pueden hacerlo. Además, las características, estructuras o rasgos particulares pueden combinarse de cualquier manera adecuada, como resultaría evidente para un experto en la materia a partir de esta descripción, en una o más modalidades.

[00359] Si bien algunas modalidades descritas aquí incluyen algunas características presentes en otras modalidades, pero no otras, se entiende que las combinaciones

de características de diferentes modalidades están comprendidas dentro del alcance de la invención y constituyen modalidades distintas, tal como lo comprenderían los expertos en la materia. Por ejemplo, en las siguientes reivindicaciones, cualquiera de las modalidades reivindicadas puede utilizarse en cualquier combinación.

[00360] Además, algunas de las modalidades se describen aquí como un método o una combinación de elementos de un método que puede ser implementado por un procesador de un sistema informático o por otros medios para llevar a cabo la función. Así, un procesador con las instrucciones necesarias para llevar a cabo dicho método o elemento constituye un medio para su ejecución. Así mismo, un elemento descrito en este documento, perteneciente a una modalidad de aparato, es un ejemplo de un medio para llevar a cabo la función que realiza dicho elemento con el fin de implementar la invención.

[00361] En la presente descripción se detallan varios aspectos específicos. No obstante, se entiende que algunas modalidades de la invención pueden llevarse a cabo sin necesidad de estos detalles. En otros casos, no se han mostrado en detalle métodos, estructuras y técnicas bien conocidos para facilitar la comprensión de esta descripción.

[00362] Téngase en cuenta que cuando se describe un método que incluye varios elementos, por ejemplo, varios pasos, no se implica ningún orden de dichos elementos, por ejemplo, de dichos pasos, a menos que se indique específicamente.

[00363] En el contexto de esta descripción, la palabra “que comprende” y sus construcciones gramaticales asociadas significan “que incluye principalmente, pero no necesariamente de forma exclusiva”, “que tiene” o “que incluye”, y no “que consiste únicamente en”. Las variantes de la palabra “que comprende”, como “comprender” y “comprende”, tienen significados correspondientemente variados.

[00364] De igual modo, cabe señalar que el término “acoplado” no debe interpretarse como limitante a conexiones directas. Se pueden utilizar los términos

“acoplado” y “conectado”, junto con sus derivados. Debe entenderse que estos términos no son sinónimos, aunque pueden serlo. Por lo tanto, el alcance de la expresión “un dispositivo A acoplado a un dispositivo B” no debe limitarse a dispositivos o sistemas donde una entrada o salida del dispositivo A esté conectada directamente a una entrada o salida del dispositivo B. Significa que existe una ruta entre el dispositivo A y el dispositivo B, la cual puede incluir otros dispositivos o medios intermedios. Además, “acoplado a” no implica dirección. Por consiguiente, la expresión “un dispositivo A está acoplado a un dispositivo B” puede ser sinónima de la expresión “un dispositivo B está acoplado a un dispositivo A”. “Acoplado” puede significar que dos o más elementos están en contacto físico o eléctrico directo, o que dos o más elementos no están en contacto directo entre sí, pero aun así cooperan o interactúan.

[00365] En toda la presente descripción, a menos que se especifique lo contrario, el uso de adjetivos ordinales como “primero”, “segundo”, “tercero”, “cuarto”, etc., para describir objetos comunes o relacionados, indica que se hace referencia a diferentes instancias de esos objetos comunes o relacionados, y no pretende implicar que los objetos así descritos deban proporcionarse o colocarse en un orden o secuencia determinados, ya sea temporal, espacial, jerarquizado o de cualquier otra manera.

[00366] Aunque la invención ha sido descrita con referencia a ejemplos específicos, los expertos en la materia comprenderán que la invención puede materializarse en muchas otras formas.