Datavant, Inc. (Petitioner)
v.
Vigilytics LLC (Patent Owner)

Patent Owner's Demonstratives

Case Nos. IPR2024-00307, -00311, -00381, -00382
U.S. Patent Nos. 9,965,651, 10,109,375, 10,886,012, 9,665,685

April 3, 2025

Before Honorable Administrative Patent Judges
Thomas L. Giannetti, Brian J. McNamara, and Steven M. Amundson

FISH.

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

Table of Contents

Overview of the Vigilytics Patents…….....….. 3

Ground 2: Evenhaim-Settimi / Evenhaim-Settimi-Murphy.….10

Ground 1: Evenhaim / Evenhaim-Murphy……..…….. 39

Ground 3: Dick-Landi-Murphy……..…….. 47

*All citations in these demonstratives are to papers/exhibits of record in IPR2024-00307 (’651 Patent) unless otherwise indicated.

Overview of the Vigilytics Patents

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

Overview of the Technology

• Vigilytics' patents describe novel solutions for providing third parties requested information that characterizes medical and/or healthcare data for a specific group of individuals who have not authorized disclosure of their protected health information (PHI) under HIPAA.

• Vigilytics was the first to solve the problem of how to provide such information for a **specific** group of individuals based on analysis of **de-identified** medical and/or healthcare data without authorization.

• Vigilytics' patents describe:

- **(1)** Isolating personally identifying information (PII) of the specific individuals in the group from the individuals' de-identified medical and/or healthcare data, thereby maintaining privacy while complying with HIPAA (e.g., using a 3+ party system);
- **(2)** Producing tokens by irreversibly transforming identifying information for specific individuals identified in a request to match those individuals to their de-identified medical and/or healthcare data associated with corresponding tokens; and
- **(3)** Analyzing the individuals' de-identified medical and/or healthcare data to produce irreversibly de-identified information that characterizes the medical and/or healthcare data for the group (e.g., producing risk scores characterizing a risk of insuring specific individuals or producing reports characterizing health outcomes for specific individuals following their discharge or other encounter with a healthcare facility).

•Datavant cites **no prior art** that produces medical and/or healthcare information for a **specific** group of individuals based on analysis of **de-identified** data.

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

[EX1001; POR, 2-7]

Overview of the Technology

•Vigilytics' patents protect patient privacy in part by isolating the identities of the specific individuals in the group from their de-identified medical and/or healthcare data. No entity can access both pieces of information.

By using the tokens **202** and request tokens **204**, no parties other than the authorized source sites **138** can associate de-identified data **140** with the identity of any person in the group of patients **134** or discharges (e.g., the group **154**). The hospital **150** requesting the report **212** can never receive data associated with individuals. Further, the information in the report **212** may be processed into metrics that characterize a large group and cannot be used to infer information about individuals. De-identified data records may be provided to the data processor **210** (e.g., a third party data processor), but that third party may not have access to any personally identifiable information **118** about the group **154**. Nor may any party with access to the de-identified data **140**, other than the healthcare providers **138**, also have access to the de-identifier **142**. These features of the system **200** maintain the privacy of the protected health information **130**.

EX1001, 5:26-42

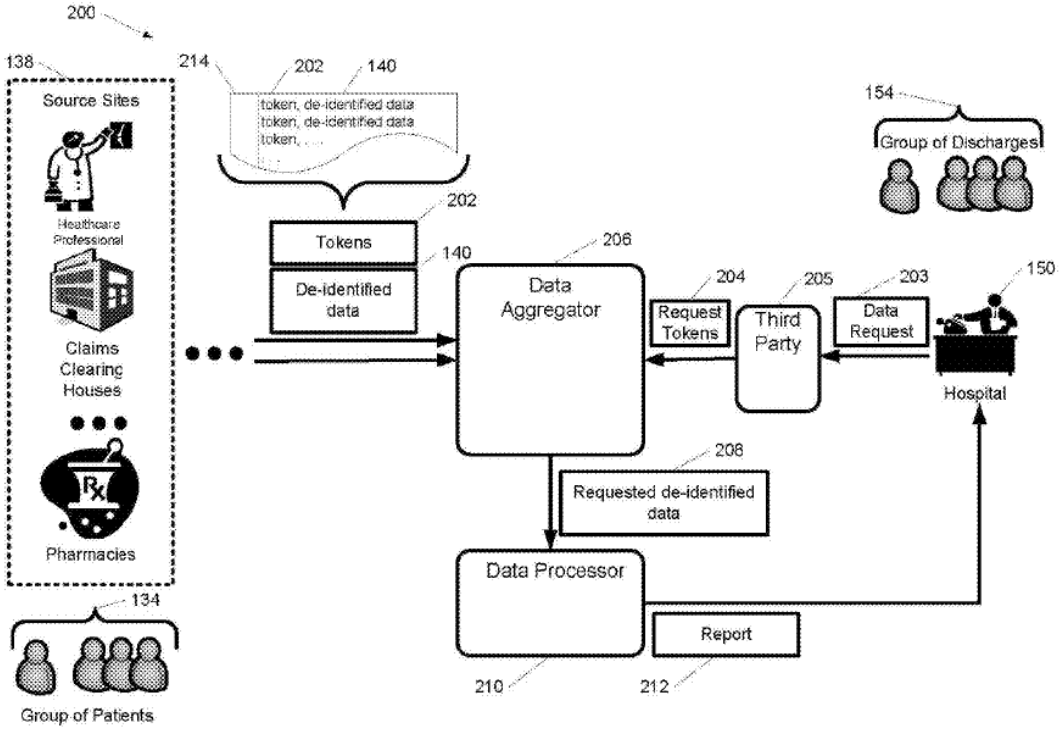
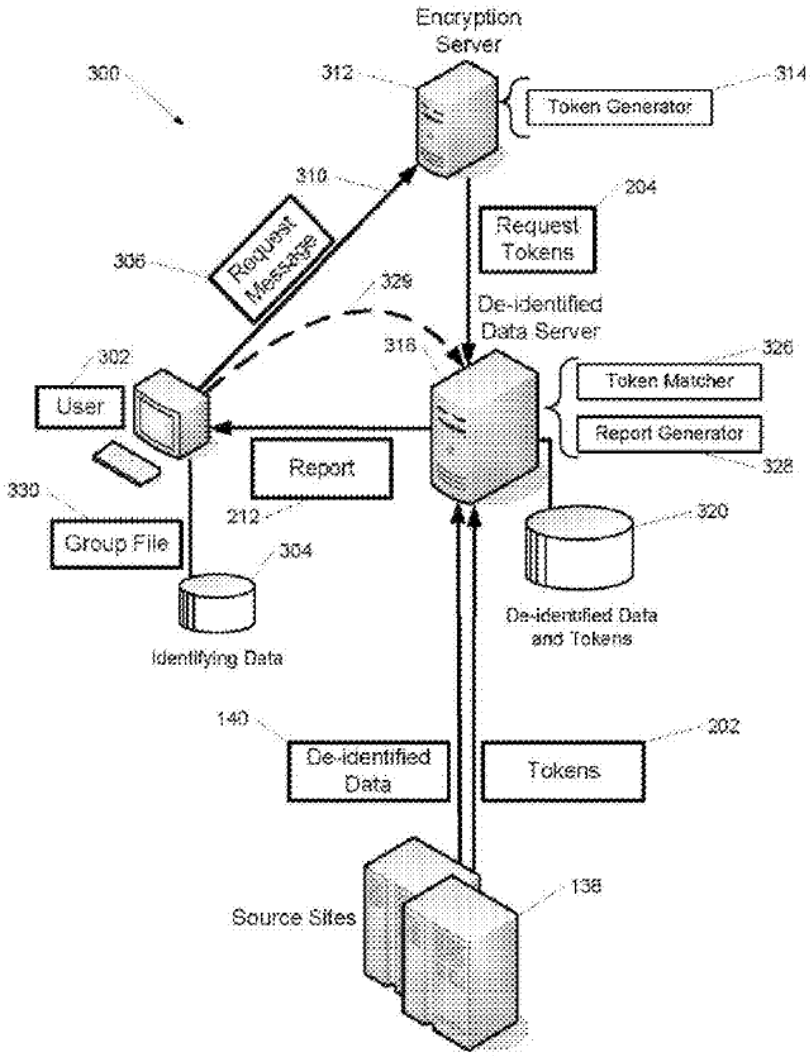


FIG. 2

EX1001, FIG. 2

Overview of the Technology



EX1001, FIG. 3

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

[POR, 5]

HIPAA Informs a POSA's Understanding of the Claimed Subject Matter

•Vigilytics' technology overcomes challenges resulting from restrictions imposed by the Health Insurance Portability and Accountability Act (HIPAA) to produce medical and/or healthcare information for a group of specific individuals while protecting the individuals' privacy.

•The problems and solutions described in the specifications of the Vigilytics patents arise specifically in the context of HIPAA. *See, e.g.*, EX1001, 1:25-27 ("relates to processing and transmitting personal data, the dissemination of which is restricted by federal law"), 1:28-62, 4:40-45.

•The claims of the Vigilytics patents likewise recite language reflecting concepts that arise specifically in the context of HIPAA. *See, e.g.*, EX1001, Claim 1 ("information that characterizes a group of multiple individuals from a set of ***de-identified*** medical data"; "providing the produced medical information … ***absent authorization*** from the individuals").

•Dr. Rubin explained that a POSA would read the specification and claims of the Vigilytics patents through the lens of HIPAA. *See* EX2001, ¶68; EX1025, 75:19-79:5.

•Datavant and Dr. Clark have not disputed that the subject matter of the Vigilytics patents operates in an environment involving HIPAA. *See, e.g.*, Pet., 2-3; EX1003, ¶¶34-36; EX2009, 13:18-44:15, 17:10-18.

Representative Independent Claims

’651 Patent (Claim 1)

Claim 1	
1.a	A computer-implemented method comprising: receiving a request from a user device for information that characterizes a group composed of multiple individuals from a set of de-identified medical data representing characteristics of the individuals pre- or post-encounter with a healthcare facility,
1.b	wherein the request contains identifying data for each member in the group of multiple individuals who have had an encounter with a healthcare facility;
1.c	producing, by a hardware processor, a request token for each individual included in the group of individuals;
1.d	comparing the request tokens to tokens associated with the de-identified data to find matching tokens and
1.e	to determine whether the group includes at least a minimum number of individuals to be included in a report in response to the request wherein the minimum number of individuals are selected according to at least a privacy-preserving rule that disallows same individuals to be included within a defined period of time;
1.f	producing the requested information that characterizes the group,
1.g	an identity of the individuals being unattainable from the produced information; and
1.h	providing the produced medical information to report the characteristics of the group of individuals absent authorization from the individuals in the group of individuals identified in the request

’685 Patent (Claim 1)

Claim 1	
1.a	A computer-implemented method comprising: receiving a request from a user device for medical information that characterizes a group composed of multiple individuals from a set of de-identified medical data representing pre- or post-encounter characteristics of the individuals,
1.b	wherein the request contains identifying data for each member in the group of multiple individuals;
1.c	producing, by a hardware processor, a request token for each individual included in the group of individuals;
1.d	comparing the request tokens to tokens associated with the de-identified medical data to find matching tokens;
1.e	producing the requested medical information, the identity of the individuals being unattainable from the produced medical information without the individuals in the group having authorized production of such medical information that characterizes the group; and
1.f	providing the produced medical information to report the characteristics of the group of individuals absent authorization from the individuals in the group of individuals identified in the request containing identifying data.

Representative Independent Claims

’375 Patent (Claim 1)

Claim 1	
1.a	A computer-implemented method comprising: receiving a request message for healthcare information that characterizes a group composed of a number of individuals,
1.b	wherein the request message for healthcare information is sent to an encryption server that is configured to determine if the number of individuals in the request is at least a defined minimum number of individuals,
1.c	and wherein the request message includes a batch number and at least one report option;
1.d	producing, by a processor of an encryption server, a unique request token from identifying data for each individual included in the group,
1.e	wherein the request tokens are encrypted before being transmitted over a network;
1.f	comparing the request tokens to tokens associated with healthcare data that includes de-identified information to find matching tokens;
1.g	producing the healthcare information with the identity of the individuals being unattainable from the produced healthcare information
1.h	based on a minimum number of matched tokens; and
1.i	providing the produced healthcare information to report the characteristics of the group absent authorization from the individuals in the group.

’012 Patent (Claim 1)

Claim 1	
1.a	A computer-implemented method comprising: producing, by a processor of the user computing device, an encrypted request message for healthcare information that characterizes a group composed of multiple individuals,
1.b	the encrypted request message including a unique token produced by the user device for each individual included in the group;
1.c	sending the encrypted request message to a server to produce the healthcare information,
1.d	the server is configured to compare the unique request tokens to unique tokens associated with healthcare data to find matching tokens,
1.e	the unique tokens associated with healthcare data are absent identities of individuals,
1.f	the healthcare information being produced based on a minimum number of matched tokens and the identity of the individuals being unattainable from the produced healthcare information; and
1.g	receiving, at the user computing device, the produced healthcare information to report the characteristics of the group
1.h	absent authorization from the individuals in the group.

Ground 2: The Evenhaim-Settimi / Evenhaim-Settimi-Murphy Combinations are Deficient

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

Overview of Evenhaim

Evenhaim sought to analyze "the original, most complete data set" from the data source entities

"Classic" Model

By contrast, an example of "classic" model of health analytics calls for aggregation of de-identified patient-level or physician-level data which allow the central analysis model only (see FIGS. **1** & **2**). The common method used by existing healthcare analytics companies (such as IMS, NDCHealth and Verispan) is based on the following steps: (A) De-identify & Aggregate—patient data is de-identified (some data is removed or grouped together) and data is reported typically at the physician level. In some cases the dataset includes longitudinal de-identified patient-level data (such as Verispan and Dendrite) (B) Collect from various sources—data is collected from multiple healthcare organizations into a data warehouse (C) Analyze—syndicated data reports and custom studies are produced.

Evenhaim's "Crossix" Method

The instant Crossix "method" is based on an opposite sequence: (A) Analyze—the healthcare organization (such as health plan) runs an analysis software that receive as an input the required analysis and aggregation level as well as possibly a list of identifiable patients or physicians for which analysis is requested. This analysis is performed on the original, most complete data set; (B) De-identify & Aggregate—Once name matching and analysis is complete data is aggregated and, if needed, de-identified; (C) and Collect from various sources—Analysis responses from multiple organization are composed to deliver the analyses requested.

EX1005 (Evenhaim), 17:28-52.

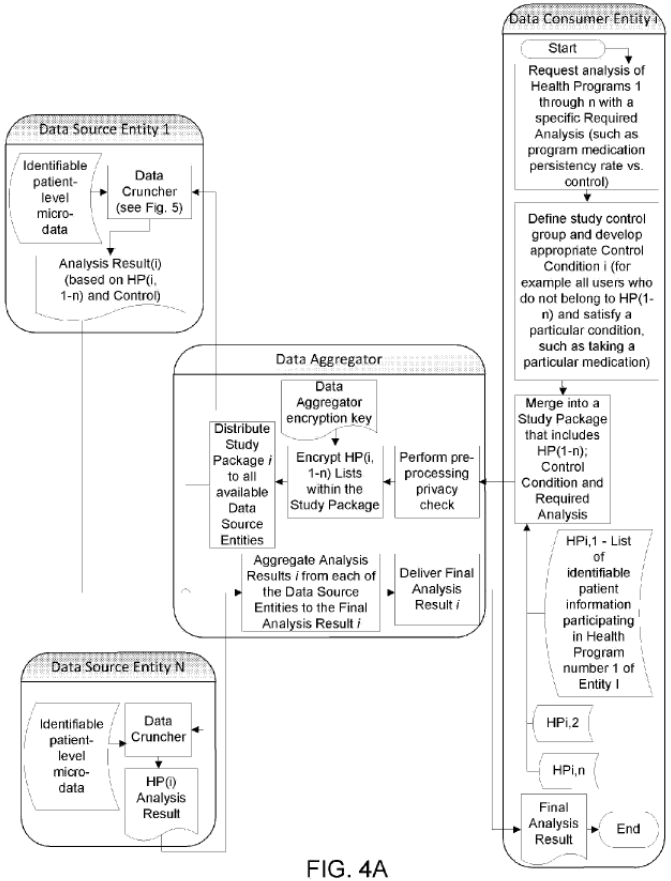
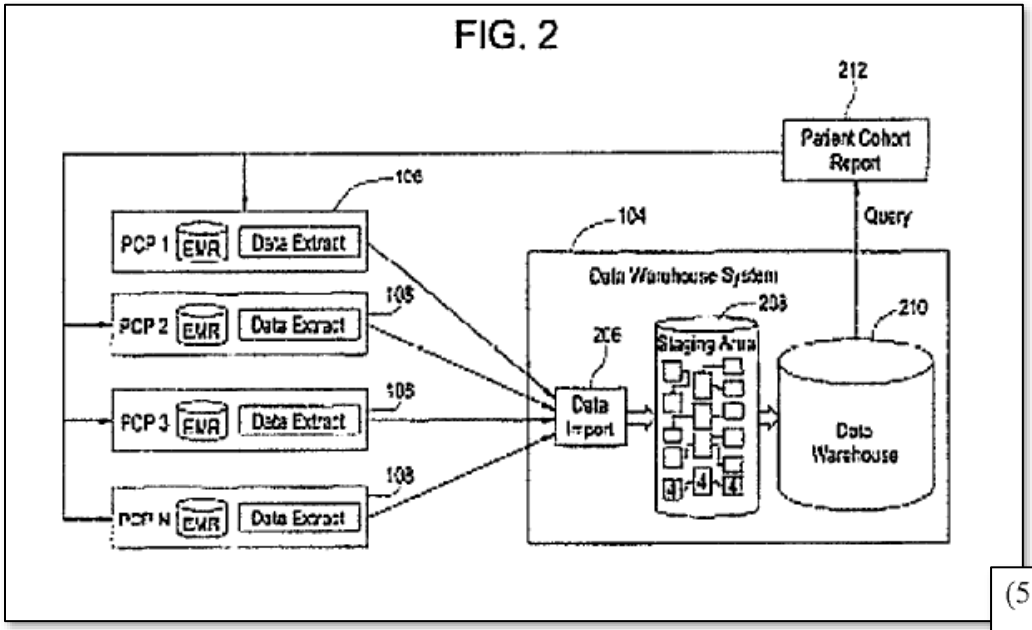


FIG. 4A

EX1005 (Evenhaim), FIG. 4A

Overview of Settimi



EX1006 (Settimi), FIG. 2

(57) **ABSTRACT**

Certain embodiments provide systems and methods for searching electronic medical data to identify a pool of potential study participants. Certain embodiments provide a method including the steps of receiving a plurality of search criteria, executing a search of electronic medical data based on each of the plurality of search criteria to create a pool of potential study participants and outputting the pool of potential study participants. In certain embodiments, the method may further include normalizing the plurality of search criteria according to at least one vocabulary. In certain embodiments, the method may allow a user to select the plurality of search criteria from a list of allowable criteria.

EX1006 (Settimi), Abstract

Overview of Proposed Evenhaim-Settimi Combination

• Key modifications to Evenhaim's system in the proposed combination with Settimi:

1. **De-identify** patients' medical and/or healthcare records **before** a search for individuals is initiated;
2. Consolidate de-identified medical and/or healthcare records in a **central repository**; and
3. Configure Evenhaim's data aggregator (e.g., Crossix) to produce encoded patient identifiers (i.e., EPIDs, the alleged "tokens") for each patient identified in a request for medical and/or healthcare information, and **use the EPIDs to search Settimi's central repository of de-identified records**.

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

[Pet., 38-42]

Datavant Fails to Show that a POSA Would be Motivated to Combine Evenhaim and Settimi as Alleged

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

The Combination is the Product of Impermissible Hindsight

Multiple Indicia of Hindsight:

1. A POSA would not find it obvious to make the combination because no known solution existed for matching a **specific** group of individuals to their **de-identified** medical and/or healthcare information while maintaining patient privacy in compliance with HIPAA.
2. Combination violates Evenhaim's principle of operation that involves searching original, high-resolution patient data, and correspondingly violates Evenhaim's teaching away from searching de-identified data.
3. Datavant's stated motivation for improving search efficiency by centralizing patient records is not justified against the cost of violating Evenhaim's principle of operation and teaching away.

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

[POR, 26-28, 40-46; POSR, 1-17]

The Combination is the Product of Impermissible Hindsight

Multiple Indicia of Hindsight:

4. Unlike Vigilytics' solution, the combination breaches patients' privacy and allows their identities to be discovered without authorization by combining patients' personally identifying information (PII) and de-identified medical and/or healthcare records at the requester's server.

5. The combination is illegal under HIPAA because it requires covered entities to disclose reversibly encrypted identification codes (Settimi's EPIDs). A POSA would not be motivated to implement an illegal system.

> "[T]he prejudice of hindsight bias often overlooks that the genius of invention is often a combination of known elements which in hindsight seems preordained."
>
> *Power Integrations, Inc. v. Fairchild Semiconductor Int'l, Inc.*, 711 F.3d 1348, 1368 (Fed. Cir. 2013)

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

[POR, 26-28, 40-46; POSR, 1-17; -00382 POSR, 1]

Hindsight Indicia 1: Searching Settimi's De-Identified Database for Specific Individuals Was Not Obvious

- Evenhaim acknowledges existing sources of de-identified medical and/or healthcare records, but found no solution for a third-party to search de-identified records for a group of specific individuals.
- Datavant's contention that it would be obvious to use tokens for this purpose is unsupported by contemporaneous evidence and belied by Evenhaim.

Evenhaim:

Most existing analytics services in healthcare are based on a model where patient-level health transactions are de-identified and then aggregated and processed for analysis. Clearly through this process, high-resolution raw data that includes the maximum amount of information is reduced to a lower resolution data to preserve privacy and possibly other interests. Many companies, including IMS Health, Verispan (a joint venture between Quintiles and McKesson), Pharmetrics, Dendrite and others employee this model. This existing health analytics model may provide either physician level behavior or de-identified patient-level behavior and possibly a combination of both. However, while this model is very effective in analyzing longitudinal patient behavior and, in some cases, matching it to a known prescriber (such as physician), no linkage can be made to a consumer health program that only impact a subset of the market (unless the program correlate well with a limited region or cover a known group of physicians). One cannot query or analyze parameters that are no longer in the lower resolution dataset. Therefore, the existing healthcare analytics model does not provide the capability to aggregate the behavior of a group of patients by a predefined list of consumers participating in a health program.

EX1005, 16:45-67

Hindsight Indicia 1: Searching Settimi's De-Identified Database for Specific Individuals Was Not Obvious

- Settimi does not teach or allow third parties to search the central repository for specific individuals using EPIDs.
- Settimi's central repository is instead searched by specifying attributes for a desired cohort of patients whose identities are unknown. The EPIDs are then used by covered entities to re-identify patients in the cohort.
- Datavant's proposed use of Settimi's EPIDs by a third-party for searching the central repository is unsupported.
- No embodiment in Settimi allows a third-party to convert PII for specific individuals to EPIDs to search the central repository.

Dr. Rubin:

133. In more detail, Settimi describes techniques for querying de-identified medical data stored in the central warehouse to surface patients who fit the clinical trial criteria specified in the request. For example, to identify candidate patients for a clinical trial, a user can query for "patients with a particular disease that were treated with a particular medication," "patients of a particular age and sex who have particular test results," or "all women with heart disease who are taking a hormone replacement therapy drug." EX1006, [0036]. A cohort of patients who meet the conditions specified in the request are identified from de-identified medical records stored in the central warehouse, and an indication of the cohort is returned to the requester or the primary care provider (PCP)—although only the PCP has the ability to re-identify patients by decrypting the EPIDs. *Id.*, [0037].

EX2001, ¶133

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

[POR, 40-44; POSR, 5-6]

Hindsight Indicia 1: Searching Settimi's De-Identified Database for Specific Individuals Was Not Obvious

- **Settimi allows only the covered entity (i.e., the PCP or data source entity) that created the EPID and de-identified the patient's record in the first instance to search for specific individuals using the EPID:**

> **[0047]** In certain embodiments, a user may search for one or more patient records within EMR by invoking a "find" dialog or search function. The user may search by the EPID, for example, and enter or select an EPID number to activate a search. The corresponding patient chart may be retrieved and displayed. Thus, a patient may be re-identified for an authorized healthcare provider who has been identified and verified.

EX1006, ¶47

Hindsight Indicia 2: The Combination Violates Evenhaim's Principle of Searching Original, High-Resolution Data

- **Dr. Clark's conclusory assertions that it would be obvious to produce tokens for searching de-identified records for specific individuals is not credible.**

> Q. Do you have any systems before the Vigilytics patent in which a third party encrypted PII to generate request tokens that could be matched against a de-identified dataset?
>
> A. Not from memory.

EX2009, 70:7-11

> Q. Have you ever implemented a system that performed a pre- or post-encounter analysis?
>
> A. I mean, I've implemented expert systems before, but I don't believe I implemented an expert system in healthcare, no.

EX2009, 39:18-22

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

[POSR, 6, 20-21; POR, 40-44; EX1015, ¶12]

Hindsight Indicia 2: The Combination Violates Evenhaim's Principle of Searching Original, High-Resolution Data

- **Evenhaim explicitly sought to search original patient data before de-identification.**
- **Datavant's combination disregards Evenhaim's express preferences.**

Approach Disparaged by Evenhaim:

By contrast, an example of "classic" model of health analytics calls for aggregation of de-identified patient-level or physician-level data which allow the central analysis model only (see FIGS. **1** & **2**). The common method used by existing healthcare analytics companies (such as IMS, NDCHealth and Verispan) is based on the following steps: (A) De-identify & Aggregate—patient data is de-identified (some data is removed or grouped together) and data is reported typically at the physician level. In some cases the dataset includes longitudinal de-identified patient-level data (such as Verispan and Dendrite) (B) Collect from various sources—data is collected from multiple healthcare organizations into a data warehouse (C) Analyze—syndicated data reports and custom studies are produced.

Evenhaim's "Crossix" Method:

The instant Crossix "method" is based on an opposite sequence: (A) Analyze—the healthcare organization (such as health plan) runs an analysis software that receive as an input the required analysis and aggregation level as well as possibly a list of identifiable patients or physicians for which analysis is requested. This analysis is performed on the original, most complete data set; (B) De-identify & Aggregate—Once name matching and analysis is complete data is aggregated and, if needed, de-identified; (C) and Collect from various sources—Analysis responses from multiple organization are composed to deliver the analyses requested.

EX1005, 17:28-52

Hindsight Indicia 2: The Combination Violates Evenhaim's Principle of Searching Original, High-Resolution Data

ADVANTAGES, OBJECTS AND BENEFITS OF THE INVENTION

Ergonomic Issues: Preferred embodiments of the instant invention allow analysis of "source-entity" raw data at is original, most detailed form (high resolution data) including full access to all of the privacy sensitive data currently at its disposal while simultaneously maintaining existing privacy restrictions to the aggregator processor. In addition, high-resolution analysis may be performed at multiple different "source-entities" each of which preserves its privacy restrictions, yet under certain conditions the data can be aggregated by the aggregator processor to provide a more comprehensive analysis.

Economic Issues: Preferred embodiments of the instant invention allow exploitation of an order of magnitude more value potential from the data currently resident at the "source-entity" processors while only adding nominal expenses at the "aggregator" processor. Furthermore, expenses at the "aggregator" processor are essential to define and provide new avenues of access to the ensemble of privacy sensitive data located at the "source-entity" processors.

EX1005, 6:12-32

Hindsight Indicia 2: The Combination Violates Evenhaim's Principle of Searching Original, High-Resolution Data

Dr. Rubin:

136. The Petition's proposed combination of Evenhaim-Murphy with Settimi would not have been obvious to a POSITA for the additional reason that it would have violated Evenhaim's express directive and principle of operation for searching on original, micro-level data before de-identification and aggregation. EX1005, 6:15–34 ("instant invention allow[s] analysis of 'source-entity' raw data at its original, most detailed form (high resolution data) including full access to all of the privacy sensitive data currently at its disposal"), 17:42–52 ("analysis is performed on the original, most complete data set"), 17:28–41 (contrasting Evenhaim's purported invention with "classic" approaches that called for "aggregation of de-identified patient" before searching).

EX2001, ¶136

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

[POR, 44-46]

Hindsight Indicia 2: The Combination Violates Evenhaim's Principle of Searching Original, High-Resolution Data

Federal Circuit:

"[E]ven if a reference is not found to teach away, its **statements regarding preferences are relevant to a finding regarding whether a skilled artisan would be motivated to combine** that reference with another reference."

"[B]y completely disregarding certain teachings as ill-defined 'subjective preferences,' the Board's approach invited the '**distortion caused by hindsight bias**' into the fold. … We have observed that 'the prejudice of hindsight bias' often overlooks that the 'genius of invention is often a combination of known elements which in hindsight seems preordained.'

Polaris Indus., Inc. v. Arctic Cat, Inc., 882 F.3d 1056, 1069 (Fed. Cir. 2018)

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

[POR, 26-30, 44-46; POSR, 1-8]

Hindsight Indicia 2: Evenhaim Teaches Away from the Combination's Use of De-Identified Data

Evenhaim: De-identified Data is "Useless"

(I) In-sufficiency of target consumer program measurement while data exist because of privacy issues—The challenge of measuring the impact of a consumer health program becomes significant whenever the health program sponsor does not have the full healthcare information of the target population at their disposal. Blocked by both access to data as well as privacy challenges, sponsoring organizations have to assess the impact of their efforts with very limited methods. As described above in this section, HIPAA provides substantial limitation on Personal Health Information and existing de-identification method may render the information useless for the purpose of measuring the impact of health programs. Naturally, with limited measurement abilities, less resources are directed by sponsors to valuable health programs such as compliance programs.

EX1005, 4:34-48

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

Hindsight Indicia 2: Evenhaim Teaches Away from the Combination's Use of De-Identified Data

Evenhaim further observed that existing models that relied on **de-identified data** for longitudinal searches were inadequate because "[o]ne **cannot query or analyze parameters that are no longer in the lower resolution database**" (see 16:63-64), especially as such models "**preclude[] external researchers from benefiting from the richness of the 'source-entity' data**." EX1005, 16:63-64, 7:46-52.

"A reference may be said to teach away when a person of ordinary skill, upon reading the reference, would be discouraged from following the path set out in the reference, or would be led in a direction divergent from the path that was taken by the applicant."

DePuy Spine, Inc. v. Medtronic Sofamor Danek, Inc., 567 F.3d 1314, 1327 (Fed. Cir. 2009)

Hindsight Indicia 3: Datavant's Stated Motivation to Improve Search Efficiency through Centralization is Deficient

- **Datavant cites no evidence that the speculative efficiency gains flowing form a centralized search would have justified the high cost that Evenhaim attributes to searching de-identified medical records.**

Dr. Clark:

137. While Evenhaim sends the queries for healthcare data to various healthcare organizations, health plans (e.g., HMOs), and pharmacy benefit managers, etc. Settimi provides the added advantage of consolidation information from healthcare sources into a central repository for queries. *See* Ex. 1005, 15:15-24, 16:5-21, Fig. 4; Ex. 1006, Figs. 1 & 2, [0027]. The repository in Settimi provides one central database for conducting searches and also uses a "common vocabulary" that makes searching easier and more reliable (e.g., the searcher does not need to know the proper search terms used by different plans/providers). Ex.

EX1003, ¶137

Vigilytics:

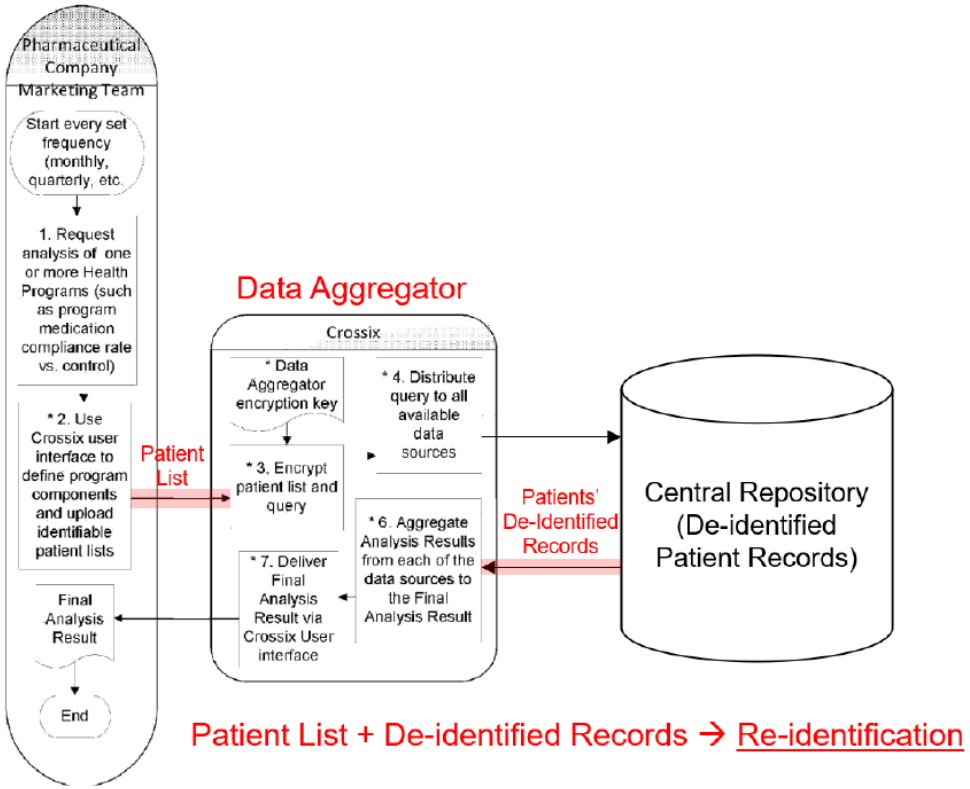
high-resolution records ***before de-identification***. EX2001, ¶¶125, 136-137. Even if the patient records were required to be de-identified before their release to the central repository to comply with HIPAA, the Reply cites ***no*** evidence that the speculative efficiency gains flowing from a centralized search would have justified the high cost that Evenhaim attributes to searching de-identified medical records as opposed to the original records in their most complete form. The law is

POSR 2-3

[Pet., 41; POSR, 2-3]

Hindsight Indicia 4: The Combination Presents a Substantial Re-Identification Risk That a POSA Would Not Pursue

Evenhaim-Settimi-Murphy Combination:



Evenhaim-Settimi-Murphy Illustration

[POSR, 9]

Vigilytics Patents:

By using the tokens **202** and request tokens **204**, no parties other than the authorized source sites **138** can associate de-identified data **140** with the identity of any person in the group of patients **134** or discharges (e.g., the group **154**). The hospital **150** requesting the report **212** can never receive data associated with individuals. Further, the information in the report **212** may be processed into metrics that characterize a large group and cannot be used to infer information about individuals. De-identified data records may be provided to the data processor **210** (e.g., a third party data processor), but that third party may not have access to any personally identifiable information **118** about the group **154**. Nor may any party with access to the de-identified data **140**, other than the healthcare providers **138**, also have access to the de-identifier **142**. These features of the system **200** maintain the privacy of the protected health information **130**.

EX1001, 5:26-42

Hindsight Indicia 4: The Combination Presents a Substantial Re-Identification Risk That a POSA Would Not Pursue

Dr. Rubin:

Petitioner's interpretation. But Evenhaim is entirely silent in this regard. Worse, the embodiment invented in the Petition would itself risk a HIPAA violation if the data source entities released de-identified patient data to the data aggregator that included the patient's token, as would occur to permit linking of data records at the aggregator. EX1005, 12:53–13:20. In this case, the data aggregator would be capable of comparing the token it received from the source entity to the token it earlier sent in the queries, thereby permitting the aggregator to correlate the patient's name or other identifiers from the patient list to de-identified patient-level data released from the source entities in violation of the patient's privacy and regulations such as the HIPAA Privacy Rule. *Id.*

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

[POR, 16; POSR, 8-9]

Hindsight Indicia 4: The Combination Presents a Substantial Re-Identification Risk That a POSA Would Not Pursue

- Dr. Clark conceded that it "would not be prudent" to combine names with health conditions and that he "wouldn't recommend anyone do that."

Dr. Clark:

Q. If the third-party possessed both the de-identified data and the names of the individuals in the group, how does re-identification occur?

A. If I have your name and I have a health condition, it's possible that I could, depending upon whether they are using pseudonyms or not or whether the name was Joe Smith, that I may or may not be able to re-identify you, but I would not implement a system in which the patient's names were available. That's for sure.

Q. And why would you not do that?

A. Because I would not want someone with a unique name to have their name out there with a health condition that then would allow me to infer the protected health information. That would not be prudent.

Q. That would be a re-identification risk?

A. I think it could be. I mean, I -- I wouldn't recommend anyone to do that.

EX2009, 30:17-31:7

Hindsight Indicia 5: The Combination Violates HIPAA by Requiring Disclosure of Reversibly Encrypted Identification Codes (Settimi's EPIDs)

- Settimi's EPIDs are formed by **reversibly encrypting patient identifying information (PII)** (e.g., PID, name, social security number).
- Settimi teaches that only the PCP's password string is hashed—i.e., not the PII, which is reversibly encrypted.

[0041] A novel EPID **308** is assigned to each patient based on the PID **304** associated with the patient and a password entered by the PCP. The PID **304** to EPID **308** mapping is not maintained persistently. As depicted in the exemplary embodiment shown in FIG. **3**, a password string **312** is supplied by the PCP via a password encryption user interface **310** on the PCP user system **110**. This password string **312** is known only to the PCP and is required in order to decode the EPID **308** into a PID **304**. The user at the PCP site must have the password string **312** to obtain the PID **304** and this password string **312** must be re-entered each time a patient is to be re-identified. The password encryption user interface **310** may be a graphical user interface. In an exemplary embodiment of the present invention, the user entered password string **312** is encoded using the two-fish algorithm. The two-fish algorithm, as known in the art, is a secret-key block cipher cryptography algorithm that is designed to be highly secure and highly flexible. It utilizes a single key for both encryption and decryption and is often referred to as symmetric encryption. The encoding is performed by patient identifier encoding software **306** located on the PCP system **108**. The patient identifier encoding software **306** also hashes the encoded password string to produce a sixteen-digit number. This sixteen-digit number is numerically added to the PID **304** to create the EPID **308**. Other methods of creating the EPID **308** from the PID **304** may be utilized with an exemplary embodiment of the present invention (e.g. Rivest, Shamir and Adelman, RSA, algorithm based on patient name, age and social security number, etc.) as long as the EPID may only be decoded at the PCP site.

EX1006, [0041]

[POSR, 12-16; POR, 51-52]

Hindsight Indicia 5: The Combination Violates HIPAA by Requiring Disclosure of Reversibly Encrypted Identification Codes (Settimi's EPIDs)

HIPAA Privacy Rule (45 CFR § 164.514(c))

(c) ***Implementation specifications: Re-identification.*** A covered entity may assign a code or other means of record identification to allow information de-identified under this section to be re-identified by the covered entity, provided that:

(1) ***Derivation.*** The code or other means of record identification is not derived from or related to information about the individual and is not otherwise capable of being translated so as to identify the individual; and

(2) ***Security.*** The covered entity does not use or disclose the code or other means of record identification for any other purpose, and does not disclose the mechanism for re-identification.

EX2006, 2

Dr. Rubin:

individual identifying information." EX1008, 2:46–48; *id.*, 5:13–19. By generating and releasing reversibly Encrypted IDs based on individual identifying information of patients, Landi violates HIPAA's proscription against the use of codes for re-identification that are "derived from or related to information about the individual."[4] 45 C.F.R. §164.514I(1). A POSITA therefore would not have turned to Landi for its specific de-identification methodology that involves creating Encrypted IDs because Landi's approach would not have achieved the Petition's stated goal of achieving HIPAA compliance. Landi instead invites a HIPAA violation, which a POSITA would have sought to avoid.

EX2001, ¶144

Dr. Clark:

Q. So Landi's encrypted ID does not comply with the derivation requirement. Correct?

A. I'm not sure that's true because you have to have access to the key in order to do it; but generally speaking, in that embodiment, you could argue that it's not combined with a derivation requirement.

EX2009, 54:23-55:1-3

The Petition's reliance on Settimi in Ground 2 suffers from a similar defect that spoils the proposed Evenhaim-Settimi combination. Like Landi, Settimi generates its encoded patient identifier (EPID) based on information about the individual. EX1006, [0038], [0041]–[0043], [0056].

[POSR, 12-16; POR, 51-52]

Hindsight Indicia 5: The Combination Violates HIPAA by Requiring Disclosure of Reversibly Encrypted Identification Codes (Settimi's EPIDs)

- **HHS Guidance identifies no reversible encryption functions that would comply with HIPAA Privacy Rule. Only suitable function identified is a one-way (hash) function.**
- **Emphasizes that disclosure of keys associated with hash function is not permitted.**

Can an Expert determine a code derived from PHI is de-identified?

There has been confusion about what constitutes a code and how it relates to PHI. For clarification, our guidance is similar to that provided by the National Institutes of Standards and Technology (NIST)[29], which states:

"*De-identified information can be re-identified (rendered distinguishable) by using a code, algorithm, or pseudonym that is assigned to individual records. The code, algorithm, or pseudonym should not be derived from other related information* about the individual, and the means of re-identification should only be known by authorized parties and not disclosed to anyone without the authority to re-identify records. A common de-identification technique for obscuring PII [Personally Identifiable Information] is to use a one-way cryptographic function, also known as a hash function, on the PII.*

**This is not intended to exclude the application of cryptographic hash functions to the information."*

In line with this guidance from NIST, a covered entity may disclose codes derived from PHI as part of a de-identified data set if an expert determines that the data meets the de-identification requirements at §164.514(b)(1). The re-identification provision in §164.514(c) does not preclude the transformation of PHI into values derived by cryptographic hash functions using the expert determination method, provided the keys associated with such functions are not disclosed, including to the recipients of the de-identified information.

EX1016, 30-31

[POSR, 14]

Datavant Fails to Show that the Evenhaim-Settimi / Evenhaim-Settimi-Murphy (ESM) Combinations Disclose Multiple Claim Limitations

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

ESM Fails to Disclose "A Privacy-Preserving Rule that Disallows Same Individuals to be Included within a Defined Period of Time" ('651 Element [1e])

Murphy:

As the numbers that are being returned to the query tool are obfuscated, a history table of true numbers is maintained. This history table is seen only by the system, and contains for each query the date/time, the user ID, and the number of patients truly returned by the queries. A trigger each time a query is performed queries this history table. The query looks at the numeric results of queries performed by that user to see if the exact same number of patients were returned in previous queries within a specific time period. Depending on the time range employed for obfuscation, when a certain number of queries return the exact same number of patients, the user is prevented from performing further queries. Eventually, the account is reviewed and the lock is released if it is determined that the same number of patients were returned by the queries purely by coincidence.

EX1009, 554

Dr. Rubin:

128. Petitioner's obviousness contentions with respect to "wherein the minimum number of individuals are selected according to at least a privacy-preserving rule that disallows same individuals to be included within a defined period of time"—as recited in elements [7.h], [1.e], and [14.g]—are based on gross mischaracterizations of Murphy and should be rejected. The Petition contends that Murphy describes "***auditing queries in a history table*** to check the date/time and the number of individuals returned by the queries ***to disallow repeated use of the same individual in a query***." Pet., 27. This is entirely false. To "disallow repeated use of the same individual" using the history table, Murphy's system would need to track identities of the patients/individuals in the history table - yet Murphy never describes such features. EX1009, 554. Murphy's discussion of the history table is

EX2001, ¶128

Independent claims 7 and 14 of the '651 Patent recite similar language and the deficiencies in ESM similarly apply to these additional claims.

[POR, 35-40; POSR, 17-19]

ESM Fails to Disclose "Producing[] … a Request Token for Each Individual Included in the Group of Individuals" and "Comparing the Request Tokens to Tokens Associated with the De-Identified Data to Find Matching Tokens" ('651 Elements [1c]-[1d])

Vigilytics:

Datavant is still unable to show in the Reply any prior art that produces a request token for each individual identified in a request for medical information, and that then compares the request tokens to tokens associated with de-identified data to find matching tokens, as recited in [1.c]-[1.d]. To this point, Evenhaim does not rely on matching tokens in this manner but instead searches original patient data before de-identification. *Supra*, §II.A.1; POR, 26-28. Settimi's central database is likewise deficient since it is queried by demographic or clinical criteria to identify patient cohorts and is never queried by tokens to identify medical records for a specific group of patients.[5] *Supra*, §II.A.3; POR, 41-44. Lacking any express disclosure of these features, Datavant resorts to uncorroborated assertions that it would nonetheless be obvious for "the patient records [to] be queried using tokens because tokens (*e.g.*, 'EPIDs') are the only language Settimi's centralized database understands for identifying records."

***Arendi S.A.R.L. v. Apple Inc.*, 832 F.3d 1355, 1362 (Fed. Cir. 2016)** ("[O]ur cases repeatedly warn that references to 'common sense' … cannot be used as a wholesale substitute for reasoned analysis and evidentiary support, especially when dealing with a limitation missing from the prior art references specified").

Independent claims 7 and 14 of the '651 Patent, and independent claims 1, 7, and 14 of the '375, '012, and '685 Patents, recite similar language and the deficiencies in ESM similarly apply to these additional claims.

[POSR, 20-21]

ESM Fails to Disclose "The Unique Tokens Associated with Healthcare Data are Absent Identities of Individuals ('012 Element [1e]) and "An Identity of the Individuals Being Unattainable From the Produced Information" ('651 Element [1g])

Datavant contends that the identity of the individuals is "unattainable" from Settimi's EPIDs (which are part of the produced information in the combination), and argues with respect to similar features in Ground 3 that the identity of an individual corresponding to reversibly encrypted PII is rendered unattainable "by protecting the decryption key." Reply, 23-24; EX1006, [0027] ("re-identified at the PCP"), [0037], [0041] ("symmetric encryption"), [0042]-[0048]. But the claims do not condition the claimed unattainability to only those entities that lack a decryption key. POR, 63-64. The '651 Patent instead distinguishes reversibly encrypted information for which the identities of individuals remain attainable from irreversibly deidentified information for which the identities of individuals are unattainable. EX1001, 7:36-38, 4:17-21, 7:13-21 (differentiating reversibly encrypted request message 306 carrying PII from irreversibly deidentified data 140/204/212); EX2009, 46:16-48:14 (Dr. Clark describing difference between reversible encryption and irreversible de-identification). Evenhaim-Settimi-Murphy thus fails to disclose or render obvious [1g].

[POSR, 19]

Independent claims 7 and 14 of the '651 Patent, and independent claims 1, 7, and 14 of the '375, '012, and '685 Patents, recite similar language and the deficiencies in ESM similarly apply to these additional claims.

ESM Fails to Disclose "Providing the Produced Medical Information … Absent Authorization from the Individuals" ('651 Element [1h])

Evenhaim:

relevant disease information. These pharmaceutical health programs are promoted to consumers of that product who then opt-in to participate in the program. The participants of each program allow, among other things, the pharmaceutical manufacturer to analyze their data in aggregate. Some of the

EX1005, 15:34-37

Health Program—a program (used as specific example for the preferred embodiment of the instant invention) that affects a subset of the overall potential population. Typically patients, consumers or healthcare professionals will opt-in to participate in such a program and if the organization sponsoring it is not covered by HIPAA, the sponsoring organization will adhere to its published privacy policy. Typically Health

EX1005, 5:46-52

Dr. Clark:

Q. What does it mean to opt in to a health program?

A. That means to agree to participate.

Q. So if patient is agreeing to participate in the health program, they would be authorizing the use of their medical information for that program. Correct?

A. I would expect so.

EX2009, 77:12-18

Independent claims 7 and 14 of the '651 Patent, and independent claims 1, 7, and 14 of the '375, '012, and '685 Patents, recite similar language and the deficiencies in ESM similarly apply to these additional claims.

[POSR, 21-22]

Ground 1: The Evenhaim / Evenhaim-Murphy Combinations are Deficient

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

Evenhaim Fails to Disclose Matching Specific Individuals to "De-identified" Medical and/or Healthcare Information

- All claims of the ’685, ’375, and ’651 patents operate on “**de-identified**” medical and/or healthcare data for a group of specific individuals identified in a request.
- By contrast, Evenhaim searches the original, full patient records **before de-identification**. Evenhaim instead teaches away from searching de-identified data. Evenhaim therefore does not disclose or render obvious any claim of the ’685, ’375, or ’651 patents.

Approach Disparaged by Evenhaim:

By contrast, an example of “classic” model of health analytics calls for aggregation of de-identified patient-level or physician-level data which allow the central analysis model only (see FIGS. **1** & **2**). The common method used by existing healthcare analytics companies (such as IMS, NDCHealth and Verispan) is based on the following steps: (A) De-identify & Aggregate—patient data is de-identified (some data is removed or grouped together) and data is reported typically at the physician level. In some cases the dataset includes longitudinal de-identified patient-level data (such as Verispan and Dendrite) (B) Collect from various sources—data is collected from multiple healthcare organizations into a data warehouse (C) Analyze—syndicated data reports and custom studies are produced.

Evenhaim’s “Crossix” Method:

The instant Crossix “method” is based on an opposite sequence: (A) Analyze—the healthcare organization (such as health plan) runs an analysis software that receive as an input the required analysis and aggregation level as well as possibly a list of identifiable patients or physicians for which analysis is requested. This analysis is performed on the original, most complete data set; (B) De-identify & Aggregate—Once name matching and analysis is complete data is aggregated and, if needed, de-identified; (C) and Collect from various sources—Analysis responses from multiple organization are composed to deliver the analyses requested.

EX1005, 17:28-52

[POR, 26-28; POSR, 24]

Evenhaim Fails to Disclose Matching Specific Individuals to "De-identified" Medical and/or Healthcare Information

- Datavant's citation to Evenhaim's "identity-free micro-data" is misplaced. Identity-free micro-data results in Evenhaim from de-identification of the "extract" **after** performing the search and after patients' records are identified based on PII.

Reply:

> include specific individuals by name or ID or the like[]"). Also, Evenhaim acknowledges that the "micro-level records" (*i.e.*, patient level data) that it seeks to maintain are not harmed by the removal of PII inasmuch as Evenhaim references "*identity-free* micro-data." *Id.*, 10:29-32, 12:36-52, Fig. 4 ("micro data"), 12:35-42;

Reply, 7

Evenhaim:

> According to a second preferred embodiment of the instant
> 30 invention, agglomerating the extract includes filtering out portions of the extract so that only identity-free micro-data or identity-free aggregated data remains.

EX1005, 10:29-32

[POSR, 24]

Evenhaim Fails to Disclose Matching Specific Individuals to "De-identified" Medical and/or Healthcare Information

- Datavant's observation that "Evenhaim describes searches in which the patient is identified other than by his/her name" does not imply that the search is performed on de-identified data. Evenhaim instead explains that a variety of personally identifying information (PII) about a user can be used in a search of raw patient data.

Reply:

> Specifically, Evenhaim describes searches in which the patient is identified other than by his/her name. EX1005, 8:30-43 ("a list of names or a combination of a few

Reply, 7

Evenhaim:

> Simply stated, the aggregator may accept a query that includes lots of identity specifics (e.g. a list of names or a combination of a few fields that together can allow identification)—the type of query that one does not expect to be answerable in any privacy preserving fashion. This sub-step

EX1005, 10:29-32

[POSR, 7-8]

Evenhaim Fails to Disclose "Producing[] … a Request Token for Each Individual Included in the Group of Individuals" and "Comparing the Request Tokens to Tokens Associated with the De-Identified Data to Find Matching Tokens" ('651 Elements [1c]-[1d])

102. Encrypting entire patient lists (and queries), as disclosed in Evenhaim, is far different from producing a token for each patient named in the list, which would be inefficient and unnecessary for data transmission security in any event. Evenhaim is clear that only the patient lists, and not identifiers for each individual in the list, are encrypted at the data aggregator, thus permitting the data source entities to decrypt the lists with an appropriate decryption key. EX1005, 14:3–7 ("***Returning to the abovementioned aspects of preferable encryption***, it is certainly strongly advised that regardless of the general level of encryption elected, ***a parameter list including 'individuals'*** should carry a stronger level of ***encryption***."), 15:61–65 ("***Encrypting the lists*** using any industry grade encryption method; and Delivering the lists of participants together to a query engine with a capability to ***decrypt the lists*** at the time of processing …"), FIG. 5 (showing decryption of each "patient identifier ***list***").

[POR, 14-16]

Evenhaim Fails to Disclose "Providing the Produced Medical Information … Absent Authorization from the Individuals" ('651 Element [1h])

Evenhaim:

relevant disease information. These pharmaceutical health programs are promoted to consumers of that product who then opt-in to participate in the program. The participants of each program allow, among other things, the pharmaceutical manufacturer to analyze their data in aggregate. Some of the

EX1005, 15:34-37

Health Program—a program (used as specific example for the preferred embodiment of the instant invention) that affects a subset of the overall potential population. Typically patients, consumers or healthcare professionals will opt-in to participate in such a program and if the organization sponsoring it is not covered by HIPAA, the sponsoring organization will adhere to its published privacy policy. Typically Health

EX1005, 5:46-52

Dr. Clark:

Q. What does it mean to opt in to a health program?

A. That means to agree to participate.

Q. So if patient is agreeing to participate in the health program, they would be authorizing the use of their medical information for that program. Correct?

A. I would expect so.

EX2009, 77:12-18

[POSR, 21-22, 25]

Evenhaim-Murphy Fails to Disclose "A Privacy-Preserving Rule that Disallows Same Individuals to be Included within a Defined Period of Time" ('651 Element [1e])

Murphy:

As the numbers that are being returned to the query tool are obfuscated, a history table of true numbers is maintained. This history table is seen only by the system, and contains for each query the date/time, the user ID, and the number of patients truly returned by the queries. A trigger each time a query is performed queries this history table. The query looks at the numeric results of queries performed by that user to see if the exact same number of patients were returned in previous queries within a specific time period. Depending on the time range employed for obfuscation, when a certain number of queries return the exact same number of patients, the user is prevented from performing further queries. Eventually, the account is reviewed and the lock is released if it is determined that the same number of patients were returned by the queries purely by coincidence.

EX1009, 554

Dr. Rubin:

128. Petitioner's obviousness contentions with respect to "wherein the minimum number of individuals are selected according to at least a privacy-preserving rule that disallows same individuals to be included within a defined period of time"—as recited in elements [7.h], [1.e], and [14.g]—are based on gross mischaracterizations of Murphy and should be rejected. The Petition contends that Murphy describes "***auditing queries in a history table*** to check the date/time and the number of individuals returned by the queries ***to disallow repeated use of the same individual in a query***." Pet., 27. This is entirely false. To "disallow repeated use of the same individual" using the history table, Murphy's system would need to track identities of the patients/individuals in the history table - yet Murphy never describes such features. EX1009, 554. Murphy's discussion of the history table is

EX2001, ¶128

[POR, 35-40; POSR, 17-19]

Evenhaim Fails to Disclose "Unique Tokens Associated with Healthcare Data are Absent Identities of Individuals" ('012 Element [1e])

Evenhaim:

an "outsider". Intrinsically, in the context of the instant invention, the parameters of the list may include identity disclosing specifics—which probably do not pass ordinary criteria of even nominal privacy thresholds, and/or broader variables—which probably do pass ordinary criteria of normal, rigid, or strict privacy thresholds.

These identity-disclosing specifics may be such things as name, personal identity number or simple data combinations that would allow breach of privacy if applies to disclose identity. What is important in the further application of the instant method is that these identity-disclosing specifics are part of a large enough list in the processing of a query so that the query result will present sufficient statistical distribution to protect the "reverse engineering" of the result back to any individual in the initial query list. For example, the instant method can ask about the status of Tom, Dick, Harry, and a further collection of individuals—and obtain results from that query—so long as the statistical character of those results will not allow correlation of a result specific back to an individual. Thus it may be that the query includes a list of 10,000 personal IDs and the results show that these individuals belong to group A with 60% probability, Group B with 25% probability, and one of Groups C-E with the remaining 15% probability. Accordingly, transmitting the parameter list may include transmitting a sufficiently large list of identity disclosing specifics.

Simply stated, the aggregator may accept a query that includes lots of identity specifics (e.g. a list of names or a combination of a few fields that together can allow identification)—the type of query that one does not expect to be answerable in any privacy preserving fashion. This sub-step

EX1005, 8:4-34

['012 POR, 14-15, 28; '012 POSR, 24-25]

Ground 3: The Dick-Landi-Murphy Combination is Deficient

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

Overview of Dick

- Searches archived medical information repositories
- Discloses de-identification techniques

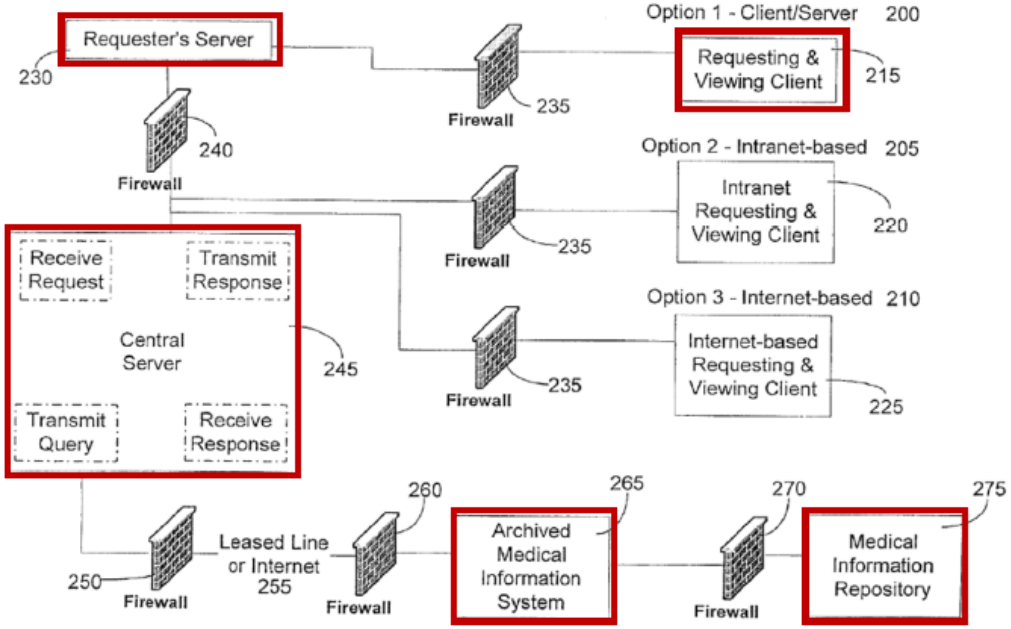


FIG. 3

EX1007 (Dick), FIG. 1

Dick performs searches with queries containing personally identifying information (PII)

- Requests include individual identification information or identification information for a group

[0024] Each request for medical information includes identification of a subject and, where required by law, a signed information release form. A subject might consist of a human individual or group of humans. The subject is the target of the search for medical information. The identification of the subject could be by way of name, patient number, social security number, driver's license number, address, phone number, biometric identification or any other identification or combination of identification characteristics capable of being correlated with stored medical information, if it exists.

EX1007(Dick), [0024]

- Searching performed using Master Person Index

[0029] A Universal Master Person Index (UMPI) may exist at the central servers **245** and will consist of a set of demographics and other specific information for identifying an individual, including pointers to other remote systems which may have relevant information about that individual. Similarly, an L-MPI (local MPI) may reside on the AMIS **265** and this will contain demographic and other specific information for the identification of the individual whose medication information resides on that particular AMIS **265**. The purpose in both instances of the Master Person Index (MPI) is to enable the identification of an individual for whom a request is being made and to increase the speed and efficiency of retrieval.

EX1007(Dick), [0029]

[EX2001, ¶¶78-84; POR, 46]

Dick's Requester's Server

- The RVC 215 can "receive a request from an authorized user responding to prompts from software executing on requester's server 230." (EX1007, [0022]).

[**0022**] Each RVC operates to receive a request for medical information according to its configuration. Generally, each RVC will receive such a request via an authorized user's responses to prompts generated by executing software displayed on the RVC video display. More specifically, a client-server RVC **215** would receive a request from an authorized user responding to prompts from software executing on requestor's server **230** or on RVC **215**. An intranet-based RVC **220** would receive a request from an authorized user responding to prompts from software executed by RVC **220**. An internet-based RVC **225** would receive a request from an authorized user responding to prompts from internet browser software executed by RVC **225** wherein the browser software is executing instructions received by an internet website accessed through the browser software.

EX1007, [0022]

Overview of Landi

- Landi perceived a problem with prior de-identification approaches in that they lacked any "mechanism by which patient identification can be recovered, if necessary."
- Landi proposed techniques for encrypting PHI such that patients could be re-identified by decrypting the encrypted ID.
- Landi describes "conventional methods for de-identifying" to include "simply stripping" PII.

TECHNICAL FIELD OF THE INVENTION

The present invention relates, in general, to systems and methods for protecting patient privacy when health care information is shared between various healthcare entities and, in particular, to systems and methods that implement secured key encryption for de-identifying patient data to ensure patient privacy, while allowing only the owners of the patient data records and/or other legally empowered entities to re-identify subject patients of de-identified data records, when needed.

EX1008, 1:13-22

Conventional methods for de-identifying patient data include simply stripping all information from the patient data records that can be used to determine the identity of a patient, or replacing such patient identifying information with something else (e.g. replace the actual name with the string "name"). With such methods, although the patient data records are de-identified, there is no mechanism by which patient identification can be recovered, if necessary.

EX1008, 2:19-26

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

[POR, 47; EX2001, ¶¶85-90]

Overview of Landi

FIG. **1** illustrates a high-level schematic diagram of a system (**10**) according to one exemplary embodiment of the invention wherein a secured encryption scheme is implemented for protecting patient privacy when patient data is shared between different entities. In general, the exemplary system (**10**) comprises a plurality of patient data processing systems (**20-1~20-**n), a central patient data processing system (**30**), a trusted broker system (**40**), and a central surveillance system (**50**). The patient data processing systems (**20-1~20-**n) are operated at different clinical sites by different healthcare organizations (including, but not limited to, doctors, health care providers, institutions, associations, organizations, hospitals, etc.). In the exemplary embodiment, each clinical cite releases de-identified patient data records to a third-party entity that operates the central patient data processing system (**30**), wherein the collected de-identified data records can be processed for purposes of, e.g., research, health care monitoring, etc. In addition, the central surveillance system (**50**) may be operated and controlled by a governmental agency (such as the CDC), or any other entity that is authorized by laws or regulations for example, to re-identify subject patients that are associated with de-identified patient data records maintained and processed by the central patient data processing system (**30**).

EX1008, 4:11-34

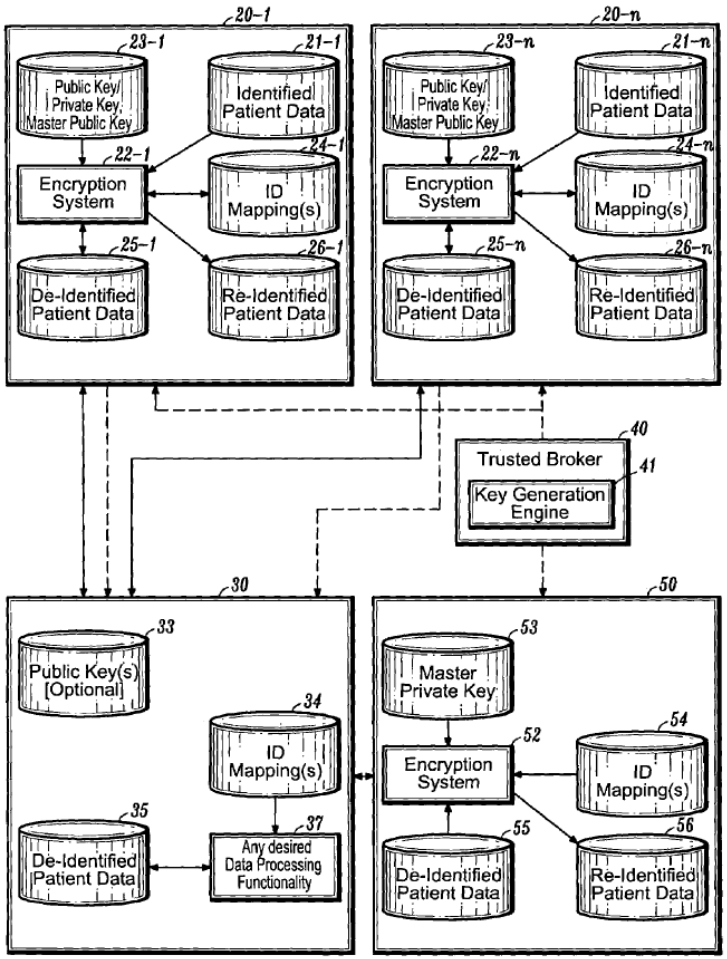


FIG. 1

EX1008, FIG. 1

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

[POR, 47; EX2001, ¶¶85-90]

Overview of Dick-Landi Combination

Key modifications proposed to Dick's system in view of Landi:

1. Consolidate de-identified medical and/or healthcare records in a centralized repository; and
2. Configure Dick's requester's server to produce encrypted IDs (i.e., the alleged "tokens") for each patient identified in a request for medical and/or healthcare information, and use the encrypted IDs to search Landi's centralized repository of de-identified records.

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

[Pet., 57-62]

Datavant Fails to Show a POSA Would be Motivated to Combine Dick and Landi

Dick-Landi Combination is the Product of Impermissible Hindsight:

1. Dick already teaches de-identification techniques, leaving no motivation to combine with Landi.
2. Dick's system does not violate HIPAA, leaving no motivation to produce tokens at Dick's requester's server for the stated purpose of complying with HIPAA.
3. A POSA would be deterred from making the combination because no known solution existed for matching a **specific** group of individuals to their **de-identified** medical and/or healthcare information while maintaining patient privacy in compliance with HIPAA.
4. Unlike Vigilytics' solution, the combination breaches patients' privacy and allows their identities to be discovered without authorization by combining patients' personally identifying information (PII) and de-identified medical and/or healthcare records at the requester's server.
5. The combination is illegal under HIPAA because it requires covered entities to disclose reversibly encrypted identification codes (Landi's encrypted IDs). A POSA would not be motivated to implement an illegal system.

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

[POR, 46-70]

Dick already teaches de-identification techniques

Petition:

> and processes" but Dick does not describe any specific de-identification methods. Ex. 1007, [0013], [0046]; Ex. 1003, ¶55, 196-199. Therefore, a POSA relying on Dick would naturally look elsewhere in order to implement the de-identification called for in Dick. Ex. 1003, ¶200.

Pet., 58

- Dr. Clark conceded that Dick's teachings of de-identification methodologies was sufficient for a POSITA
- No reason to look to Landi for de-identification methodologies

> 11 Q. Would a person of skill know how to implement
> 12 the de-identifying methodologies?
> 13 A. Should.
>
> 21 Q. Would a person of skill see a need to look
> 22 beyond the Dick reference itself to implement
> 23 de-identifying methodologies?
> 24 A. No. I think a person of skill would have
> 25 knowledge of HIPAA and the need to de-identify data.

EX2008, 80:11–25

No HIPAA Violation in Dick That Would Inspire Need for Tokens

- Datavant alleges potential HIPAA violation in sending names (PII) of individuals in request to data source as a motivation to produce tokens from the patient names and combine Dick and Landi.
- But no HIPAA violation exists, so no motivation to combine.

Petition:

A POSA would have understood that creating encrypted tokens in Dick's system using Landi's method complies with HIPAA. Ex. 1008, 1:58-2:26, 3:52-57, 9:20-27, 10:23-31, 11:64-12:67; Ex. 1007, [0004], [0024], [0026]; Ex. 1003, ¶204. For example, it would potentially violate HIPAA to just send PII to be used for such a query directly to the centralized data server source (45 C.F.R. § 164.502(d)), so in order to avoid a potential HIPAA violation it was common practice to use a public encryption key to create tokens on a requesting server when sending a query. Ex. 1003, ¶213.

Pet., 65

Dr. Rubin:

used a public encryption key to create tokens on a requesting server."). But that need is entirely illusory, conspicuously devoid of evidentiary support. Petitioner's own expert verified that there is no HIPAA violation in sending only PII in a query to the requester's server:

16 Q. Sure. So I'll rephrase my question.
17 Are you -- are you aware of any provisions of
18 HIPAA that would be violated by a noncovered entity
19 transmitting patient names to covered entities?
20 A. I mean, I'd have to look at the specific
21 example. But if you just sent the name and didn't have
22 the healthcare data with it in any way, then it wouldn't
23 violate. But if you link that with any medical
24 procedure or any healthcare information, then you would
25 have a HIPAA violation.

EX2008, 57:16–25

154. There is NO rule in HIPAA that prevents the transmission of PII in a query when the query includes no PHI. EX2006, 164.514(d). This so-called

EX2001, ¶¶153-154

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

[POR, 56-57]

No HIPAA Violation in Dick That Would Inspire Need for Tokens

- Dr. Clark testified there is no HIPAA violation when sending names

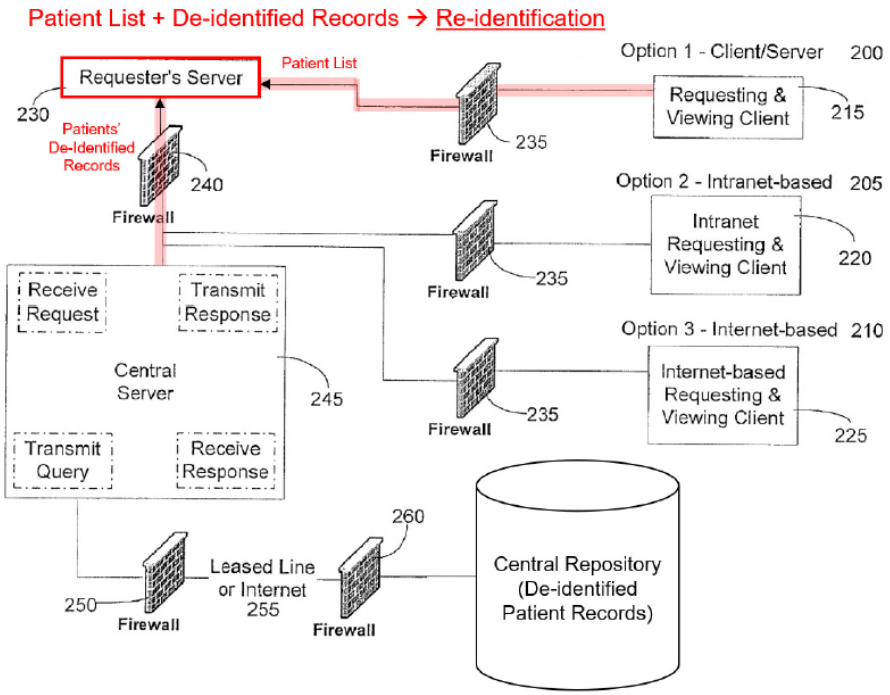
Dr. Clark:

16 Q. Sure. So I'll rephrase my question.
17 Are you -- are you aware of any provisions of
18 HIPAA that would be violated by a noncovered entity
19 transmitting patient names to covered entities?
20 A. I mean, I'd have to look at the specific
21 example. But if you just sent the name and didn't have
22 the healthcare data with it in any way, then it wouldn't
23 violate. But if you link that with any medical
24 procedure or any healthcare information, then you would
25 have a HIPAA violation.

EX2002, 57:15--25

The Combination Presents a Substantial Re-Identification Risk That a POSA Would Not Pursue

Dick-Landi-Murphy Combination:



Dick-Landi-Murphy Illustration

[POSR, 28]

Vigilytics Patents:

By using the tokens **202** and request tokens **204**, no parties other than the authorized source sites **138** can associate de-identified data **140** with the identity of any person in the group of patients **134** or discharges (e.g., the group **154**). The hospital **150** requesting the report **212** can never receive data associated with individuals. Further, the information in the report **212** may be processed into metrics that characterize a large group and cannot be used to infer information about individuals. De-identified data records may be provided to the data processor **210** (e.g., a third party data processor), but that third party may not have access to any personally identifiable information **118** about the group **154**. Nor may any party with access to the de-identified data **140**, other than the healthcare providers **138**, also have access to the de-identifier **142**. These features of the system **200** maintain the privacy of the protected health information **130**.

EX1001, 5:26-42

[POR, 58-59; POSR, 9, 28]

The Combination Presents a Substantial Re-Identification Risk That a POSA Would Not Pursue

Dr. Rubin:

159. Third, the proposed combination would also result in a violation of patients' privacy rights under HIPAA or other applicable regulations due to the fact that the requester's server would hold all three pieces of information at the same time: the names of the patients received from the request, the corresponding encrypted IDs, and the patients' de-identified healthcare information received in the response. EX2008, 76:6–78:19. The MIR/AMIS being covered entities would not be permitted to send de-identified medical information to the requester's server because the requester's server is not authorized to have such information. EX2008, 74:9–17. Dick never proposed that the requester's server would maintain patients' PHI, and the Petition has certainly not established that the requester's server would be authorized to do so. EX2008, 71:20–72:6. Accordingly, it would be impermissible for the MIR/AMIS to release all of this information to the requester's server that could be used to link patient names with their corresponding healthcare records. This clear privacy and HIPAA violation means that a POSITA would not consider this option at all.

EX2001, ¶159

[POR, 58-59; POSR, 9, 28]

The Combination Presents a Substantial Re-Identification Risk That a POSA Would Not Pursue

- Dr. Clark conceded that it "would not be prudent" to combine names with health conditions and that he "wouldn't recommend anyone do that."

Dr. Clark:

Q. If the third-party possessed both the de-identified data and the names of the individuals in the group, how does re-identification occur?

A. If I have your name and I have a health condition, it's possible that I could, depending upon whether they are using pseudonyms or not or whether the name was Joe Smith, that I may or may not be able to re-identify you, but I would not implement a system in which the patient's names were available. That's for sure.

Q. And why would you not do that?

A. Because I would not want someone with a unique name to have their name out there with a health condition that then would allow me to infer the protected health information. That would not be prudent.

Q. That would be a re-identification risk?

A. I think it could be. I mean, I -- I wouldn't recommend anyone to do that.

EX2009, 30:17-31:7

The Combination Violates HIPAA by Requiring Disclosure of Reversibly Encrypted Identification Codes (Landi's Encrypted IDs)

•Landi's encrypted IDs are formed by **reversibly encrypting patient identifying information (PII)** (e.g., names, DOB, social security number).

Landi:

public keys and/or a master public key (**23-1**~**23**-n). More specifically, each healthcare entity can use its own public key or both its own public key and the master key to encrypt one or more items of patient identifying information contained in the respective patient medical records owned/managed by the healthcare entity), to thereby generate an Encrypted ID for each subject patient. An Encrypted ID of a given patient is a unique reproducible encrypted version of patient identifying information that is sufficient to uniquely identify the patient. The patient identifying information, which is encrypted to generate the Encrypted ID, may include one or more items of patient identifying information, including, but not limited to, patient name, social security number, and/or address.

EX1008, 5:7-19

[POSR, 12-16; POR, 51-52]

The Combination Violates HIPAA by Requiring Disclosure of Reversibly Encrypted Identification Codes (Settimi's EPIDs)

HIPAA Privacy Rule (45 CFR § 164.514(c))

(c) ***Implementation specifications: Re-identification.*** A covered entity may assign a code or other means of record identification to allow information de-identified under this section to be re-identified by the covered entity, provided that:

(1) ***Derivation.*** The code or other means of record identification is not derived from or related to information about the individual and is not otherwise capable of being translated so as to identify the individual; and

(2) ***Security.*** The covered entity does not use or disclose the code or other means of record identification for any other purpose, and does not disclose the mechanism for re-identification.

EX2006, 2

Dr. Rubin:

individual identifying information." EX1008, 2:46–48; *id.*, 5:13–19. By generating and releasing reversibly Encrypted IDs based on individual identifying information of patients, Landi violates HIPAA's proscription against the use of codes for re-identification that are "derived from or related to information about the individual."[4] 45 C.F.R. §164.514I(1). A POSITA therefore would not have turned to Landi for its specific de-identification methodology that involves creating Encrypted IDs because Landi's approach would not have achieved the Petition's stated goal of achieving HIPAA compliance. Landi instead invites a HIPAA violation, which a POSITA would have sought to avoid.

EX2001, ¶144

Dr. Clark:

Q. So Landi's encrypted ID does not comply with the derivation requirement. Correct?

A. I'm not sure that's true because you have to have access to the key in order to do it; but generally speaking, in that embodiment, you could argue that it's not combined with a derivation requirement.

EX2009, 54:23-55:1-3

[POR, 51-52; POSR, 25-27]

DLM Combination Fails to Provide Multiple Claim Limitations

- Non-exhaustive list of claim limitations not disclosed:
  - "A Privacy-Preserving Rule that Disallows Same Individuals to be Included within a Defined Period of Time" ('651 Element [1e])
  - "Producing[] … a Request Token for Each Individual Included in the Group of Individuals" and "Comparing the Request Tokens to Tokens Associated with the De-Identified Data to Find Matching Tokens" ('651 Elements [1c]-[1d])
  - "The Unique Tokens Associated with Healthcare Data are Absent Identities of Individuals ('012 Element [1e]) and "An Identity of the Individuals Being Unattainable From the Produced Information" ('651 Element [1g])
  - "Determine Whether the Group Includes at Least a Minimum Number of Individuals to be Included in a Report in Response to the Request" ('651 Element [1e])
  - "The Provided Information is Medical Data in Conformity with [HIPAA]" ('651 Claim 2)

*Other independent claims of the challenged patents likewise recite various claim limitations similar to those listed above. DLM combination fails to provide these limitations for similar reasons.

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

Overview of Murphy

Our solution to thwarting an attack on this kind of system was to take advantage of the following set of properties of the query tool, and inventing a solution that fit around these properties. Although the fact that these properties must exist in the query tool limits the usefulness of the solution, these are properties many query tools would have (on the web for example) where aggregate numbers of patients need to be returned. The required properties are:

1) It is sufficient that the query tool only returns numbers of distinct patients.

2) The number of distinct patients is useful even when the number is within a probable range of 2 or 3 patients. This range can be varied to obtain optimized signal to noise ratios for detecting an attack.

EX1009, 553

In an abstract sense, the key to the data obfuscation method is to make attempts to discern individuals in the database a recognizable pattern by a computer algorithm. The method works by having the server not pass true numbers of patients to the query tool, but rather numbers that would be (if the same query were performed repeatedly) distributed around the true number of patients. For example, if one queried to find out how many people carried the diagnosis of manic-depression at Partners Healthcare Inc., one might get the result of 3,487, 3,488, 3,488, 3,490, or 3,489 patients. The exact nature of this range is the subject of much discussion below, but the range would be quasi-centered on the true number of 3,489. In order to find the true number, one would need to run the same query repeatedly. It is here where a computer algorithm would recognize this attack and mark the user whose queries returned the same true (not reported) number repeatedly as a suspicious pattern.

EX1009, 553-54

Datavant Errs in Relying on Murphy as "Block[ing]" Queries to Meet "Minimum Number" Claim Limitations

Example Limitations:

- '651 Element [1.e] - "determine whether the group includes at least a minimum number of individuals to be included in a report in response to the request"
- '375 Element [1.b] - "an encryption server that is configured to determine if the number of individuals in the request is at least a defined minimum number of individuals"
- '375 Elements [1.g]-[1.h] - "producing the healthcare information ... based on a minimum number of matched tokens"
- '012 Element [1.f] - "the healthcare information being produced based on a minimum number of matched tokens"

'651 Petition:

Murphy makes clear that, in complying with HIPAA, an algorithm preferably blocks any queries that return fewer than 3 patients to prevent the user from associating a patient with his/her record. Ex. 1009, 555 ("One vulnerability is that we allow a large number of queries to be performed that return zero patients.… Therefore, we modify our range algorithm to have a wider obfuscation at values close to zero, and indicate the greater range by returning '<3' when the obfuscated number is less than 3."); Ex. 1003, ¶89.

Pet., 27

Murphy Does Not "Block Queries"

Dr. Rubin:

175. First, the Petition's assertion that Murphy's algorithm "blocks queries" when the number of patients is less than 3 is incorrect. EX2008, 68:10-69:2. For results around zero, Murphy teaches modifying the "range algorithm to have a wider obfuscation at values close to zero, and [indicating] the greater range by returning '<3' when the obfuscated number is less [than] 3." EX1009, 555. Widening the obfuscation range in these cases is far different from blocking a query as alleged. To the contrary, the query tool always returns a numerical result to the query by applying an obfuscation algorithm. The obfuscation algorithm for results less than three involves providing an indication that the obfuscated result is less than three, rather than returning an absolute value or no value at all. The query is still answered, however, as the result informs the requester that the number of matching patients is near zero, and even this result bears a relationship to the true number of patients identified in response to the request. Importantly, Murphy's query tool does not return actual patient records but instead returns numerical values. *Id.* So, any numerical result, including '<3,' constitutes a response to the query. *Id.* Accordingly, Murphy does not teach "blocking queries," as argued by the Petition.

EX2001, ¶175

Murphy:

Although this method may seem overly complex in some ways, the task of obfuscating data in a foolproof manner is extremely difficult. The method we have described also has some vulnerabilities that need to be explored. One vulnerability is that we allow a larger number of queries to be performed that return zero patients. Unfortunately, naï ve users who are performing incorrect queries will often have zero patients returned. Having the table-watching agent terminate access to these users would be too sensitive a response. However, attacks around zero are more likely to succeed. Therefore we modify our range algorithm to have a wider obfuscation at values close to zero, and indicate the greater range by returning "<3" when the obfuscated number is less that 3.

EX1009, 555

Murphy Criticizes De-identification and Re-identification

An approach to preventing the identification of individuals in the database could be to de-identify patients in the database with a system that complies with HIPAA's 18 points for data de-identification (2). Although this would be a good start, there are two sets of problems with this. The first set of problems results from the destruction of important clinical data. This results mostly from the destruction of location information (zip codes) and the destruction of date information. Just attempting to recruit patients in the general area of the hospital for research studies relies on zip codes, not to mention more complex epidemiological studies. The destruction of date information is equally problematic. The HIPAA regulations state that dates should be rounded to the month if not to the year. However, very often a query will want to identify which event occurred first as a means of finding cause and effect. For example, in a study on drug adverse events, one may want to find all cases where a rise in liver function tests followed the initiation of a drug by 1-3 days. The second set of problems is that even with the destruction of all the above information, there is still no guarantee that someone cannot be identified in the database. As greater quantities of data are added, the chance that a set of clinical values (such as an SGOT of 367 and an SGPT of 1325) will not be able to uniquely define a patient becomes very low.

EX1009, 553

Another approach might be to use table ambiguation by suppression of selected cells (3). However, this approach would also result in the destruction of large amounts of clinical data. Because cells that uniquely distinguish a patient from other patients are the cells to be suppressed, the most interesting clinical data would be specifically targeted for destruction. This would severely limit the usefulness of the data for many purposes, such as has been extensively analyzed by Ohno-Machado et. al. (4). Furthermore, this approach is also susceptible to re-identification of patients that have supposedly been de-identified (5).

EX1009, 553

- Murphy rejects options of de-identifying
- Murphy rejections techniques susceptible to re-identification

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

[POR, 11-14, 35-40, 68-70]

Thank You

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE