

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

Case No. IPR2024-00226
U.S. Patent No. 9,118,641

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



DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

1

Exhibit 2009
Datavant v. Vigilytics
IPR2024-00226

Table of Contents

Overview of the '641 Patent and Challenged Claims	3
Ground 1: Evenhaim.....	9
Ground 2: Evenhaim-Settimi.....	25
Ground 3: Dick-Landi.....	40
Ground 4: Dick-Landi-Murphy.....	63

Overview of the '641 Patent and the Challenged Claims

Overview of the '641 Patent

58. To address this conundrum, the '641 Patent disclose novel methods and systems to implement specific token generation and anonymization approaches to achieve a new computer capability: systems able, in response to a request for information regarding a group of individuals who have not authorized PHI disclosure, to provide to third parties (e.g., insurance companies or professional employer organizations, known as "PEOs") irreversibly de-identified healthcare information from which the third party may calculate a risk score used in setting an insurance premium for the group, while complying with federal privacy regulations, including HIPAA.¹ See, e.g., *id.* at 4:25–30. This was a significant advancement because prior to the invention, signed authorizations had to be collected from individuals in the request, which was an impediment to a viable product. Although tokenized health data had been used for other purposes prior to

EX2001 (Dr. Rubin Dec), ¶158.

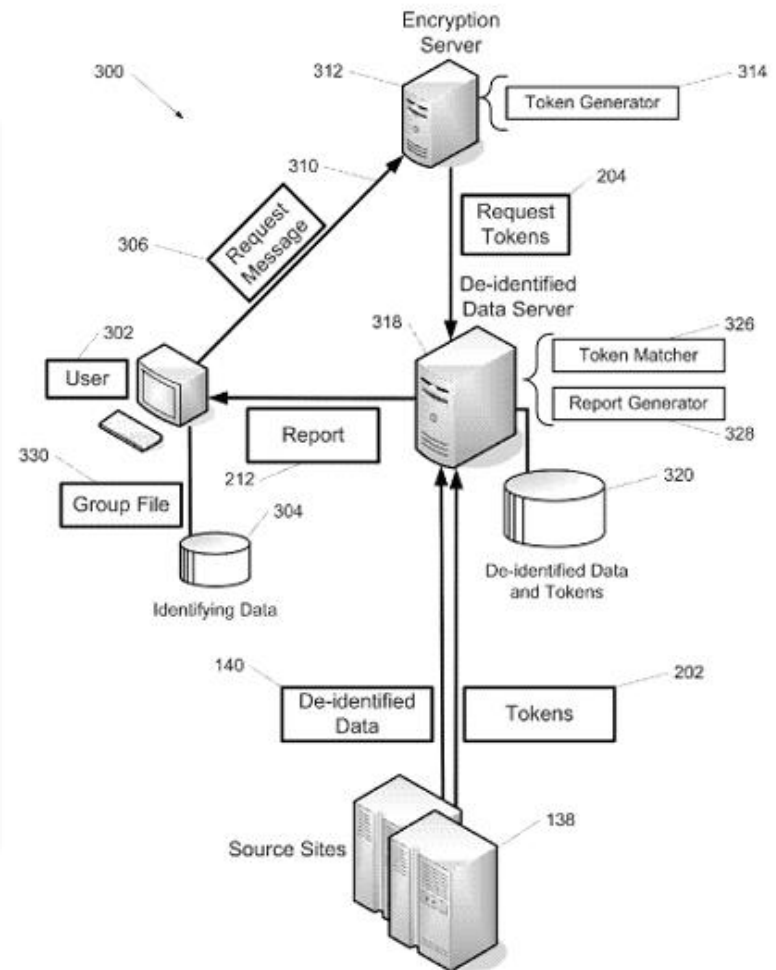


FIG. 3

EX1001, FIG. 3 (cited in 00226-12 (POR), 6)

Challenged Claims of the '641 Patent

ii. The Grounds on Which the Challenge is Based

Ground	Claims	Basis for Unpatentability
1	1-19	Obvious over U.S. Pat. No. 7,823,207 ("Evenhaim") (Ex. 1005)
2	1-19	Obvious over Evenhaim in view of U.S. Pat. Appl. Publ. No. US 2007/0294111 ("Settimi") (Ex. 1006)
3	7-10 and 13	Obvious over U.S. Pat. Appl. Publ. No. US 2002/0111833 ("Dick") (Ex. 1007) in view of U.S. Pat. No. 7,519,591 ("Landi") (Ex. 1008)
4	1-6, 11, 12, and 14-19	Obvious over Dick in view of Landi and "A Security Architecture for Query Tools used to Access Large Biomedical Databases" by Murphy and Chueh ("Murphy") (Ex. 1009)

Pet., 9 (cited in 00226-12 (POR), 6)

Independent Claim 1

1. A computer-implemented method comprising:
receiving a request for healthcare information that characterizes a group composed of multiple individuals, the healthcare information being assembled from healthcare data that comprises de-identified information;
producing, by a processor of an encryption server, tokens, wherein the identity of the individuals in the group is unattainable from the tokens, each token being associated with an individual in the group, each token being associated with a corresponding token in a set of de-identified healthcare data, and each token in the set of de-identified healthcare data being associated with healthcare data for a corresponding individual;

when the group includes at least a minimum number of individuals, producing the requested healthcare information that characterizes the group by identifying the tokens in the set of de-identified healthcare data and the associated de-identified healthcare data that correspond to the produced tokens, wherein the produced tokens and the tokens in the set of de-identified data are similarly encrypted, and the produced healthcare information is absent personally identifiable information and has been produced without the individuals in the group having authorized production of such healthcare information that characterizes the group; and
providing the produced healthcare information to report the characteristics of the group composed of multiple individuals absent authorization at any time from the individuals in the group of individuals to provide the requested healthcare information.

EX1001, 10:25-54.

Independent Claim 7

7. A system comprising:
an encryption server, having a processor and a memory, for producing a token for each individual included in a group composed of multiple individuals identified in a request for healthcare information that characterizes the group, wherein the identity of the individuals in the group is unattainable from the tokens, each token being associated with a corresponding token in a set of de-identified healthcare data, and each token in the set of de-identified healthcare data being associated with healthcare data for a corresponding individual;
a data server for producing the requested healthcare information that characterizes the group by identifying the tokens in the set of de-identified healthcare data and the associated de-identified healthcare data that correspond to the produced tokens, wherein the produced tokens and the tokens in the set of de-identified data are similarly encrypted, and the produced healthcare information is absent personally identifiable information and has been produced without the individuals in the group having authorized production of such healthcare information that characterizes the group; and
providing the produced healthcare information to report the characteristics of the group composed of multiple individuals absent authorization at any time from the individuals in the group of individuals to provide the requested healthcare information.

EX1001, 11:8-35.

Independent Claim 14

14. One or more computer readable storage devices storing instructions that are executable by a processing device, and upon such execution cause the processing device to perform operations comprising:

- receiving a request for healthcare information that characterizes a group composed of multiple individuals, the healthcare information being assembled from healthcare data that comprises de-identified information;
- producing, by a processor of an encryption server, tokens, wherein the identity of the individuals in the group is unattainable from the tokens, each token being associated with an individual in the group, each token being associated with a corresponding token in a set of de-identified healthcare data, and each token in the set of de-identified healthcare data being associated with healthcare data for a corresponding individual;

- when the group includes at least a minimum number of individuals, producing the requested healthcare information that characterizes the group by identifying the tokens in the set of de-identified healthcare data and the associated de-identified healthcare data that correspond to the produced tokens, wherein the produced tokens and the tokens in the set of de-identified data are similarly encrypted, and the produced healthcare information is absent personally identifiable information and has been produced without the individuals in the group having authorized production of such healthcare information that characterizes the group; and
- providing the produced healthcare information to report the characteristics of the group composed of multiple individuals absent authorization at any time from the individuals in the group of individuals to provide the requested healthcare information.

EX1001, 12:1-32.

Ground 1: Evenhaim

Overview of Evenhaim: Analyze “Original, Most Complete Data Set”

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 (cited by Sur-Reply, 6).

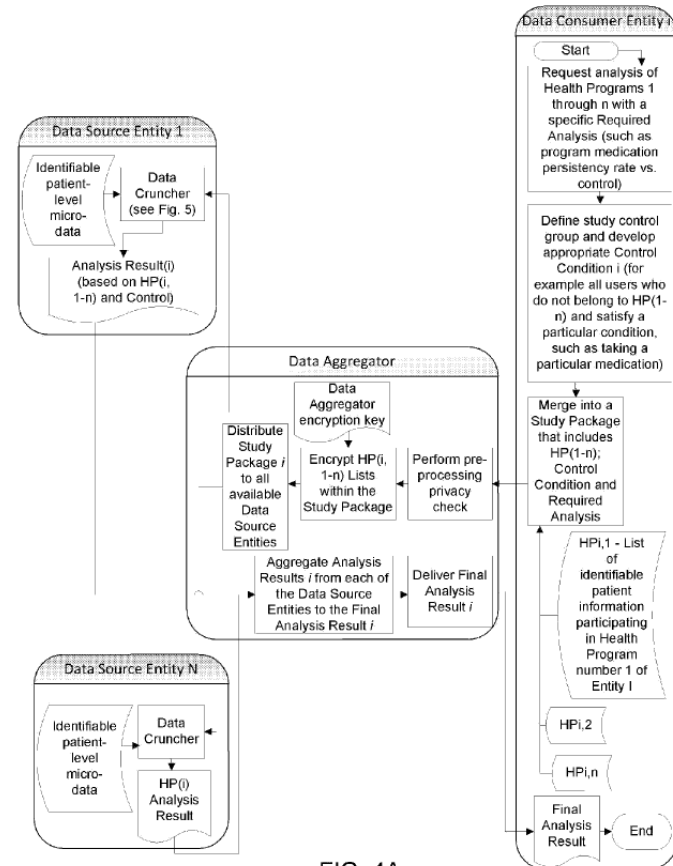


FIG. 4A

EX1005 (Evenhaim), FIG. 4A (cited by Sur-Reply, 5).

Overview of Evenhaim: Analyze “Original, Most Complete Data Set”

ADVANTAGES, OBJECTS AND BENEFITS OF THE INVENTION

Ergonomic Issues: Preferred embodiments of the instant invention allow 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 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 (Evenhaim), 6:12-32 (cited by POR, 26, 39; Sur-Reply, pp.4-6, 16).

Overview of Evenhaim: Analyze “Original, Most Complete Data Set”

Evenhaim discloses a system that allows for searching of PHI at source entities such as pharmacies or healthcare providers. EX1005, 1:14–19, Abstract.

Unlike the '641 Patent, Evenhaim relies on the source entities to search PHI directly and to de-identify patient-level data only after they are searched.

EX1005, 6:15–25, 17:43–52; EX2001, 76–79.

POR, pp.12-13.

entity processors. EX1005, 6:56–67, 7:1–27. The source-entity processors search their original micro-level data—including patients' PHI—to identify records responsive to the search request. EX1005, 6:15–25, 17:43–52. Once patient level data has been identified responsive to the requests, the source-entity processors de-identify those records and return them in “files” to the aggregator processor. EX1005, 7:20–24. The aggregator processor aggregates files received from each of multiple source-entity processors and reports an agglomerated extract of the files to the consumer-entity in response to the query. EX1005, 9:13–27. The aggregator collects and assembles identity protected reports from multiple source entities such that the statistical significance of conclusions are enhanced. *Id.*, 10:7-11.

EX2001 (Dr. Rubin Dec), ¶77 (cited by POR, 12-13).

Overview of Evenhaim: Analyze “Original, Most Complete Data Set”

119. Evenhaim’s data aggregator also would not convey request tokens to the data source entities because the data source entities are configured to search original micro-level patient data from which patient names or IDs in the request can be matched against patient names or IDs in the data source entities’ original data records. EX1005, 6:15–34, 9:47–51, 17:41–52. There is no token matching at the data source entities in Evenhaim not only because it is unnecessary, but also because it would be contrary to Evenhaim’s specific goal of searching original micro-level patient data *before it is de-identified or aggregated*. *Id.*; see also EX1005, 17:47–48 (“This analysis is performed on the original, most complete

EX2001 (Dr. Rubin Dec), ¶119 (cited by POR, 23-27).

8	Q. The queries are sent to the data source
9	entities, correct?
10	A. Yes.
11	Q. And the data source entities are covered
12	entities under HIPAA, correct?
13	A. Generally speaking, yes.
14	Q. And so the data source entities would,
15	themselves, maintain PII for patients, correct?
16	A. Well, that's sort of the point of the
17	reference, right?

EX2002 (Dr. Clark Depo), 51:8-17 (cited by POR, 27).

Evenhaim Teaches Away From Searching De-Identified Data

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.

Sur-Reply, 6 (citing EX1005, 17:28-52).

As described above in this section, HIPAA provides substantial limitation on Personal Health Information and existing de-identification method may render the information useless
45 for the purpose of measuring the impact of health programs.

EX1005, 4:42-45 (cited by Sur-Reply, 11).

agent.” EX1005, 12:53–66. Evenhaim thus teaches away from the exact approach that Petitioner and Dr. Clark improperly seek to inject in Evenhaim’s system to force a read on the challenged claims of the ’641 Patent.

EX2001 (Dr. Rubin Dec), ¶119 (cited by POR, 23-27).

no longer in the lower resolution dataset. Therefore, the exist-
65 ing healthcare analytics model does not provide the capability to aggregate the behavior of a group of patients by a pre-defined list of consumers participating in a health program. This is only one example of the limitations of the existing healthcare analytics model that is addressed by the instant protocol.

EX1005, 16:65-17:3 (cited by Sur-Reply, 4, 6).

Evenhaim Teaches Away From Searching De-Identified Data

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

- “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.”

Sur-Reply, 4 (citing EX1005, 17:28-52).

Datavant's Reply Quotes Evenhaim's Description of "Identity-Free Micro-Data" Out of Context

Reply:

like[]"). Thus, Evenhaim acknowledges that original, unaltered patient names are not required to achieve its goal. Also, Evenhaim acknowledges that the "micro-level records" that it seeks to maintain are not harmed by the removal of PII inasmuch as Evenhaim acknowledges the option of "identity-free micro-data." *Id.*, 10:29-32, 12:36-52, Fig. 4 ("micro data"); EX1014, 61:17-67:2 (distinguishing micro-level data from physician level data); *see* EX1006, [0038], [0043].

Reply, 8.

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 (Evenhaim), 10:29-32 (cited by Reply, 8).

Sur-Reply:

41 of Evenhaim on several occasions, but this portion of Evenhaim refers to the operation of "agglomerating *the extract*," i.e., an operation performed by the data aggregator on a search extract received from the data source entities after a search on the original patient data has generated data subject to the extract.

Sur-Reply, 7.

Evenhaim's Encryption of Patient Lists Does Not Produce Tokens for Individuals for Use in Searching De-Identified Data

106. Encrypting entire patient lists (and query), 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. See, e.g., 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”).

EX2001 (Dr. Rubin Dec), ¶106 (cited by POR, 23-29).

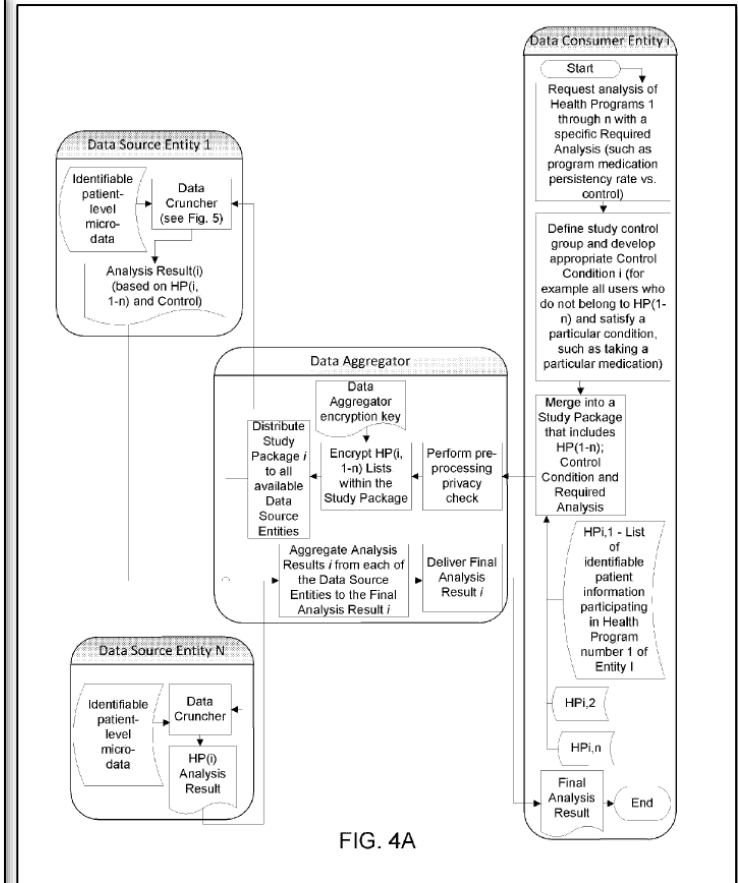


FIG. 4A

EX1005, FIG. 4A (cited by POR, 13-14).

Evenhaim's "Interim Data Merger Embodiments" Do Not Involve Producing Tokens for Individuals for Use in Searching De-Identified Data

Evenhaim:

25 Returning now to FIGS. 3 and 8 (and Mutatis Mutandis to
their respective program storage devices), it can now be
appreciated that there are numerous locations where an
encryption of individual identification and/or merger of data
for individuals could be facilitated. By way of non-limiting
30 example, on the "aggregator" processor side, the encryption
and/or merger could be enable during the sub-steps: receiving

EX1005, 12:24-30 (cited by POR, 21-29).

110. To be clear, Evenhaim never once discloses or suggests that its
"interim data merger embodiments" would involve the data aggregator producing
tokens for individual patients that would be shared with the data source entities as
alleged. Dr. Clark repeatedly cites Evenhaim's statement that "the encryption of
individual identification and/or merger of data for individuals could be facilitated
... on the 'aggregator' processor side" (see EX1005, 12:24-29), but this one line is
far too vague to support anything like the detailed scheme that Petitioner and Dr.
Clark attempt to inject into Evenhaim as part of their mapping to the '641 claims.

EX2001 (Dr. Rubin Dec), ¶110 (cited by POR, 21-29).

32. The steps of "receiving a respective file, or aggregating the plurality of files"
notably occur upon the aggregator's receipt of *response(s)* to the queries and are
not at all related to encrypting the query itself. See *id.*, 7:3-6, FIG. 3. It would
make no sense for the data aggregator to produce a token for each patient in the
patient list that it sends with the query to the data source entities, especially since
the patient list and query themselves contain no PHI that would require de-
identification under HIPAA. The data source entities also already know the
patients' identities and maintain their PHI, and there is simply no reason that the
data aggregator would need to disguise the identities of the patients in a request
that contains no PHI from data source entities that already maintain PHI for the
patients.

EX2001 (Dr. Rubin Dec), ¶110 (cited by POR, 21-29).

Evenhaim's "Interim Data Merger Embodiments" Do Not Involve Producing Tokens for Individuals for Use in Searching De-Identified Data

109. The Petition's and Dr. Clark's additional citations to Evenhaim's "interim data merger enabled embodiments" are similarly unavailing and fail to show that the data aggregator ever generates encrypted tokens for individuals identified in a request that would be sent in a query to any data source entity. EX1003, ¶39; EX1005, 12:4–13:20. As an initial matter, Evenhaim's "interim data merger enabled embodiments" involve encryption in an entirely different context from encrypting the patient list and query for data transmission security, as described elsewhere in Evenhaim. Indeed, as noted above, the patient lists encrypted at the data aggregator for data transmission security are *decrypted* by the data source entities, while the data merger enabled embodiments discussed in col. 12 of Evenhaim specifically cautions that the party receiving the encrypted data should *not* be capable of decrypting that data. Compare EX1005, 13:33–35 ("every step requiring a transmitting of data ... preferably include[s] respective encryption *and decryption*") and 15:63 ("decrypt the *lists*") with 12:16–20 ("special care must be taken that the entity performing the linkage is *not capable*

of knowing the key and decryption function, because knowing these facets would enable a breach of the individual's privacy"). It would be non-sensical for the data source not to decrypt a patient list and query that the aggregator has encrypted for data transmission security. The notion that decrypting patient lists at the source entity would implicate a breach of privacy of the type described at col. 12:16–20 is similarly non-sensical, especially as the patient list contains no PHI and Evenhaim expressly discloses embodiments where such decryption is performed without breaching patient privacy. See, e.g., EX1005, FIG. 5 ("Decrypt patient identifier list"). Petitioner's and Dr. Clark's arguments ignore these facts and repeatedly conflate encryption for data transmission security with the encryption described at column 12 of Evenhaim.

EX2001 (Dr. Rubin Dec), ¶109 (cited by POR, 21-29).

EX2001 (Dr. Rubin Dec), ¶109 (cited by POR, 21-29).

Evenhaim's "Interim Data Merger Embodiments" Do Not Involve Producing Tokens for Individuals for Use in Searching De-Identified Data

113. Producing tokens for individual patients at the data aggregator to query the data source entities is also non-sensical because it would require the data source entities to take inefficient steps of either (i) decrypting the tokens before searching for the appropriate patients' data records or (ii) pre-generating corresponding tokens for patient data records that could be compared against the tokens received from the data aggregator. Considering the significance of these steps, one would reasonably expect Evenhaim to describe them, or even mention them at all, if the interim-data merger enabled embodiments were intended to cover 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 identifier 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.*

EX2001 (Dr. Rubin Dec), ¶113 (cited by POR, 21-29).

Evenhaim Does Not Match Request Tokens for Individuals Against Similarly Encrypted Tokens Associated with De-identified Data

Dr. Rubin:

122. The Petition cites Dr. Clark's declaration testimony that "[t]here is an embodiment [in Evenhaim] that does not require tokens, but Evenhaim also describes using a token matching system to allow for querying de-identified patient records." EX1003, ¶19. But this is wrong; the latter embodiment simply does not exist.

EX2001 (Dr. Rubin Dec), ¶122 (cited by POR, 31-32).



Evenhaim:

33 Thus there are at least two different situations where data from the data source is returned already aggregated, after analysis as well as a second option where data is release at the micro-level (each person's data), but the person identity information, for the matched population, is replaced with a
40 unique encryption key, such as a one way hash or Advanced Encryption Standard (AES) or the likes. This compatible encryption for the identifying micro-level data preserves the

EX1005, 12:39-42 (cited by POR, 31).

Evenhaim Does Not Match Request Tokens for Individuals Against Similarly Encrypted Tokens Associated with De-identified Data

Dr. Rubin:

124. Thus, the first embodiment involves data being aggregated at the data source before it is returned to the data aggregator, and the second situation involves de-identifying individual patient level data that can be returned to the data aggregator and later merged or linked by the aggregator itself. Notably, neither situation involves the data source entities performing token matching on de-identified patient data as alleged. Dr. Clark's characterization of these embodiments is completely at odds with Evenhaim's actual disclosure. EX1003, ¶40.

EX2001 (Dr. Rubin Dec), ¶124 (cited by POR, 31-32).

Evenhaim Does Not Match Request Tokens for Individuals Against Similarly Encrypted Tokens Associated with De-identified Data

126. Dr. Clark also distorts Evenhaim's descriptions of data linking *at the aggregator* to wrongly contend that the data source entities would create tokens associated with patient healthcare data for a matching process. *See* EX1003, ¶40 ("this token matching system 'preserves the ability to known [sic] that two records belong to the same identity, but preserves the privacy of the identity of that person.'"). *Id.*, 12:42–45. But Dr. Clark fails to acknowledge that Evenhaim is instead referring to the *aggregator's* (not the data sources') ability to link de-identified patient records belonging to the same patient:

If the data sources are careful to encrypt the identifying portions of each record using a common encryption "key", then further down the data processing circuit it will be possible to link records related to a common individual without compromising the identity of that individual. Of course, special care must be taken that the entity performing the linkage is not capable of knowing the key and decryption function;

EX1005, 12:12–18.

EX2001 (Dr. Rubin Dec), ¶126 (cited by POR, 31-32).

127. Evenhaim's separate references to the "data sources" and the entity that is "further down the data processing circuit," make clear that the latter entity is different from the data sources. And because Evenhaim seeks to prevent the compromise of individual identities at the entity "further down the processing circuit," the phrase cannot encompass the data sources, which themselves are covered entities under HIPAA that are already aware of such identities and already maintain PHI. EX1005, 5:12–28. As such, Dr. Clark's statement that a "POSA would understand that 'further down the data processing circuit' describes the subsequent use of the de-identified records for querying," is simply incorrect. At the very least, the cited portions of Evenhaim do not show that any token matching would take place at the data sources.

EX2001 (Dr. Rubin Dec), ¶127 (cited by POR, 32-33).

Evenhaim's Data Source Entities Decrypt the Patient Lists to Recover Patient Names for Use in Name Matching Algorithms

120. Evenhaim also specifically illustrates in FIG. 5 a data cruncher module of a data source entity that decrypts a patient identifier list for the very purpose of allowing names from the lists to be matched against names in the original data records. EX1005, Fig. 5. In fact, name matching or matching based on similar individually identifying information is the only approach described in Evenhaim for processing the queries received at the data source entities. See EX1005, 9:47–51 (“If the parameter list included specific individuals by name or ID or the likes, “Crunching together” may involve *name matching through “fuzzy logic” and other name matching algorithms of the population* defined by the parameter list with the source-entity database names, in addition to the other steps defined above.”). By contrast, Evenhaim describes no embodiment where Petitioner’s proposed “token matching” occurs at the data source entities. This again is not surprising, as Evenhaim specifically sought to have the data source entities respond to queries based on analysis of *raw data* in its original, most detailed form. EX1005, 6:16–17 (“the instant invention allow[s] analysis of

EX2001 (Dr. Rubin Dec.), ¶120 (cited by POR, 28).

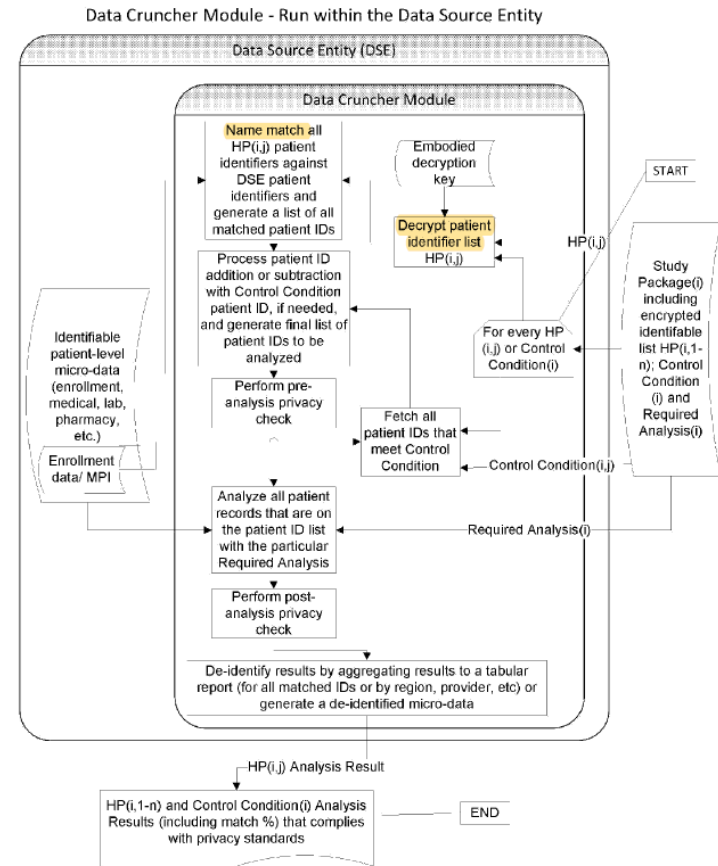


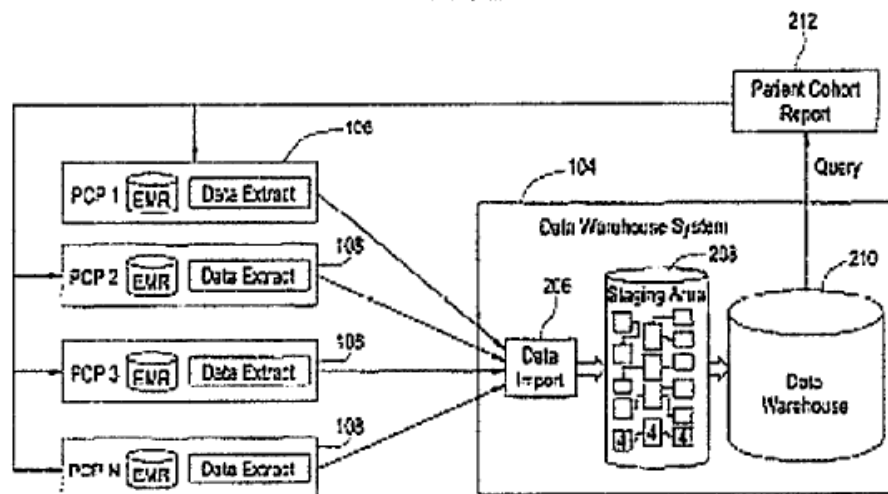
FIG. 5

EX1005, FIG. 5 (cited by POR, 18).

Ground 2: Evenhaim-Settimi

Overview of Settimi

FIG. 2



EX1006 (Settimi), FIG. 2 (cited by Sur-Reply, 14)

(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 (cited by POR, 35)

Settimi's Centralized Warehouse Stores Fully De-Identified Data under HIPAA or Limited Data Sets

[0027] Certain embodiments provide a secure process for sending de-identified patient information from an ambulatory patient care provider (PCP) site to a data warehouse system where the patient data may be analyzed and compared with a wider range of patient data. The terms “de-identified patient information” and “de-identified patient data” as used in this document refer to both fully de-identified data as defined by HIPAA and limited data set data as defined by HIPAA. A limited data set is protected health information for research, public health and health care

EX1006 (Settimi), [0027] (cited by Sur-Reply, 11)

[0040] In an alternate exemplary embodiment of the present invention, the following identifiers are removed or transformed in order to create de-identified data that would be classified under the HIPAA definition as limited data set information: direct identifiers such as name, postal address (other than city, state and zip code), social security number and medical records numbers. In the limited data set information implementation of the present invention some identifying information may remain such as dates of examination, documentation, diagnosis, prescription and lab test results.

EX1006 (Settimi), [0040] (cited by Sur-Reply, 11)

[0039] The extraction process is performed by application software located on the PCP system 108 and may be executed in the background on a periodic basis (e.g., at 2 a.m. every night, at 2 a.m. every Saturday). In this manner, the extraction process will be less likely to interfere with existing software located on the PCP system 108. The extraction process may also be initiated by a remote system (e.g., the data warehouse system 104) and may include full or incremental back-up schemes. In an exemplary embodiment of the present invention, the following identifiers are removed or transformed in order to create de-identified data that would be classified under the HIPAA definition as fully de-identified data: name, geographic subdivisions smaller than a state including street address, city, county, precinct, zip code (down to the last three digits), dates directly related to an individual (e.g., birth date), phone and fax numbers, electronic mail addresses, health plan number, account number, certificate/license number, device identifier and serial numbers, unified resource locator (URL), internet protocol (IP) address, biometric identifiers, full face photograph, and other unique identifying numbers, characteristics or codes.

EX1006 (Settimi), [0039] (cited by Sur-Reply, 11)

Settimi's Centralized Warehouse Stores Fully De-Identified Data under HIPAA or Limited Data Sets

HIPAA De-Identified Data (45 CFR § 164.514(b))

- (b) *Implementation specifications: Requirements for de-identification of protected health information.* A covered entity may determine that health information is not individually identifiable health information only if:
- (1) A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable:
 - (i) Applying such principles and methods, determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and
 - (ii) Documents the methods and results of the analysis that justify such determination; or
- EX2006 (cited by Sur-Reply, 12)

HIPAA Limited Data Sets(45 CFR § 164.514(e))

- (e)
- (1) *Standard: Limited data set.* A covered entity may use or disclose a limited data set that meets the requirements of paragraphs (e)(2) and (e)(3) of this section, if the covered entity enters into a data use agreement with the limited data set recipient, in accordance with paragraph (e)(4) of this section.
 - (2) *Implementation specification: Limited data set:* A limited data set is protected health information that excludes the following direct identifiers of the individual or of relatives, employers, or household members of the individual:
 - (3) *Implementation specification: Permitted purposes for uses and disclosures.*
 - (i) A covered entity may use or disclose a limited data set under paragraph (e)(1) of this section only for the purposes of research, public health, or health care operations.
 - (ii) A covered entity may use protected health information to create a limited data set that meets the requirements of paragraph (e)(2) of this section, or disclose protected health information only to a business associate for such purpose, whether or not the limited data set is to be used by the covered entity.
 - (4) *Implementation specifications: Data use agreement –*
 - (i) *Agreement required.* A covered entity may use or disclose a limited data set under paragraph (e)(1) of this section only if the covered entity obtains satisfactory assurance, in the form of a data use agreement that meets the requirements of this section, that the limited data set recipient will only use or disclose the protected health information for limited purposes.
 - (ii) *Contents.* A data use agreement between the covered entity and the limited data set recipient must:

EX2006 (cited by Sur-Reply, 12)

Evenhaim Teaches Away From the Combination's Proposed Searching of Fully De-Identified Data

143. The Petition's proposed combination of Evenhaim and Settimi would not have been obvious to a POSITA for the additional reason that it would have violated Evenhaim's express preference 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 (Dr. Rubin Dec.), ¶143 (cited by POR, 39-40)

144. For example, regarding the combination, the Petition contends that “[o]nce de-identified and stored in the central repository, the healthcare records can be queried using a token matching scheme ... using the token (EPID) to find the associated records for a specific patient without revealing the patient's identity.” Pet., 35. As described in the preceding section, Settimi never actually discloses searching the data warehouse using the token (EPID). But even if Settimi did disclose these features, the combination still would have required a non-obvious modification to Evenhaim's system in light of Evenhaim's desire to search “raw data at its original, most detailed form.” The data records stored in Settimi's central warehouse are not maintained in their original, most detailed form and do not include PHI. Not only is the data de-identified, thereby leaving the data records without useful information that Evenhaim sought to make available for search, but Settimi confirms that the data imported to the data warehouse before a search would be performed is aggregated, cleansed, and normalized—all contrary to Evenhaim's express teachings. EX1006, [0035] (“The data import function 206 aggregates and cleanses de-identified patient data from multiple sites and then stores the data into a staging area 208. Data received from multiple PCP systems 108 is *normalized*” and “*combined together* into a relational database”). Ground 2 thus fails for this additional reason.

EX2001 (Dr. Rubin Dec.), ¶144 (cited by POR, 39-40)

Evenhaim Teaches Away From the Combination's Proposed Searching of Fully De-Identified Data

As described above in this section, HIPAA provides substantial limitation on Personal Health Information and existing de-identification method may render the information useless
45 for the purpose of measuring the impact of health programs.

EX1005 (Evenhaim), 4:42-45 (cited by Sur-Reply, 4)

ment of the present invention, the following identifiers are removed or transformed in order to create de-identified data that would be classified under the HIPAA definition as fully de-identified data: name, geographic subdivisions smaller

EX1006 (Settimi), [0039] (cited by Sur-Reply, 11)

Evenhaim Teaches Away From the Combination's Proposed Searching of Limited Data Sets

As described above in this section, HIPAA provides substantial limitation on Personal Health Information and existing de-identification method may render the information useless
45 for the purpose of measuring the impact of health programs.

EX1005 (Evenhaim), 4:42-45 (cited by Sur-Reply, 4)

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
65 to aggregate the behavior of a group of patients by a pre-defined list of consumers participating in a health program. This is only one example of the limitations of the existing healthcare analytics model that is addressed by the instant protocol.

EX1005 (Evenhaim), 16:63-17:3 (cited by Sur-Reply, 4)

Settimi's Limited Data Sets are Not "De-Identified Healthcare Data" as Claimed

7.c	wherein the identity of the individuals in the group is unattainable from the tokens, each token being associated with a corresponding token in a set of <u>de-identified healthcare data</u> ,
7.d	and each token in the set of <u>de-identified healthcare data</u> being associated with healthcare data for a corresponding individual;

Claim 7- POR, page vi

as defined by HIPAA. A limited data set is protected health information for research, public health and health care operations that excludes direct identifiers (e.g., name; postal address other than city, state and zip code; social security number; medical records numbers) but in which other identifying information may remain (e.g., dates of examination; documentation; diagnosis; prescription; lab test results). This is contrasted with fully de-identified data as defined by HIPAA, where all data that may be used to trace back to an individual patient is removed from the record. Information

134, including Jack, in a PHI database 136. A de-identifier 142 can process the PHI 136 to generate irreversibly de-identified data 140 by removing all personally identifiable information 118 or otherwise transforming the PHI 130 so that it cannot be associated with a particular person. A claims

EX1006 (Settimi), [0027] (cited by Sur-Reply, 11)

EX1001 ('641 Patent), 3:43-47 (cited by Sur-Reply 12-13)

Settimi's EPIDs Fail to Render the Identity of Individuals in the Group "Unattainable"

Beyond the problems with modifying Evenhaim to search de-identified or limited data sets as discussed above, Datavant's proposed integration of Settimi's encoded patient identifiers (EPIDs) in Evenhaim for searching the central database suffers from additional problems that are indicative of further impermissible hindsight undercutting the Petition's obviousness theories. Pet., 40-43; Reply, 5.³

³ Additionally, because Settimi's EPIDs are re-identification codes that have not been irreversibly de-identified, they fail to render "the identity of the individuals in the group [] unattainable," as required by independent claims 1, 7, and 14. EX1006, [0041]-[0043], [0044]; EX2008, 35:13-18, 37:20-23; POR, 39-40; *cf.*, EX1001, 3:43-46, 5:65-67.

face 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

EX1006 (Settimi), [0041] (cited by Sur-Reply, 14)

Sur-Reply, 13-14

7.c

wherein the identity of the individuals in the group is unattainable from the tokens, each token being associated with a corresponding token in a set of de-identified healthcare data,

Claim 7 (Element 7.c) - POR, page vi

134, including Jack, in a PHI database 136. A de-identifier 142 can process the PHI 136 to generate irreversibly de-identified data 140 by removing all personally identifiable information 118 or otherwise transforming the PHI 130 so that it cannot be associated with a particular person. A claims

EX1001 ('641 Patent), 3:43-47 (cited by Sur-Reply 12-13)

Settimi Fails to Support use of EPIDs for Searching the Centralized Warehouse

[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 (Settimi), [0047] (cited by Sur-Reply, 14)

141. Importantly, Settimi’s queries necessarily include and rely on patient characteristics to identify the cohort because Settimi attempts to eliminate the traditional “use of mass media to broadcast a need *for patients who fit a list of clinical conditions*.” EX1006, [0011]. Querying the database of Settimi using EPIDs—as Petitioner contends—would be non-sensical at least because EPIDs do not represent patient characteristics or clinical conditions that would allow for identification of a cohort for clinical trials or other research as intended. EX1006, [0038]. Rather, as Petitioner and Dr. Clark admit, the database in Settimi is searched using a “common vocabulary” that normalizes terminology in the data from multiple sources. Pet., 36; EX1003, ¶137. Petitioner does not contend (and it would be meaningless to do so) that this common vocabulary includes EPIDs. It does not. Indeed, if particular EPIDs that fit the clinical study requirements were known to the entity generating the query, there would be no reason for the query in the first place. Accordingly, much like Evenhaim, Settimi never discloses querying a database for de-identified healthcare data using tokens or performing so-called “token matching” at a de-identified data server, and Petitioner’s premise for combining the two references—namely, that “a POSA would readily understand that the search would involve matching the tokens sent in the query to the tokens already in the relevant healthcare record” is not correct. Pet., 41; Ex. 1003, ¶153.

EX2001 (Dr. Rubin Dec.), ¶141 (cited by POR, 37)

Datavant Fails to Show How the Evenhaim's Aggregator Could Produce EPIDs When the PID and Password for Generating EPIDs is Known Only to PCPs

Second, Evenhaim's data aggregator would not be able to produce EPIDs to query Settimi's central database because EPIDs can only be generated by an entity that possesses confidential information known only to the PCPs. Settimi explains that only the PCP produces an EPID by encrypting a patient identifier (PID) with a "password string" that "is *known only to the PCP*" and that is also "required in order to decode the EPID [] into a PID []." EX1006, [0041]; EX2008, 32:9-33:8, 33:18-34:2. The PID is likewise a unique numeric key *assigned by the PCP* "to uniquely identify an individual patient" in the PCP's EMR. *Id.*, [0038]. The data aggregator in the Evenhaim-Settimi combination would be incapable of producing an EPID for the individuals identified in a request because the aggregator would have no access to either the PID or the password necessary to produce the PID.⁴

Sur-Reply, 15

Searching by EPIDs Would be Impractical Due to Variation in EPIDs Assigned Each Patient by Different PCPs

Third, searching by EPIDs would be impractical due to variation in the EPIDs assigned to each patient by different PCPs. Because the PCP is generated based on a PCP-specific patient identifier (PID) and a PCP-specific password string, the patient's data records in the warehouse would be associated with different EPIDs depending on where the PCP originated. EX1006, [0029], [0038], [0041], FIG. 3. Consequently, a search of the warehouse by a particular EPID for a patient would fail to surface all of the patient's records from different contrary to Evenhaim's goal of comprehensively analyzing patient data from multiple sources. EX1005, 6:20-25, 10:6-11, FIGS. 4-5; EX1006, [0041].

Sur-Reply, 16

Datavant Fails to Establish that Settimi's EPIDs are HIPAA Compliant

Fourth, although Datavant posits that a POSITA would have combined Evenhaim with Settimi to “reduc[e] the risk of a HIPAA violation,” Datavant fails to show that Settimi’s EPIDs are in fact HIPAA compliant. Pet., 37. The record evidence suggests they are not. HIPAA requires that any “code” used for re-identification of a patient not be “derived from or related to information about the individual” and further forbids disclosure of re-identification codes to third parties. EX2006, 2 (45 C.F.R. §164.514(c)). Yet, it is undisputed that the EPIDs disclosed in Settimi are re-identification codes derived from a patient identifier (PID) and disclosed to third parties (e.g., a central warehouse). EX1006, [0004],

Sur-Reply, 16-17

Datavant's Reliance on the Business Associate Exception to Remedy HIPAA Violations Fails

Sur-Reply:

As the outset, the business associate argument is a new theory that contradicts, and indeed fundamentally alters, the theories originally advanced in the Petition, which were specifically framed around de-identifying data for the express purpose of complying with the HIPAA Privacy Rule under 45 C.F.R. §164.514. *See, e.g.*, Pet., 6-7, 14, 34, 37, 43. The tokenization and de-identification techniques relied upon in the Petition are only necessary under HIPAA when there is *no* business associate agreement (since PHI could otherwise be shared freely), and noticeably, the Petition never once mentions business associates. EX2008, 43:11-14, 46:7-17.

Sur-Reply, 18-19

The argument also fails on substance. No evidence/explanation exists in the record as to why it would be necessary or obvious to implement de-identification and tokens for sharing data between entities under a business associate agreement when the very purpose of a business associate agreement is to allow for the sharing of PHI that has *not* been de-identified. EX2003, 5; EX2021, 1-2; EX1020, 3. But if the Evenhaim-Settimi system did not search de-identified data or use tokens, the system could not read on limitations of the '641 claims requiring such features. EX2008, 43:11-14, 46:7-17.

Sur-Reply, 19

Petition:

At the very least, this token matching would have been obvious in view of Evenhaim, given that token matching was well known in the field. A POSA would have understood that matching the tokens was an efficient way to query de-identified data, and would ensure compliance with HIPAA. Ex. 1003, ¶¶39-41.

Pet., 14

[0048], claim 5; Ex. 1003, ¶50. Specifically, Settimi describes a centralized warehouse of de-identified patient records, collected from various patient care providers. Ex. 1006, Fig. 2, [0027]; Ex. 1003, ¶50. This central repository allows the patient data to be queried for, e.g., research, while keeping the centralized records HIPAA compliant. Ex. 1006, [0027], [0030], [0035] (“The de-identified

Pet., 34

36. Also, per Settimi, the central repository/warehouse for searching contains only de-identified healthcare records, reducing the risk of a HIPAA violation. For these

Pet., 37

Datavant's Reliance on the Business Associate Exception to Remedy HIPAA Violations Fails

Datavant also overlooks that “[c]overed entities may disclose protected health information to an entity in its role as a business associate *only to help the covered entity* carry out its health care functions – *not for the business associate’s independent use or purposes.*”⁶ Neither Evenhaim’s Crossix system (e.g., the data aggregator) nor Settimi’s central warehouse receive data that originated from the data source entities/PCPs (covered entities) for the purpose of aiding the data source entities/PCPs in carrying out their healthcare functions, as would be required under a valid business associate agreement. Evenhaim instead explains, for example, that the aggregator functions for its own independent commercial purpose including “assist[ing] ... marketing managers from pharmaceutical manufacturers or other health promotion managers” to aggregate de-identified patient data and measure the impact of a health program. EX1005, 15:3-8. Dr. Clark was likewise unable to offer any further clarity on the proposed use of business associate agreements in the combination at his deposition. EX2008, 52:10-53:1, 54:14-57:7, 53:6-54:14.

Sur-Reply, 19-20

Ground 3: Dick-Landi

Overview of Dick

- Searches archived medical information repositories
POR, 42-44
- Searches performed using personal information
 - UMPI or L-MPI
POR, 42-44
- Discloses de-identification methodologies
POR, 48-49

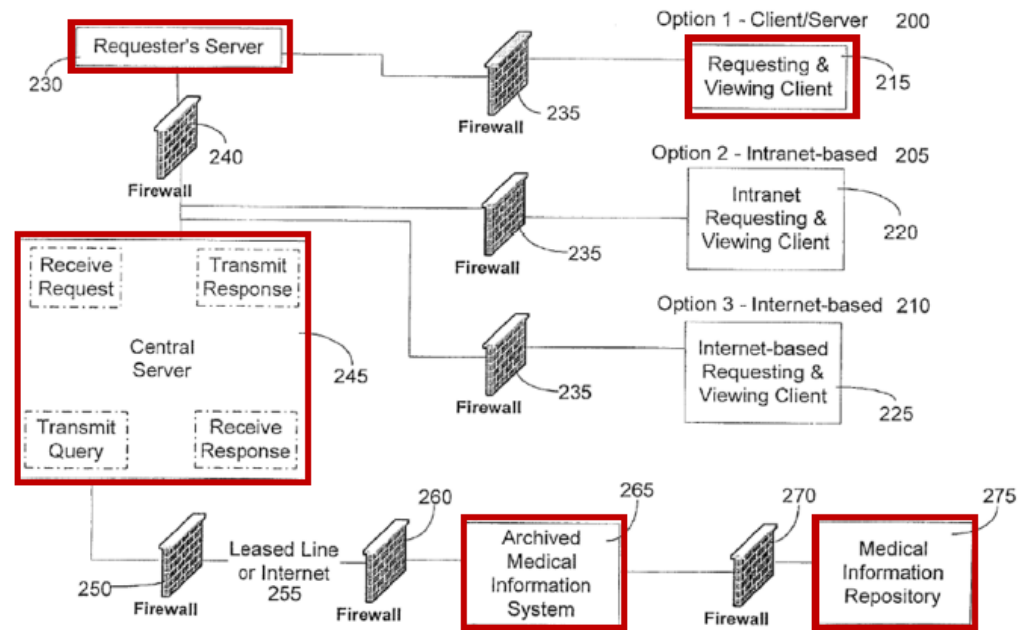


FIG. 3

EX1007 (Dick), FIG. 1 (cited by POR, 41)

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

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

POR, 42, 56

[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] (cited by POR, 42)

- Searching done using Master Person Index

POR, 42-44

[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] (cited by POR, 43)

No HIPAA Violation in Dick That Would Inspire Need for Tokens

- Petition alleges potential HIPAA violation as a motivation to combine Dick and Landi

Pet. 57-58

- But no HIPAA violation exists, so no motivation to combine

POR, 55-57

A POSA would have understood that creating encrypted tokens on the Requester's Server 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, ¶214. For example, it would potentially violate HIPAA to just send PII to be used for such a query directly to the centralized data server (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, ¶214.

Pet. 57-58

No HIPAA Violation in Dick That Would Inspire Need for Tokens

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

POR, 55-57

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 (Dr. Clark Depo), 57:15--25 (cited by POR, 62)

Dick's Requester's Server

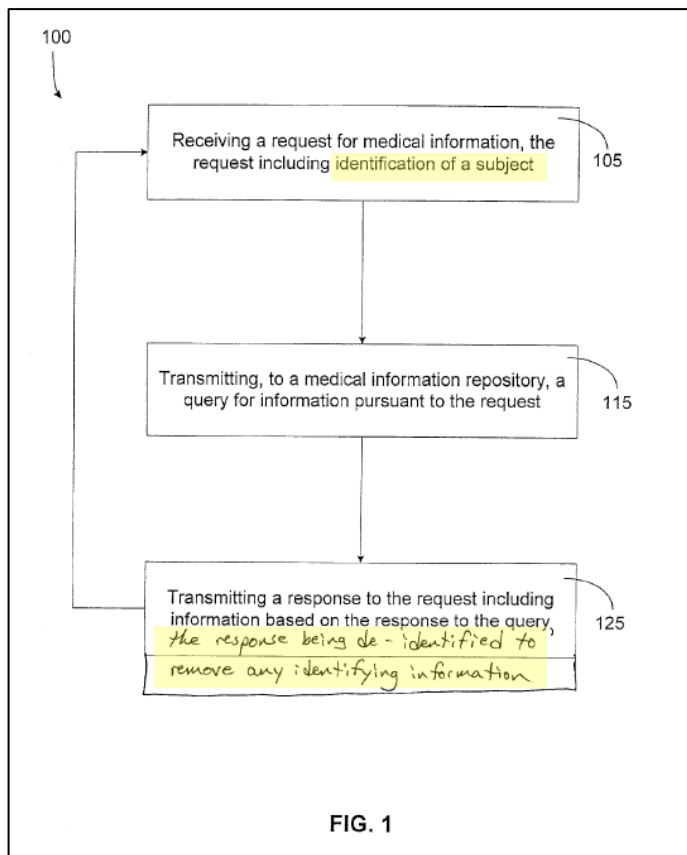
- The RVC 215 can “receive a request from an authorized user responding to prompts from software executing on requester’s server 230.” (EX1007, [0022]).
- Notably, this is the only function attributed to the requester’s server 230 in the entirety of Dick’s disclosure. (EX2001, 83–89; EX2002, 75:19–24).
POR, 42

[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(Dick), [0022] (cited by POR, 42)

Dick already teaches de-identification techniques

- Contrary to the Petition, Dick already teaches details for de-identifying techniques
POR, 43-44, 48-53



EX1007(Dick), FIG. 1 (cited by Pet., 65)

[0045] Before information from a medical information repository could be released to a research company hoping to identify individuals or groups of people or populations to solicit for pharmaceutical studies, the individuals would have to give consent or the medical information would have to be altered such that no person can be individually identified from the health information. "De-identifying" information requires more than simply removing names or addresses of patients, since in some circumstances, a combination of medical and related non-medical information could be used to individually identify a person whose information is being distributed. In order to legally distribute such information, a medical information repository would have to remove all "Identifying" information including: identifiers of the individual or of relatives, employers, or household of members of the individual associated with the record, which identifiers include names, geographic subdivisions, elements of dates related to any individual, telephone numbers, fax numbers, e-mail addresses, Social Security numbers, medical record numbers, health plan beneficiary numbers, account numbers, certificate license numbers, vehicle identifiers and serial numbers, including license plate numbers, device identifiers and serial numbers, URLs, IP addresses, biometric identifiers, photographic images, any other unique identifying number, characteristic or code. Where removing such information is not practical or renders the data useless, the medical information repository may instead release medical information upon condition that a person with appropriate knowledge and experience with generally accepted statistical and scientific principles has applied a "de-identifying" methodology to remove or make certain that no personally identifiable information is released from the repository.

[0046] The present invention contemplates the use of "de-identifying" methods and processes for allowing information from the present invention to be distributed for use in medical research. Furthermore, the benefits of the release of this information extends beyond pharmaceutical research to other areas, such as marketing and public health. The information distributed by the present invention can be

Dick already teaches de-identification techniques

- Dr. Clark conceded that Dick's teachings of de-identification methodologies was sufficient for a POSITA
POR, 48-53
- No reason to look to Landi for de-identification methodologies
POR, 48-53

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.

EX2002(Dr. Clark Depo), 80:11–25 (cited by POR, 50)

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.”

POR, 44

- Landi proposed techniques for de-identifying PHI that would allow the PHI to be re-identified.

POR, 44, 46-47, 54, 57, 63

- Landi describes “conventional methods for de-identifying” to include “simply stripping” PII

- Petition concedes that Landi “explains that prior art methods simply stripped PII.” Pet. 53.

POR, 50-51

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(Landi), 1:13-22 (cited by POR, 44)

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(Landi), 2:19-26 (cited by POR, 41)

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 site 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(Landi), 4:11-34 (cited by POR, 44-45)

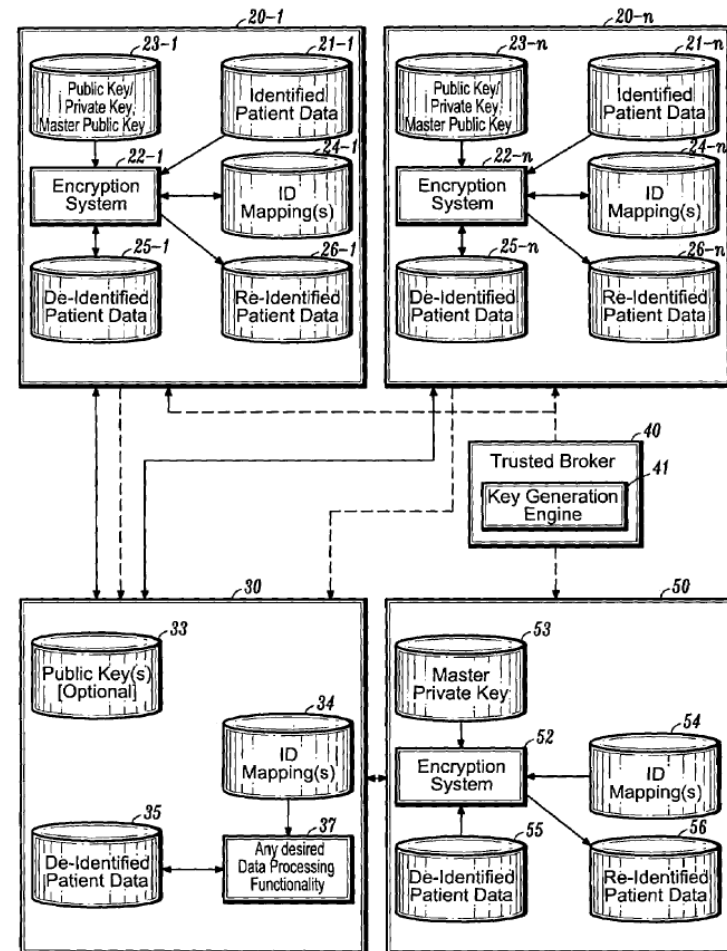


FIG. 1

EX1008(Landi), FIG. 1 (cited by POR, 55)

Landi's Encrypted IDs Are Not Absent Personally Identifiable Information

- Encrypted IDs include encrypted representation of PII
Sur-Reply, 21-22
- PII is re-identified from Encrypted IDs
 - EX1008, 2:36–40, 2:49–53, 3:9–13, 3:42–57, 4:28–34, 4:35–48, 4:54–64, 6:56–7:2, 7:30–38, 7:43–48, 8:5–10, 8:16–37, 8:38–55, 9:10–16, 9:36–42, 10:12–14, 10:32–37.

POR, 42

In one exemplary embodiment of the invention, a method for processing data includes the steps of obtaining a data record of an individual which includes individual identifying information, removing the individual identifying information in the data record to generate a de-identified data record, generating an encrypted ID for the individual, wherein the encrypted ID comprises an encrypted representation of one or more items of individual identifying information, and storing the encrypted ID with or in the de-identified data record. A decryption key is securely maintained and accessible by an authorized entity that is legally authorized or empowered to decrypt the encrypted ID in the de-identified data record to re-identify the individual.

EX1008(Landi), 2:46-48 (cited by Sur-Reply, 22)

With the encryption system operating in a de-identification mode, an initial step is to use the encryption key (e.g., public key) to generate an Encrypted ID for each subject patient of the identified patient data records (step 53). As noted above, 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. The Encrypted ID for each patient is generated such that the patient identifying information contained within the Encrypted ID cannot be reconstructed without a valid decryption key (e.g. private

EX1008(Landi), 11:6-29 (cited by Sur-Reply, 22)

Landi's Encrypted IDs Are Not Absent Personally Identifiable Information

- Dr. Clark testified that PII is not absent from the produced healthcare information.

POR, 62-63

- Instead, PII remains in the produced healthcare information as encrypted ID.

POR, 62-63

```
23      Q.  And so that de-identified information in Landi
24  is -- includes the encrypted ID, correct?
25      A.  It could.
1      Q.  In the combination, the encrypted ID would be
2  routed to the requesting client through the requester's
3  server?
4      A.  That's possible, sure.
5      Q.  That's consistent with your description of the
6  combination?
7      A.  I think so.
```

EX2002(Dr. Clark Depo), 93:23-94:7 (cited by POR, 62-63)

Landi's Encrypted IDs Are Not Absent Personally Identifiable Information

- Dr. Rubin explained that PII remains in Landi's encrypted ID
POR, 63

169. During his deposition, Dr. Clark clarified that in Landi, as applied to the combination, personally identifiable information (PII) is not “absent” from the produced healthcare information. Instead, the PII remains in the produced healthcare information as encrypted information (e.g., an encrypted ID). EX2002, 86:9–24. Additionally, as mentioned before, Landi's encryption methodology is designed to permit de-identification of PII. *See, supra* §VIII.D. Thus, consistent with Dr. Clark's explanation, PII would be available in the produced healthcare information and could be read simply by applying the decryption algorithm and re-identification techniques expressly described in Landi. EX1008, 2:36–40, 2:49–53, 3:9–13, 3:42–57, 4:28–34, 4:35–48, 4:54–64, 6:56–7:2, 7:30–38, 7:43–48, 8:5–10, 8:16–37, 8:38–55, 9:10–16, 9:36–42, 10:12–14, 10:32–37. The Dick-Landi combination therefore does not teach that the PII is absent from the produced healthcare information as claimed in elements [1.h], [7.h], or [14.h].

EX2001 (Rubin Dec.), ¶169 (cited by POR, 63)

Landi's Encrypted IDs Are Not Absent Personally Identifiable Information

- Datavant conflates the “absence” of PII as claimed with the accessibility of the PII
Sur-Reply, 22
- PII is absent from the produced healthcare information in the '641 patent when it is “irreversibly de-identified”, unlike Landi's encrypted ID
Sur-Reply, 22

decrypt it.” Reply, 24. But Datavant's argument conflates the “*absence*” of PII as claimed with the *accessibility* of PII (not claimed). Under Datavant's logic, the money in a safe stocked with cash would be “absent” from the safe just because a key is needed to unlock it. Just as cash is plainly present (not absent) in the safe, PII is also present (not absent) in Landi's encrypted ID, and Datavant does not contend otherwise. EX1008, 2:46-48, 5:18-19, 11:6-29, FIG. 3. Nor could Datavant contend otherwise, as this distinction implicates a key aspect of the '641 Patent's disclosure, which explains how produced healthcare information (e.g., report 212) does not simply limit access to PII data found therein; it instead generates a report based on irreversibly deidentified information, with no PII, consistent with the “absence” of PII required by the plain language of elements [1.h], [7.h], and [14.h]. See, e.g., EX1001, 6:8-15 (“report generator 328 [] process[es] the requested de-identified data 208 in a way that leaves it irreversibly de-identified”), 4:52-54, 7:11-15.

Sur-Reply, 22

Landi's Does Not Teach Token Matching

- No express disclosure of token matching
 - Cited portions of Landi discuss de-identification techniques, not token matching

POR, 60-62

- Landi does not discuss searching by Encrypted IDs
 - FIG. 1: Central Server 30 not accessible by external entities

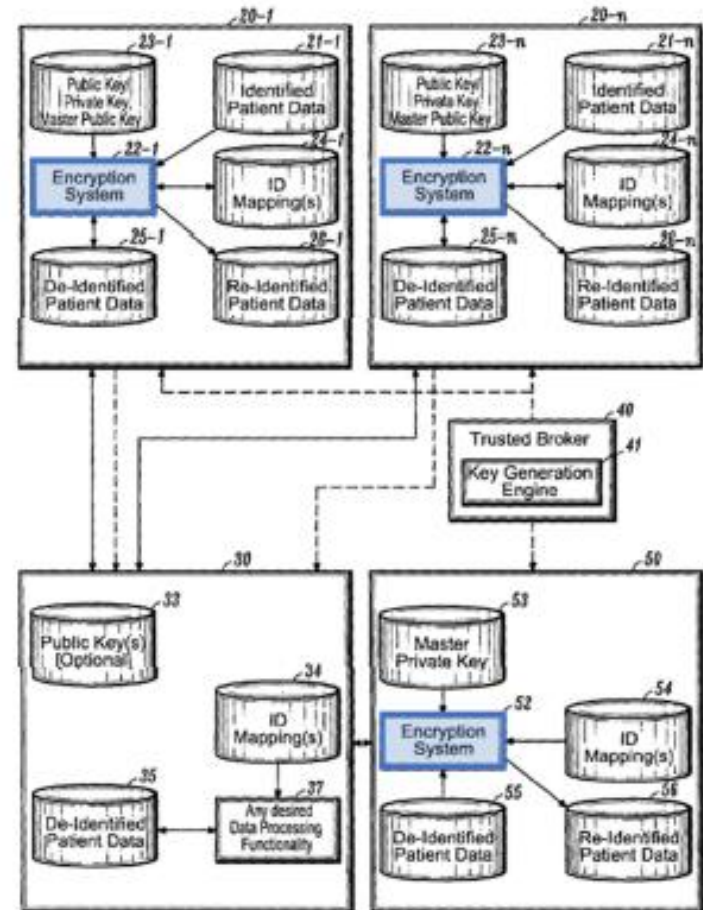


FIG. 1

EX1008(Landi), FIG. 1 (cited by POR, 55)

Landi Does Search the Centralized Database for Known Individuals Using an Encrypted ID

It is to be appreciated that in the exemplary system of FIG. 1, a secure public/private key encryption scheme enables non-restricted, free dissemination of public keys that are used for generating de-identified patient that is stored in de-identified data records which are disclosed or otherwise shared with third party entities, while protecting patient privacy by allowing only those holders of valid private keys (e.g., owners of the patient data, or legally empowered entities) to decode the de-identified patient data and re-identify patients, when necessary. The public keys are freely distributable (no HIPAA issues) to healthcare organizations for de-identification of patient data. Moreover, the public keys can be used to produce a unique, reproducible encrypted version of patient identifying information. Since the encryption is reproducible, the same patient can be recognized at different times (i.e., consistent de-identification can be achieved over episodic collection events) even though the identity of the person remains unknown.

EX1008(Landi), 8:23-55 (cited by POR, 61)

- Landi uses the same encryption algorithm and public key to encrypt the same PII to achieve consistency in the assignment of encrypted IDs to the same patients' records from different providers or at different times.
- Landi's centralized database is not searched by encrypted ID.

POR, 60-61

The Dick-Landi Combination Is Flawed

- Landi's encrypted IDs are not HIPAA compliant
- No HIPAA Violation in Dick
 - This is the only motivation articulated by the petition to replace Dick's requester's server with an encryption server
- The combination compromises HIPAA security
 - Placing the encryption at the requester's server or central server means that either entity can re-identify
- Dick already teaches de-identification techniques, so no motivation to combine with Landi

Landi's Encrypted IDs Are Not HIPAA Compliant

- (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, § 164.514(c) (cited by POR, 52)

HIPAA prohibits derivation of re-identification codes from PII

Landi's Encrypted IDs Are Not HIPAA Compliant

- Dr. Clark failed to independently verify whether Landi's representation regarding HIPAA compliance was accurate

Sur-Reply, 23-24

3	Q. Did you check whether the author of Landi is a
4	HIPAA expert before stating in your declaration that a
5	POSA would have no reason to doubt Landi's statement
6	regarding HIPAA compliance?
7	A. Of course not. I didn't -- I didn't pull
8	their biography, Counselor.
9	Q. Well, did you check whether Landi's encrypted
10	IDs comply with the provisions in subsection (c) of the
11	HIPAA Privacy Rule?
12	A. I did not because the statement that I made
13	also allowed for associate agreements, so I had no
14	reason to doubt that the system could be implemented in
15	a HIPAA-compliant manner.

EX2008(Dr. Clark 2nd Depo), 69:3-15 (cited by Sur-Reply, 24)

Landi's Encrypted IDs Are Not HIPAA Compliant

- Dr. Clark agreed that Landi's encrypted ID violates HIPAA's proscription against deriving re-identification codes from PII

Sur-Reply, 25

6	Q. Let's consider a scenario where there is no
7	business associate agreement in place. If we generated
8	Landi's encrypted ID in this manner, would it violate
9	the re-identification provision that we just reviewed
10	from the HIPAA Privacy Rule?
11	A. Strictly speaking, if you did not have an
12	associate agreement, then you would not be compliant
13	with the derivation requirement.

EX2008 (Dr. Clark 2nd Depo), 71:6-13 (cited by Sur-Reply, 25)

A POSITA Would Not Destroy Dick's HIPAA Compliance by Using Landi's Encrypted IDs

Dr. Clark's admission is striking considering that HIPAA compliance was of paramount importance to a POSITA. See, e.g., EX2008, 61:8-23; EX1003, ¶¶34-36, 201, 203, 214-215, EX2001, ¶¶64, 72, 161-162. And this importance is apparent from even Dick, the primary reference in Ground 3, whose "purpose" is described in the Petition as "compliance with HIPAA." Pet., 53. Even Dr. Clark emphasized Dick's goal of HIPAA compliance and agreed that "it wouldn't make sense" to modify Dick to implement features that were not HIPAA compliant. EX2008, 62:9-18; EX1003, ¶201 ("Dick describes or suggests de-identification for complying with HIPAA."), ¶¶202-203. In this context, a POSITA would not have been motivated to implement Landi's non-compliant encrypted IDs in Dick's system as Datavant proposes, and Ground 3 should thus be rejected in its entirety. EX2001, ¶¶169-173, 195-198.

Sur-Reply, 25

Like Dick, Landi's purpose is compliance with HIPAA. Ex. 1008, 1:58-2:18, 3:52-57, 9:20-29, 10:3-31, 11:64-12:67. And, both references are directed to using

Pet, 53

12	Q. Given that one of Dick's goals was HIPAA
13	compliance, would a POSITA be motivated to implement
14	features in Dick's system that were not HIPAA compliant?
15	A. That wouldn't make sense to me. But, again,
16	as we've discussed, there are lots of ways to achieve
17	HIPAA compliance which are not specific to the Privacy
18	Rule alone.

EX2008(Dr. Clark 2nd Depo), 62:12-18 (cited by Sur-Reply, 25)

Dick-Landi Combination Breaches Patient Privacy by Allowing Encryption Server to Correlate Data to Identify Patients

- Replacing Dick's Requester's Server with Landi's Encryption System would violate HIPAA
POR, 58-59
- Dr. Clark testified the encrypted IDs are routed through the requester's server
 - MIR/AMIS violates HIPAA
 - Requester's server would be able to link patient names to their medical records
POR, 58-59

```
23      Q.  And so that de-identified information in Landi
24  is -- includes the encrypted ID, correct?
25      A.  It could.

1      Q.  In the combination, the encrypted ID would be
2  routed to the requesting client through the requester's
3  server?
4      A.  That's possible, sure.
5      Q.  That's consistent with your description of the
6  combination?
7      A.  I think so.
```

EX2002(Dr. Clark Depo), 93:23-94:7 (cited by POR, 58-59)

Dick-Landi Combination Breaches Patient Privacy by Allowing Encryption Server to Correlate Data to Identify Patients

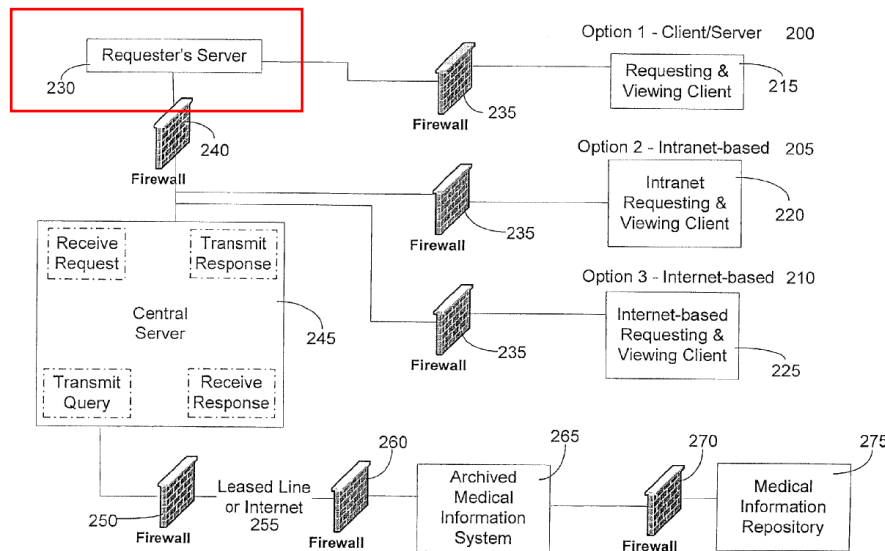


FIG. 3

EX1007(Dick), FIG. 3; POR, 41

Dr. Rubin:

192. The proposed combination would 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. EX2002, 93:5-94:7. 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. 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. 2002, 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(Dr. Rubin Dec), ¶192; POR, 58-59

Ground 4: Dick-Landi-Murphy

Datavant Errs in Relying on Murphy as “Block[ing]” Queries to Meet “Minimum Number” Limitations of Claims 1-6, 11, 12, 14-19

1. A computer-implemented method comprising:

when the group includes at least a minimum number of individuals, producing the requested healthcare information that characterizes the group by identifying the tokens in the set of de-identified healthcare data and the associated de-identified healthcare data that correspond to the produced tokens, wherein the produced tokens and the tokens in the set of de-identified data are similarly encrypted, and the produced healthcare information is absent personally identifiable information and has been produced without the individuals in the group having authorized production of such healthcare information that characterizes the group; and

5. The computer-implemented method of claim 1, in which producing the requested healthcare information includes identifying at least a minimum number of corresponding tokens.

11. The system of claim 7, in which producing the requested healthcare information includes identifying at least a minimum number of corresponding tokens.

12. The system of claim 7, in which producing the requested healthcare information includes determining if the group includes at least a minimum number of individuals.

Claims 14 and 18 parallel claims 1 and 5, respectively
EX1001('641 Patent), cl.1,5,11,12

Petition:

Murphy makes clear that, in complying with HIPAA, an algorithm preferably blocks queries that return fewer than 3 patients to prevent the user from associating the patient with a particular record. Ex. 1009, 555. (“[W]e 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, ¶242. Ensuring a minimum number of tokens are identified would have been a simple design choice to comply with HIPAA. Ex. 1003, ¶¶66, 239-40.

Pet, 69

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(Murphy), 553 (cited by POR, 64)

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(Murphy), 553-54 (cited by POR, 64-65)

Murphy Does Not “Block Queries”

Dr. Rubin:

201. First, the Petition’s assertion that Murphy’s algorithm “blocks queries” when the number of patients is less than 3 is incorrect. 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.*

EX2001(Dr. Rubin Dec), ¶201 (cited by POR, 48)

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, naïve 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 than 3.

EX1009(Murphy), 555 (cited by POR, 65, 67 and Sur-Reply, 27)

Dr. Clark Acknowledged “Possibility” that Murphy Returns Obfuscated Result Regardless of Minimum No. Threshold

9 Q. And at page 555, in the right-hand column, at
10 the last sentence of the paragraph that begins with
11 "although," it states, (as read) "Therefore, we modify
12 our range algorithm to have a wider obfuscation at
13 values close to 0 and indicate the greater range by
14 returning '<3' when the obfuscated number is less than
15 3."
16 Correct?
17 A. That's what it says.

EX2008(Dr. Clark 2nd Depo), 76:9-17 (cited Sur-Reply, 27)

Q. So what I'm driving at is that Murphy is saying here that the result returned to the user is obfuscated both when the result is less than the threshold and greater than the threshold; is that correct?
A. That's one possibility that you -- I mean, it's talking about generally, as I said, thresholding, and so one possibility is to never return an exact number.
Q. There's an obfuscated result regardless of whether it's above or below the threshold, correct?
A. That's one possibility.

EX2008(Dr. Clark 2nd Depo), 77:5-16 (cited by Sur-Reply, 27)

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(Murphy), 553 (cited by POR, 68)

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(Murphy), 553 (cited by POR, 68)

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

POR, 68

Thank You