Internal distribution code:

(A) [-] Publication in OJ
(B) [-] To Chairmen and Members
(C) [-] To Chairmen
(D) [X] No distribution

Datasheet for the decision of 12 March 2019

Case Number: T 1248/12 - 3.5.01

Application Number: 05718937.5

Publication Number: 1761893

IPC: G06F11/30

Language of the proceedings: EN

Title of invention:
A PRIVACY PRESERVING DATA-MINING PROTOCOL

Applicant:
Crossix Solutions LLC

Headword:
Privacy preserving data mining/CROSSIX

Relevant legal provisions:
EPC Art. 56

Keyword:
Inventive step - de-identifying data (no - not technical)

Decisions cited:
T 0424/03, T 0769/92, T 0641/00

Europäisches Patentamt
European Patent Office
Office européen des brevets

Beschwerdekammern

Boards of Appeal

Chambres de recours

Boards of Appeal of the
European Patent Office
Richard-Reitzner-Allee 8
85540 Haar
GERMANY
Tel. +49 (0)89 2399-0
Fax +49 (0)89 2399-4465

Case Number: T 1248/12 - 3.5.01

D E C I S I O N
of Technical Board of Appeal 3.5.01
of 12 March 2019

Appellant: (Applicant)	Crossix Solutions LLC 201 E. 86th 17c New York, NY 10028 (US)
Representative:	Lermer, Christoph LermerRaible Patent- u. Rechtsanwalts PartGmbB Lessingstrasse 6 80336 München (DE)
Decision under appeal:	**Decision of the Examining Division of the European Patent Office posted on 21 February 2012 refusing European patent application No. 05718937.5 pursuant to Article 97(2) EPC.**

Composition of the Board:

Chairman	W. Chandler
Members:	A. Wahrenberg
	P. Schmitz

Summary of Facts and Submissions

I. This appeal is against the decision of the examining division to refuse the European patent application No. 05718937.5 (published as WO 2005/094175 A2) on the grounds that the subject-matter of claim 1 of the main and three auxiliary requests did not involve an inventive step (Article 56 EPC).

II. The examining division took the view that the claimed privacy-preserving data mining protocol was not technical, essentially for the reasons that it could be carried out without technical means, and served a non-technical purpose (compliance with legal requirements). Starting from a notorious data processing system, the examining division found that the implementation of the non-technical data mining protocol would have been straightforward and obvious.

III. In the statement setting out the grounds of appeal, the appellant requested that the decision of the examining division be set aside and that a patent be granted on the basis of the refused main request or one of the three refused auxiliary requests.

IV. In the communication accompanying the summons to oral proceedings, the Board preliminarily agreed with the examining division that the claimed data mining protocol was not technical. Moreover, the Board considered that the definition of the different types of processing (aggregating, agglomerating, and "crunching together") in claim 1 was not clear.

V. The Board held oral proceedings on 12 March 2019, during which the appellant filed a fourth auxiliary

request. The appellant's final requests were, consequently, that the decision under appeal be set aside and that a patent be granted on the basis of the main request, or one of the first to fourth auxiliary requests.

VI. Claim 1 of the main request reads:

A Privacy Preserving Data-Mining Protocol, operating between a secure "aggregator" data processor and at least one of "source-entity" data processor, wherein the "aggregator" and the "source-entity" processors are interconnected via an electronic data-communications topology, and the protocol includes the steps of:

A) on the side of the "aggregator" processor:

(i) from a user interface--accepting a query against a plurality of the predetermined attributes and therewith forming a parameter list,

(ii) via the topology--transmitting the parameter list to each of the "source-entity" processors,

(iii) via the topology--receiving a respective file from each of the "source-entity" processors,

(iv) aggregating the plurality of files into a data-warehouse,

(iv[*sic*]) using the parameter list, extracting query relevant data from the data-warehouse,

(vi) agglomerating the extract, and

(vii) to a user interface--reporting the

agglomerated extract; and

B) on the side of each processor of the at least one "source-entity" processors:

(i) accumulating data-items wherein some of the data-items have privacy sensitive micro-data,

(ii) organizing the data-items using the plurality of predetermined attributes,

(iii) via the topology--receiving a parameter list from the "aggregator" processor,

(iv) forming a file by "crunching together" the data-items according to the parameter list,

(v) filtering out portions of the file which characterize details particular to less than a predetermined quantity of micro-data-specific data-items, and

(vi) via the topology--transmitting the file to the "aggregator" processor.

VII. Claim 1 of the first auxiliary request adds "configured as a central data processing machine" immediately after the words "a secure 'aggregator' data processor", and the text "configured as a data processing machine" after "at least one "'source-entity' data processor", in the opening part of the claim.

VIII. Claim 1 of the second auxiliary request adds to the first auxiliary request "wherein said parameter list let [*sic*] the source-entity begin to participate in the user initiated process" at the end of feature B)(iii):

IX. Claim 1 of the third auxiliary request adds to the second auxiliary request:

"which may include specific individuals' names or IDs" at the end of feature A)(i);

"including de-identified results" after the word "file" in feature A)(iii);

"wherein 'crunching together' includes de-identifying results by aggregating results to a tabular report" at the end of feature B)(iv).

X. Claim 1 of the fourth auxiliary request adds to the main request:

"aggregating including bundling the responses of the individual source entities into a large source-entity de-identified data collection" at the end of feature A)(iv).

"including filtering out portions of the extract which characterize details particular to less than a predetermined quantity data items," at the end of feature A)(vi).

"involving name matching through fuzzy logic and other matching algorithms" at the end of feature B)(iv).

"implementing a privacy threshold at the source entity" at the end of feature B)(v).

XI. The appellant's arguments can be summarised as follows:

The particular configuration of processors, as well as

the interrelationship between those processors and the data mining-protocol, was based on technical considerations.

The claimed protocol used functional data structures, similar to the ones in decisions T 424/03 (Clipboard formats I/MICROSOFT) and T 769/92 (General purpose management system/SOHEI) to facilitate and manage the exchange of data between the different data processors. Furthermore, signal conversion and data transmission in an electronic system were technical features.

The data structures comprised, as shown in Figures 4A and 5:

a parameter list (study package) that was generated by the aggregator based on the user's query and forwarded to the aggregator;

the files (analysis result) that were generated by the source entities and sent to the aggregator;

the aggregated files (aggregate analysis result) generated by the aggregator.

The reception of the parameter list by the source entity triggered a particular technical function in the source entity. Furthermore, the source-entity files and the aggregator files had particular formats that were independent of the cognitive content.

The invention increased data privacy, i.e. it solved a technical problem relating to data security.

Reasons for the Decision

1. *Background*

1.1 The invention concerns data privacy in a database system.

1.2 The processing of privacy sensitive data, such as medical records, is subject to legal restrictions. For example, the Health Insurance Portability and Accountability Act 1996 (HIPAA) in the USA prevents health care providers from sharing individually identifiable health information with third parties, such as researchers or pharmaceutical companies. However, it is possible to share de-identified data that does not identify or provide information that could identify an individual.

Similar data privacy laws exist in Europe, and in other jurisdictions.

1.3 De-identification is a lossy process in that information is removed. Therefore, it might not be possible to extract certain information from the de-identified data, even if this information does not breach individual privacy. The invention seeks to overcome this problem.

2. *Main request, claim 1*

2.1 Claim 1 defines a data mining protocol that operates between an "aggregator" and a number of "source-entities". The source entities correspond to health care providers as in the example in paragraph 1.2 above. The "aggregator" is a trusted, central

processor.

Initially, the Board had doubts whether claim 1 defined the data processing performed by the various processors in a sufficiently clear manner. In particular, the Board did not see any clear difference between the steps of "aggregating", "agglomerating", and "crunching-together". During oral proceedings before the Board, the appellant explained the claimed data-mining protocol as follows:

A user, for example a researcher who wants to get information about a group of people, inputs a query including, for example, the names or IDs of the people in the group. The query (or "parameter list") is sent, via the aggregator, to the source entities that store the data.

The source-entities collect the relevant data into files (the data items are "crunched together"); they de-identify the data to a certain extent, for example by removing addresses, and send the files to the aggregator that aggregates them into a data warehouse. The aggregation further protects privacy by de-identifying the source-entities. The aggregator also extracts query-relevant data from the data warehouse, and presents a condensed ("agglomerated") extract to the user.

The Board is satisfied that the appellant's example falls within the terms of claim 1, and the Board is able to assess inventive step on the basis of this example.

3. Main request, inventive step

3.1 The examining division found that the data mining protocol in claim 1 was an administrative scheme, which, when considered on its own, constituted excluded subject-matter according to Article 52(2) and (3) EPC. The examining division could not identify any technical problem solved by the data mining. In the examining division's opinion, the aim of the data processing was rather to comply with legal requirements.

3.2 The Board shares the examining division's view that de-identifying data, by removing individually identifiable information, and by aggregating data from a plurality of sources, is not technical. It aims to protect data privacy, which is not a technical problem. The problem of data privacy is not synonymous with data security. Data privacy concerns what information to share and not to share (and making sure that only the information that is to be shared is shared), whereas data security is about how to prevent unauthorised access to information.

3.3 It is established case law that non-technical features cannot contribute to inventive step. Therefore, non-technical features may legitimately be part of the problem to be solved (T 641/00 - Two identities/ COMVIK), for example in the form of a requirement specification given to the skilled person to implement.

3.4 The examining division assessed inventive step starting from a generic data processing system. During the oral proceedings before the Board, the question of inventive step was also discussed in view of the prior art described in the application. This prior art includes at least a database system corresponding to the source-

entities in claim 1. Thus, the invention as defined in claim 1 differs from the prior art database system in that:

- the user's query goes via an "aggregator";
- the query result is de-identified, at the source, and "crunched together" into a file that is sent to the aggregator, which:
    - aggregates files from a plurality of source entities;
    - extracts query relevant data;
    - agglomerates the extract; and
    - presents the agglomerated extract.

3.5 As concluded in paragraph 3.2 above, the steps of de-identifying the data at the source and aggregating the results from a plurality of sources is part of the non-technical requirement specification to be implemented. So is the presentation of the result in a condensed form.

3.6 The skilled person having been given the task of implementing the requirement specification would provide an "aggregator processor", because that is what the requirement specification ("aggregate the results") is telling him to do. The aggregator processor and the database system (source-entities) need to communicate with each other: the source entities need to obtain the query and the aggregator processor needs to obtain the query results. The skilled person would find suitable formats for this. The Board notes that the claims do not specify any particular format beyond the use of a "list" and files. The processing performed by the source-entities (de-identifying) and aggregator (aggregating, extracting and agglomerating), and the presentation of the results to the user, does not go

beyond what the requirement specification dictates.

Thus, the skilled person would have arrived at the subject-matter of claim 1 without inventive effort.

3.7 For these reasons, the Board concludes that the subject-matter of claim 1 of the main request lacks an inventive step (Article 56 EPC).

4. *First auxiliary request*

4.1 The additional features in claim 1 of the first auxiliary request, i.e. that the aggregator is configured as a central data processing machine, and that the source-entity is configured as a data processing machine, are already implied by the features of the main request. Therefore, the Board's reasons with regard to the main request are applicable also to the first auxiliary request.

5. *Second auxiliary request*

5.1 Claim 1 of the second auxiliary request specifies that the "parameter list let [*sic*] the source-entity begin to participate in the user initiated processor". In other words, the parameter list or query triggers the database to be queried.

5.2 The Board does not see that this feature adds anything of significance over the main request. It goes without saying that, without a query, the database cannot be queried. Therefore, the same reasons apply to the second auxiliary request.

6. *Third auxiliary request*

6.1 The Board's reasoning with regard to the main request already takes into account the additional features in claim 1 of the third auxiliary request that the parameter list may include the names or ID's of individuals, and that the files received by the aggregator from the source entities include de-identified results.

6.2 The third auxiliary request further specifies that the "crunching together" includes de-identifying results by aggregating results to a tabular report. The Board does not consider that this adds anything technical, for the reasons already provided in connection with the main request, i.e. because de-identifying data is not technical.

7. *Fourth auxiliary request*

7.1 The fourth auxiliary request was filed during oral proceedings before the Board.

According to Rule 13(1) RPBA, any amendment to a party's case after it has filed its grounds of appeal or reply may be admitted and considered at the Board's discretion. The discretion shall be exercised in view of *inter alia* the complexity of the new subject matter submitted, the current state of the proceedings and the need for procedural economy.

Furthermore, amendments sought to be made after oral proceedings have been arranged shall not be admitted if they raise issues which the Board or the other party or parties cannot reasonably be expected to deal with

without adjournment of the oral proceedings (Article 13(3) RPBA).

7.2 The fourth auxiliary request adds in claim 1 several features taken from the description, for example the feature that the "crunching together" involves "name matching through fuzzy logic and other matching algorithms". The Board considers that the complexity of this feature alone is such that it cannot reasonably be dealt with in oral proceedings. Therefore, the Board holds the forth auxiliary request inadmissible.

Order

For these reasons it is decided that:

The appeal is dismissed.

The Registrar: The Chairman:

T. Buschek W. Chandler

Decision electronically authenticated