Intellectual Property Office

BL O/193/17

20 April 2017

PATENTS ACT 1977

APPLICANT	Kube Partners Limited
ISSUE	Whether patent application number GB1019825.7 complies with Section 1(2)
HEARING OFFICER	Mrs S E Chalmers

DECISION

1 Patent application GB1019825.7, entitled “Private Information Storage System“, was filed on 23 November 2010 and published as GB2485783 A on 30 May 2012.

2 The examiner has maintained throughout the proceedings that the invention as claimed in this application is excluded from patentability as a computer program and a method of doing business as such under Section 1(2) of the Patents Act 1977. The applicant has not been able to overcome this objection, despite several rounds of amendment and argument, and has requested a decision be taken on the papers. The request presents arguments setting out a main request relating to the claims of the application currently on file and an auxiliary sets of claims to be considered if the main set is found to be unallowable.

3 In view of the findings on excluded matter the examiner did not complete the search. If I find in favour of the applicant, the application will be remitted back to the examiner to continue with the substantive examination.

The invention

4 The invention relates to a scheme for storage of private information on a cloud computing platform without contravention of territorial privacy laws. It provides a method of anonymising a database of personal data by determining a deviation of each data item in data records relative to reference data items in a reference record selected according to its similarity to the data record. Deviation identifiers are assigned to each of the determined deviations in said data records to anonymise the data items. A translation table is generated mapping the data items to the deviation identifiers. The anonymised data records are then stored remotely such as on a data server of a cloud computing platform without contravening national data protection laws. The translation table is stored locally under conventional controls for personal data so as to ensure compliance with privacy laws.

5 The main request concerns the set of claims filed on 7 June 2016 which has three independent claims: claim 1 to a method of anonymising a database of personal data; claim 18 to a method of decoding a database of personal data anonymised according to the method of claim 1; and claim 32 to a database management system for decoding a subset of data items in data records from a distributed database. These claims are reproduced in an appendix to this decision. Claim 1 reads:

> *1. A method of anonymising a database of personal data, the database comprising a plurality of data records, each data record comprising a plurality of data items, the method to reduce data volume of said database of personal data stored on a local memory of a computer-implemented database management system by limiting storage of data on said local memory to translation tables, to thereby reduce computing power of said computer, the method comprising;*
>
> *for a subset of said data items in said data records, determining a deviation of each of said data items in said data records relative to reference data items in a plurality of reference records,*
>
> *wherein one of said plurality of reference records is selected for each one of said data items or subset of data items dependent on a similarity of a said data record to said reference records, wherein determining said similarity comprises:*
>
> *categorising said data items in said data records into a plurality of pools based on classification profiles defined by said reference records, wherein a data item similarity of data items in a said pool is above a threshold; and*
>
> *comparing calculated perturbation profiles of one or more of said data items in a said pool with one or more of said reference data items of said reference records,*
>
> *wherein each of said data items in said data records has a corresponding said reference data item in a said selected reference record according to a said classification profile to determine a said deviation of a said data item relative to a said reference data item in a said selected reference record,*
>
> *assigning deviation identifiers to each of said determined deviations in said data records to identify a said data item being recorded as a said determined deviation to a said reference data item and to anonymise said data items in said subset of data items in said data records;*
>
> *generating a said translation table mapping said data items in said subset and said determined deviations to said deviation identifiers;*
>
> *storing said generated translation table on said local memory; and*
>
> *storing said deviation identifiers defining said anonymised data items for said data records remotely to said generated translation table stored on said local memory,*
>
> *wherein a first data size of said generated translation table stored on said local memory is smaller than a second data size of said deviation identifiers to thereby reduce said data volume of said database of personal data stored on said local memory to reduce computing power of said computer.*

6 The auxiliary claim set filed on 10 October 2016 also includes three independent claims corresponding to those of the main request but with references to anonymising removed: claim 1 to a method of operating a database; claim 18 to a method of decoding a database of personal data stored according to the method of claim 1; and claim 32 to a database management system for decoding a subset of

data items in data records from a distributed database. These claims are reproduced in the appendix to this decision.

The law

7 The section of the Act concerning inventions excluded from patentability is Section 1(2), which reads:

> *"1(2) It is hereby declared that the following (among other things) are not inventions for the purposes of this Act, that is to say, anything which consists of –*
> *(a)...*
> *(b)...*
> *(c) a scheme, rule or method for performing a mental act, playing a game or doing business or a program for a computer;*
> *(d)...*
> *but the foregoing provision shall prevent anything from being treated as an invention for the purposes of this Act only to the extent that a patent or application for a patent relates to that thing as such."*

8 In order to decide whether an invention relates to subject matter excluded by Section 1(2), the Court of Appeal has said that the issue must be decided by answering the question of whether the invention reveals a technical contribution to the state of the art. The Court of Appeal in *Aerotel/Macrossan*[1] set out the following four-step approach to help decide the issue:

> *1) Properly construe the claim;*
> *2) Identify the actual (or alleged) contribution;*
> *3) Ask whether it falls solely within the excluded subject matter;*
> *4) Check whether the actual or alleged contribution is actually technical in nature.*

9 The operation of the approach is explained at paragraphs 40-48 of the judgment. Paragraph 43 confirms that identification of the contribution is essentially a matter of determining what it is the inventor has really added to human knowledge, and involves looking at substance, not form. Paragraph 47 adds that a contribution which consists solely of excluded matter will not count as a technical contribution.

10 The case law on computer implemented inventions has been further elaborated in *AT&T/CVON*[2] which provided five helpful signposts to apply when considering whether a computer program makes a relevant technical contribution. In *HTC v Apple*[3], Lewison LJ reconsidered the fourth of these signposts and felt that it had been expressed too restrictively. The signposts are:

> *i) whether the claimed technical effect has a technical effect on a process which is carried on outside the computer;*

---

[1] Aerotel Ltd v Telco Holdings Ltd (and others) and Macrossan's Application [2006] EWCA Civ 1371
[2] AT&T Knowledge Ventures LP and CVON Innovations Limited v Comptroller General of Patents [2009] EWHC 343
[3] HTC v Apple [2013] EWCA Civ 451

ii) whether the claimed technical effect operates at the level of the architecture of the computer; that is to say whether the effect is produced irrespective of the data being processed or the applications being run;

iii) whether the claimed technical effect results in the computer being made to operate in a new way;

iv) whether the program make the computer a better computer in the sense of running more efficiently and effectively as a computer; and

v) whether the perceived problem is overcome by the claimed invention as opposed to merely being circumvented.

11 The examiner also refers to a number of hearing officers' decisions in support of his argument including *Oka-Bi*[4].

Application of the Aerotel approach – main request

Step 1: Properly construe the claim

12 The agent and the examiner agree the construction of claim 1 is straightforward and poses no problems. Claim 1 concerns a method to anonymise a database of personal data which is applied to a subset of data items of data records to assign deviation identifiers to deviations determined between each of the data items and a corresponding reference data item of one of a plurality of reference records selected based on similarity to the data record and reference record; storing the deviation identifiers remotely and a translation table mapping the deviations to the deviation identifiers locally.

Step 2: Identify the actual (or alleged) contribution

13 In their letter of 7 June 2016 the agent proposes that the contribution is to "*anonymising a database and reducing data volume stored on a local memory of a computer-implemented database management system by limiting storage of data on the local memory to translation tables*". They elaborate that this allows "…*for enhancing privacy while reducing data volume of data stored on a local memory*". The examiner agrees with this characterisation in his report of 8 August 2016 but in my view it is overly broad as it omits reference to (i) a subset of data, (ii) deviation, (iii) reference records and (iv) remote storage all of which are essential elements of what has been added to the stock of human knowledge.

14 The agent also asserts that the invention entails a "…*new hardware structure…"* but does not elaborate how. The examiner disagrees and so do I as the typical system used to implement the embodiments shown in figure 3 is conventional.

15 So what the inventor has added to the stock of human knowledge? In my view, the contribution relates to a scheme involving anonymising a subset of data in a database for storage on a remote computing platform using deviation identifiers and limiting storage of data on the local memory to translation tables, mapping those identifiers to deviations between the personal information and reference data.

[4] BL O/402/13

Steps 3 and 4: Ask whether it the contribution falls solely within the excluded subject matter and whether it is technical

16 The examiner has argued that the invention is no more than computer program and a business method as such; I will address the question of whether the invention is a computer program first.

Computer program

17 There is no doubt in my mind that the contribution requires a computer program for its implementation. However, that does not mean that it should be immediately excluded as a computer program as such. What matters is whether or not the program provides a technical contribution. The examiner concludes that the program does not provide a technical contribution as it does not, in his opinion, satisfy any of the *AT&T/CVON* "signposts". The applicant disagrees. I will first consider each signpost in turn.

> *i) whether the claimed technical effect has a technical effect on a process which is carried on outside the computer;*

18 In the letter of 7 June 2016 the agent proposes that "*Advantages of the claimed method over the state of the art are twofold, and relate to anonymising data in a novel way with enhanced security provisions and reducing data volume stored on a local memory*". The agent goes further in their letter of 10 October 2016 alleging that "…*the architecture provides a substantial saving in terms of both overall storage required and, importantly, in terms of power consumption particularly at the server side*"; from this they conclude that signpost (i) is also met because there is a reduction in greenhouse gas.

19 The application does not set out to reduce greenhouse gas and does not substantiate the assertion that storage and power is also saved in the cloud. If there is a reduction in power consumption and production of greenhouse gases then this is incidental to the contribution which is directed to a database and any effect of which is within the computers. The first signpost does not assist the applicant.

> *ii) whether the claimed technical effect operates at the level of the architecture of the computer; that is to say whether the effect is produced irrespective of the data being processed or the applications being run;*

20 In their letter of 10 October 2016 the agent also asserts that *"…the invention can be applied generally to any type of data"* and that it operates at the level of architecture of the computer so signpost (ii) is met. The examiner disagrees.

21 The contribution is applicable to any data stored in the database for which a deviation can be determined from a corresponding reference data item. It is not applicable to any data outside of the database and is only applicable to a subset of data in the database, specifically not names or addresses in the described embodiment. The invention concerns a subset of data in a database program; it does not operate at the level of the architecture and the second signpost does not point to there being a technical effect.

iii) whether the claimed technical effect results in the computer being made to operate in a new way;

22 The agent suggests that signpost (iii) points to a technical effect because the "*database management computer system operates in an entirely new way*". I agree that the database management computer system operates in a new way but the computers involved are the same as they were before; it is the database program which is new. The third signpost is not applicable.

iv) whether the program makes the computer a better computer in the sense of running more efficiently and effectively as a computer; and

23 The reduction in data volume mentioned above is said to advantageously reduce "*computing power of the computer, making the computer run more efficiently*" in the agent's letter of 7 June 2016 and therefore at least signpost (iv) is met. This point is expanded in the letter of 10 October 2016 by proposing the contribution provides for reductions in local storage & processing, overall storage, server side processing and increases in security, speed and reliability. The application states:

"….local storage and computer power is reduced, with the increased storage demands pushed out to remote cloud computing platforms."

24 No benchmark for this reduction is given and the application does not substantiate some of the proposals. I do not find the arguments regarding the reduction in storage or power requirements at the local client persuasive. The program achieves these reductions by moving data to servers in the cloud; this strikes me as an entirely conventional way to reduce storage requirements at the local client. In any event any improvement is in comparison with running other database programs and is not a better computer per se.

25 The letter of 7 June 2016 asserts that "*it is well-established practice by the European Patent Office to consider encryption/decryption techniques as patentable per se*" and that therefore the application is allowable. No evidence is provided in support of this assertion and in any event whilst EPO practice may be persuasive, it is not binding on me. The specification states that the described approach "…*differs to encryption techniques…*" on page 3, lines 3 to 5 for example. I would agree that encryption, and corresponding decryption, techniques may be patentable but I would not go so far as to say that any claim which includes an element of data encoding which can be characterised as, or likened to, encryption will be patentable.

26 The agent states, in their letter of 10 October 2016, that the examiner's approach to signpost (iv) has been too restrictive with reference to the comments of Lewison LJ and Kitchen LJ in *HTC v Apple*, particularly in paragraphs 150, 151 and 153, which in turn refer to *Gemstar*[5], *Halliburton*[6] and *IBM*[7]. These paragraphs, and the decisions to which they refer, emphasise the need to consider what task the program (or programmed computer) performs. When a program solves a technical problem relating to the running of computers generally there is scope for a patent; otherwise

[5] Gemstar-TV Guide International Inc v Virgin Media Ltd [2010] RPC 10
[6] Halliburton Energy Services Inc's Applications [2012] RPC 129
[7] IBM/Data processing network T0006/83 [1990]

there is nothing more than the running of a computer program. The database program which affects the contribution does not make the computers involved better in the sense of being more efficient and effective computers; it is a database program which is better than another unspecified database program but that does not point to there being a technical effect. The fourth signpost does not suggest there is a technical effect.

v) whether the perceived problem is overcome by the claimed invention as opposed to merely being circumvented

27 In the letter of 10 October 2016 the agent proposes that the problems identified above are addressed rather than circumvented so signpost (v) applies. I do not agree. The contribution circumvents any local storage problems by moving the bulk of the data elsewhere. The application is predicated on problems identified with existing cloud-based solutions to storage of databases of personal information which are said to contravene territorial privacy laws and increase risks associated with password cracking or leakage of the passwords or encryption keys. Adherence to privacy laws is not a technical problem. The need for data security is addressed by applying a conventional cryptographic technique to deviations of a subset of data in the database. This could be said to circumvent the security problem by changing what is secured rather than addressing it. If I were to take the view that the scheme addresses the problem, it is only applicable to a subset of data in the database, specifically not names or addresses in the described embodiment.

28 The perceived problems and claimed solution are with databases and whilst the problem of data security extends beyond this field the solution does not. The fifth signpost does not point to there being a technical contribution.

29 Finally I have taken a step back from the arguments and looked at the claim overall. In substance the claim relates to software in the form of a database application running on conventional computers. The task the software performs involves determining deviations of a subset of the data in the database and obfuscating those deviations by translating them to an identifier before storing them on a remote computer. The context in which this arises is that of securely storing a database with personal information on remote computers in a way which does not contravene privacy laws. So does the program provide a technical contribution? I do not think so.

30 In conclusion I find that the invention defined by claim 1 of the main request is excluded by Section 1(2) of the Act as a computer program as such. The agent hasn't advanced any arguments to suggest that a different conclusion might apply to the other independent claims 18 and 32; these claims are to an equivalent decoding method and system and I find that the same conclusion applies. None of the dependent claims introduce a feature which is more than a computer program as such.

Business method

31 Having concluded that the invention concerns a computer program as such, I make no finding on the question of whether the invention is also a business method as such.

Auxiliary claims

32 The auxiliary claims introduce explicit references to the client and server sides of the system to the independent claims and remove explicit reference to anonymising the data. Performing the same analysis on the auxiliary claims leads to the same result as the additions and removal alter the form of the claims but not the substance of the invention.

Decision

33 I have found that the contribution made by the invention defined by the claims of the main request and the auxiliary request falls solely in matter excluded from patentability by virtue of Section 1(2) of the Act, namely as a program for a computer as such. Furthermore, I have read the specification carefully and I can see nothing that could be reasonably expected to form the basis of a valid claim. I therefore refuse this application under section 18(3).

Appeal

34 Any appeal must be lodged within 28 days after the date of this decision.

Mrs S E Chalmers

Deputy Director, acting for the Comptroller

Appendix

Main request independent claims

1. A method of anonymising a database of personal data, the database comprising a plurality of data records, each data record comprising a plurality of data items, the method to reduce data volume of said database of personal data stored on a local memory of a computer-implemented database management system by limiting storage of data on said local memory to translation tables, to thereby reduce computing power of said computer, the method comprising;

for a subset of said data items in said data records, determining a deviation of each of said data items in said data records relative to reference data items in a plurality of reference records,

wherein one of said plurality of reference records is selected for each one of said data items or subset of data items dependent on a similarity of a said data record to said reference records, wherein determining said similarity comprises:

categorising said data items in said data records into a plurality of pools based on classification profiles defined by said reference records, wherein a data item similarity of data items in a said pool is above a threshold; and

comparing calculated perturbation profiles of one or more of said data items in a said pool with one or more of said reference data items of said reference records,

wherein each of said data items in said data records has a corresponding said reference data item in a said selected reference record according to a said classification profile to determine a said deviation of a said data item relative to a said reference data item in a said selected reference record,

assigning deviation identifiers to each of said determined deviations in said data records to identify a said data item being recorded as a said determined deviation to a said reference data item and to anonymise said data items in said subset of data items in said data records;

generating a said translation table mapping said data items in said subset and said determined deviations to said deviation identifiers;

storing said generated translation table on said local memory; and

storing said deviation identifiers defining said anonymised data items for said data records remotely to said generated translation table stored on said local memory,

wherein a first data size of said generated translation table stored on said local memory is smaller than a second data size of said deviation identifiers to thereby reduce said data volume of said database of personal data stored on said local memory to reduce computing power of said computer.

18. A method of decoding a database of personal data anonymised according to the method of claim 1, the database comprising a plurality of data records, each data record comprising a plurality of data items, the method comprising:

retrieving deviation identifiers for a subset of said data items in remote data record from remote storage, wherein said deviation identifiers define anonymised deviations for each of said data items in said data records relative to reference data items in a plurality of reference records;

retrieving a translation table from storage, wherein said storage is distinct to said deviation identifiers, and wherein said translation table defines a mapping of

said data items in said subset and said determined deviations with said deviation identifiers;

processing said deviation identifiers using said selected reference record and said translation table to decode said database of personal data;

wherein said processing comprises performing a reverse mapping of said deviation identifiers to deviations of each of said data items in said subset of data items and, using said selected reference record and said deviations, determining said data items in said subset of data items.

32. A database management system for decoding a subset of data items in data records from a distributed database, said database management system comprising a distributed storage system, wherein said distributed storage system comprises a memory which stores said distributed database,

the distributed database comprising:

a plurality of deviation identifiers storing anonymised references to deviations in a subset of data items in data records, wherein said deviation identifiers define anonymised deviations of said subset of data items relative to reference data items in a plurality of reference records;

the database management system further comprising a processor configured to select one of said plurality of reference records for each one of said data items or subset of data items dependent on a similarity of a said data record to said reference records, wherein said processor is further configured to determine said similarity, wherein said determination of said similarity comprises:

categorising said data items in said data records into a plurality of pools based on classification profiles defined by said reference records, wherein a data item similarity of data items in a said pool is above a threshold; and

comparing calculated perturbation profiles of one or more of said data items in a said pool with one or more of said reference data items of said reference records,

and wherein each of said data items in said data records has a corresponding said reference data item in a said selected reference record according to a said classification profile to determine a said deviation of a said data item relative to a said reference data item in a said selected reference record;

the database management system further comprising:

a data dictionary comprising a translation table storing mappings from said subset of data items and said determine deviations to said deviation identifiers;

a transaction engine to retrieve one or more of said plurality of deviation identifiers from remote storage, to retrieve said translation table, and to retrieve said selected reference record, wherein said storage of said deviation identifiers is distinct to said translation table;

a query engine to process said deviation identifiers using said selected reference record and said translation table to decode said distributed database and reconstitute said subset of data items in said data records,

wherein said decoding comprises performing a reverse mapping of said deviation identifiers to deviations of each of said data items in said subset of data items and, using said selected reference record and said deviations, determining said data items in said subset of data items;

wherein said memory comprises a first, local memory which is limited to storing one or more translation tables, and wherein a first data size of a said translation table or translation tables stored on said local memory is smaller than a

second data size of said deviation identifiers to thereby reduce a data volume of data stored on said local memory to reduce computing power of a computer on which said database management system is implemented.

Auxiliary request independent claims

1. A method of operating a database held on a computer-implemented database management system comprising at least one client machine having local memory and at least one remote data server, the database comprising a plurality of data records, each data record comprising a plurality of data items,

wherein the method reduces data volume of said database stored on said local memory by limiting storage of data in said local memory to translation tables, the method comprising:

for a subset of said data items in said data records, determining a deviation of each of said data items in said data records relative to reference data items in a plurality of reference records,

wherein one of said plurality of reference records is selected for each one of said data items or subset of data items dependent on a similarity of a said data record to said reference records, wherein determining said similarity comprises:

categorising said data items in said data records into a plurality of pools based on classification profiles defined by said reference records, wherein a data item similarity of data items in a said pool is above a threshold; and

comparing calculated perturbation profiles of one or more of said data items in a said pool with one or more of said reference data items of said reference records,

wherein each of said data items in said data records has a corresponding said reference data item in a said selected reference record according to a said classification profile to determine a said deviation of a said data item relative to a said reference data item in a said selected reference record,

assigning deviation identifiers to each of said determined deviations in said data records to identify a said data item being recorded as a said determined deviation to a said reference data item;

generating a said translation table mapping said data items in said subset and said determined deviations to said deviation identifiers;

storing said generated translation table in said local memory; and

storing said deviation identifiers defining said data items for said data records remotely to said generated translation table stored on said local memory,

wherein a first data size of said generated translation table stored on said local memory is smaller than a second data size of said deviation identifiers thereby reducing said data volume of said database stored in said local memory.

18. A method of decoding a database of personal data stored according to the method of claim 1, the database comprising a plurality of data records, each data record comprising a plurality of data items, the method comprising:

retrieving deviation identifiers for a subset of said data items in remote data record from remote storage, wherein said deviation identifiers define deviations for each of said data items in said data records relative to reference data items in a plurality of reference records;

retrieving a translation table from storage, wherein said storage is distinct to said deviation identifiers, and wherein said translation table defines a mapping of

said data items in said subset and said determined deviations with said deviation identifiers;

processing said deviation identifiers using said selected reference record and said translation table to decode said database of personal data;

wherein said processing comprises performing a reverse mapping of said deviation identifiers to deviations of each of said data items in said subset of data items and, using said selected reference record and said deviations, determining said data items in said subset of data items.

32. A database management system for decoding a subset of data items in data records from a distributed database, said database management system comprising a distributed storage system, wherein said distributed storage system stores said distributed database, the distributed database comprising:

at least one client machine having local memory and at least one remote data server;

a plurality of deviation identifiers storing references to deviations in a subset of data items in data records, wherein said deviation identifiers define deviations of said subset of data items relative to reference data items in a plurality of reference records;

the database management system further comprising a processor configured to select one of said plurality of reference records for each one of said data items or subset of data items dependent on a similarity of a said data record to said reference records, wherein said processor is further configured to determine said similarity, wherein said determination of said similarity comprises:

categorising said data items in said data records into a plurality of pools based on classification profiles defined by said reference records, wherein a data item similarity of data items in a said pool is above a threshold; and

comparing calculated perturbation profiles of one or more of said data items in a said pool with one or more of said reference data items of said reference records,

and wherein each of said data items in said data records has a corresponding said reference data item in a said selected reference record according to a said classification profile to determine a said deviation of a said data item relative to a said reference data item in a said selected reference record;

the database management system further comprising:

a data dictionary comprising a translation table storing mappings from said subset of data items and said determine deviations to said deviation identifiers;

a transaction engine to retrieve one or more of said plurality of deviation identifiers from remote storage, to retrieve said translation table, and to retrieve said selected reference record, wherein said storage of said deviation identifiers is distinct to said translation table;

a query engine to process said deviation identifiers using said selected reference record and said translation table to decode said distributed database and reconstitute said subset of data items in said data records,

wherein said decoding comprises performing a reverse mapping of said deviation identifiers to deviations of each of said data items in said subset of data items and, using said selected reference record and said deviations, determining said data items in said subset of data items;

wherein said local memory is limited to storing one or more translation tables, and wherein a first data size of a said translation table or translation tables stored in

said local memory is smaller than a second data size of said deviation identifiers to thereby reduce a data volume of data stored in said local memory.