Intellectual Property Office

BL O/402/13

9 October 2013.

PATENTS ACT 1977

APPLICANT	Oka-bi Limited
ISSUE	Whether patent application number GB1201857.8 complies with Section 1(2)
HEARING OFFICER	Peter Slater

DECISION

1 Patent application GB1201857.8, entitled “Pseudonymisation/De-pseudonymisation of data“, was filed on 2 February 2012 claiming an earliest priority date of 2 February 2011. It was published as GB2488024 A on 15 August 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 as such and a method of doing business under section 1(2) of the Patents Act 1977. The applicant has not been able to overcome this objection, despite amendments to the application, and has requested a decision be taken on the papers.

The Invention

3 The potential loss of personal data represents a high degree of risk to organisations which have a legal responsibility under the Data Protection Act to protect the confidentiality of individuals such as patients in the National Health Service.

4 The invention relates to a method by which sensitive personal data can be anonymised (“pseudonymised”) to create a database which is less likely to be compromised should the data fall into the wrong hands. The process can then be reversed (“de-pseudonymised”) and the data restored to its original form for subsequent display to one or more users based upon their access rights. Various aspects of the data, for example, names and addresses may need to be kept confidential. The amount of confidential data a person can access may vary according to the person and their access rights. For example, a general practitioner may be allowed to view all the data relating a patient's file, including the patient's name and address etc., whereas a statistician who is performing statistical analysis on the dataset to identify potential trends would only be allowed to view certain aspects of the data e.g. the patient’s date of birth and gender.

5 Protection of personal data can be performed by conventional encryption techniques. However, this has been found to be computationally intensive and can often require the entire encrypted data entry to be unencrypted before it can be viewed. Furthermore, the problem becomes even more complicated when two or more users have different access rights to the data resulting in further inefficiencies in the processing of the data.

6 The present invention provides a computer implemented system for obscuring user selected attributes within a dataset, for example, a patient's name and address, details of their medical condition, their gender and date of birth etc. by replacing the data content with a computer generated pseudonym to create a pseudonymised database. Metadata is used to describe the various entries and/or columns within the dataset. This metadata is presented in the form of a list of all the data items and/or columns contained within the original dataset to an end user via an interface configured to allow the user to select those items of data which are to be stored in the new database along with those which are to be pseudonymised by having their content replaced by the computer generated pseudonyms. Two types of pseudonymisation are envisaged. A repeat pseudonymisation and a unique pseudonymisation. Where an item of data is unique to a particular patient, for example their NHS number, it is desirable to use the same repeating pseudonym to represent multiple occurrences of the same number. As I understand it, this enables data relating to a particular patient to be collated on the basis of the repeating pseudonyms without the need to restore the original dataset thus reducing the computational burden on the system. Computer executable code in the form of Structured Query Language (SQL) scripts are then created and executed in order to generate the pseudonymised database. Users with access to the database may then request permission to view the original de-pseudonymised and depending upon their access rights they will either be presented with the original data or a subset thereof pseudonymised.

7 The most recent set of claims was filed on 26 April 2013. There are 15 claims in total, of which 4 are independent. The independent claims read as follows:

> *1. A computer implemented system for enabling the pseudonymisation of a plurality of database entries, the database entries comprising attributes, the system comprising:*
> *a metadata file describing the attributes of the database entries;*
> *a user interface configured to display the metadata to a user, enabling the user to select one or more attributes from the metadata; and*
> *a code generator configured to:*
> *generate a first set of code that when executed and applied to database entries in a data source pseudonymises the attributes of that database so that repeated occurrences of a particular entry/value are given different (preferably unrelated) pseudonyms; and*
> *generate a second set of code that when executed pseudonymises attributes in that database so that repeated occurrences of a particular entry are given the same pseudonym or traceably related pseudonyms;*
> *the system further configured to enable the user to select a first set of attributes from the metadata of which it is desired to be pseudonymised in a manner so that repeated occurrences of a particular entry/value are given*

different (preferably unrelated) pseudonyms and to select a second set of attributes from the metadata of which it is desired to be pseudonymised in a manner so that repeated occurrences of a particular entry are given the same pseudonym or traceably related pseudonyms
and in response to selection of the first set of attributes the system is configure to execute and apply the first set of code to the database entries in a data source matching the metadata, thereby pseudonymising the first set of data attributes in the database entries such that repeated occurrences of a particular entry/value are given different (preferably unrelated) pseudonyms; and in response to selection of the second set of attributes to execute and apply the second set of code thereby pseudonymising the second set of data attributes in the database entries so that repeated occurrences of a particular entry are given the same pseudonym or traceably related pseudonyms.

9. A computer implemented method for enabling the pseudonymisation of a plurality of database entries, the database entries comprising attributes the method comprising:
describing the attributes of the database entries to create metadata regarding the database entries
providing to a user the metadata and enabling the user to select one or more attributes from the metadata;
enabling the user to select a first set of attributes from the metadata of which it is desired to be pseudonymised in a manner so that repeated occurrences of a particular entry/value are given different (preferably unrelated) pseudonyms and to select a second set
of attributes from the metadata of which it is desired to be pseudonymised in a manner so that repeated occurrences of a particular entry are given the same pseudonym or traceably related pseudonyms; and
generating a first set (or procedure) of code, the first set of code when executed and applied to database entries in a data source matching the metadata, pseudonymises the first set of data attributes in the database entries so that repeated occurrences of a particular entry/value are given different (preferably unrelated) pseudonyms and generating a second
set of procedure code, the second set of code when executed pseudonymises the second set of data attributes in the database entries so that repeated occurrences of a particular entry are given the same pseudonym or traceably related pseudonyms;
executing and applying the first set of code to the attributes of the database in response to selection of the first set of attributes; and
executing and applying the second set of code to the attributes of the database in response to selection of the second set of attributes.

10. A computer implemented method for accessing and enabling the pseudonymisation and de- pseudonymisation of a plurality of database entries, the database entries comprising one or more attributes the method comprising:
receiving a selection of a first set of one or more attributes to be pseudonymised in a manner so that repeated occurrences of a particular entry/value are given different (preferably unrelated) pseudonyms and to

select a second set of one or more attributes to be pseudonymised in a manner so that repeated occurrences of a particular entry/value are given the same pseudonym or traceably related pseudonyms;
pseudonymising the first and second set of attributes according to the user selection to create a pseudonymised data set;
receiving a request from a user to access the psuedonymised data set, and to view one or more pseudonymised attributes in original de-pseudonymised form, the request containing user authentication data;
determining using the user authentication data what data attributes, if any, the user is permitted to view in original de-pseudonymised form; and
de-pseudonymising the pseudonymised data attributes which the user is permitted to view.

14. A computer system for enabling pseudonymisation and de-pseudonymisation of database entries the system comprising a processor and a memory, the system configured to:
receive metadata containing names of columns/attributes of a data in a data source;
provide a user with a list of at least some of the names of columns/attributes from the metadata;
allow a user to select a first set of columns/attributes from the list the entries/values of which it is desired to be pseudonymised in a manner so that repeated occurrences of a particular entry/value are given different (preferably unrelated) pseudonyms and to select a second set of columns/attributes from the list the entries/values of which it is desired to be pseudonymised in a manner so that repeated occurrences of a particular entry/value are given the same pseudonym or traceably related pseudonyms; and
to use the selections to generate a first code procedure, which code when executed and applied to data in a data source matching the metadata will pseudonymise entries/values under the columns/attributes selected to be in the first set such that repeated occurrences of a particular entry/value are given different (preferably unrelated) pseudonyms, and will pseudonymise entries/values under the columns/attributes selected to be in the second set such that repeated occurrences of a particular entry/value are given the same pseudonym or traceably related pseudonyms and generate a second code procedure which when executed and applied to the results of the application of the first code procedure and/or the data in the data source populates a database for access by users with pseudonyms corresponding to columns/attributes selected as part of the first or second set and preferably also with entries/values from the data source corresponding to at least one column/attribute not selected by the user as part if the first or second set.

The Law

8 The examiner has raised an objection under section 1(2)(c) of the Patents Act 1977 that the invention is not patentable because it relates to a program for a computer as such; the relevant provisions of this section of the Act are shown in bold below:

> ***1(2) It is hereby declared that the following (amongst other things) are not inventions for the purpose of the 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 provisions shall prevent anything from being treated as an invention for the purposes of the Act only to the extent that a patent or application for a patent relates to that thing as such.

9 As explained in the notice published by the UK Intellectual Property Office on 8 December 2008[1],, the starting point for determining whether an invention falls within the exclusions of section 1(2) is the judgment of the Court of Appeal in *Aerotel/Macrossan*[2].

10 The interpretation of section 1(2) has been considered by the Court of Appeal in *Symbian Ltd's Application*[3]. *Symbian* arose under the computer program exclusion, but as with its previous decision in *Aerotel*, the Court gave general guidance on section 1(2). Although the Court approached the question of excluded matter primarily on the basis of whether there was a technical contribution, it nevertheless (at paragraph 59) considered its conclusion in the light of the *Aerotel* approach. The Court was quite clear (see paragraphs 8-15) that the structured four-step approach to the question in *Aerotel* was never intended to be a new departure in domestic law; that it remained bound by its previous decisions, particularly *Merrill Lynch*[4] which rested on whether the contribution was technical; and that any differences in the two approaches should affect neither the applicable principles nor the outcome in any particular case. But the *Symbian* judgment does make it clear, that in deciding whether an invention is excluded, one must ask does it make a technical contribution? If it does then it is not excluded.

11 Subject to the clarification provided by *Symbian*, it is therefore still appropriate for me, to proceed on the basis of the four-step approach explained at paragraphs 40-48 of *Aerotel/Macrossan* namely:

1) Properly construe the claim

2) Identify the actual contribution (although at the application stage this might have to be the alleged contribution).

3) Ask whether it falls solely within the excluded matter, which (see paragraph 45) is merely an expression of the "as such" qualification of section 1(2).

4) If the third step has not covered it, check whether the actual or alleged contribution is actually technical.

---

[1] http://www.ipo.gov.uk/pro-types/pro-patent/p-law/p-pn/p-pn-computer.htm
[2] *Aerotel Ltd v Telco Holdings Ltd* and *Macrossan's Application* [2006] EWCA Civ 1371; [2007] RPC 7
[3] *Symbian Ltd v Comptroller-General of Patents*, [2009] RPC 1
[4] *Merrill Lynch's Application* [1989] RPC 561

12 The operation of this test is explained at paragraphs 40-48 of the decision. 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 46 explains that the fourth step of checking whether the contribution is technical may not be necessary because the third step should have covered the point.

Construing the claims

13 The first step of the test is to construe the claims. I do not think this presents any real problems since both the applicant and the examiner appear to agree as to the meaning of the claims.

Identify the actual contribution

14 For the second step, it is necessary to identify the contribution made by the invention. Paragraph 43 of *Aerotel/Macrossan* explains that this is to be determined by asking what it is - as a matter of substance not form - that the invention has really added to the stock of human knowledge having regard to the problem to be solved, how the invention works and what its advantages are.

15 The examiner considers the contribution to relate to a computer program for the "processing of data entries so that their content is obscured when displayed to the viewer, based on the selection of specific types of entry by a user according to the level of data security required. It has the advantage that personally identifiable data is obscured such that it cannot be inferred from the other data provided, or easily compromised by third parties."

16 The applicant on the other hand considers the contribution to lie in an improved method of pseudonymisation of data and/or an improved method of creating a confidential database in which users can view data and collate related data entries using the repeat pseudonyms without the need to restore the original data set by de-pseudonymisation of the entries. Furthermore, users with access to the pseudonymised data cannot easily infer relationships between the data and thereby de-pseudonymise entries without access to the key making the data less prone to compromise.

17 There appears to be some degree of commonality between the examiner's assessment of the contribution and that of the applicant, both have identified key aspects of the invention which when combined arrive at what I consider to be the true contribution. In my opinion, the invention provides a computer implemented method of pseudonymisation of data in which user selected data entries are obscured by replacing their content with computer generated pseudonyms. Users can then view the data either in its original form or its pseudonymised form depending upon their level of access rights, and are able to collate related data entries using the repeat pseudonyms without the need to restore the original data set by de-pseudonymisation of the entries which has the effect of reducing the computational burden on the system. This also has the advantage that users with access to the pseudonymised data cannot easily infer relationships between the data and thereby de-pseudonymise entries without access to the key making the data less prone to compromise. I do not think the contribution extends to a new method of

encryption in the traditional sense. Indeed, merely substituting the content of data entries with computer generated pseudonyms using known techniques available in existing Microsoft software and giving the user the opportunity to select from one of a number of known algorithms does not seem to suggest that the contribution includes a new method of encryption per se.

Does the contribution fall solely within excluded subject matter? Is the contribution technical in nature?

Computer program

18 There is no doubt in my mind that the contribution requires a computer program for its implementation. However, the mere fact that the invention is effected in software 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.

19 The examiner does not consider the contribution to be technical in nature as it does not, in his opinion satisfy, any of the *AT&T/Cvon*[5] "signposts" which were proposed by Lewison J as an aid in determining whether a computer program had a technical effect. He therefore concludes that the invention relates to a computer program which provides no technical contribution and as such is excluded.

20 The applicant's arguments are set out in some detail in their letters of the 26 April 2013 and 15 July 2013 respectively, and I do not intend to repeat them here in their entirety suffice to say that drawing on the judgment in *Halliburton*[6], the applicant argues that the contribution insofar as it relates to an improved method of pseudonymisation of data and/or an improved method of creating a confidential database is inherently technical in nature and thus does not fall solely within any one of the excluded categories.

21 So does the program provide a technical contribution? I do not think so. It is true to say that the invention provides a new method of pseudonymising data and that it improves the security of that data insofar as it makes it more difficult for users with access to the pseudonymised data to easily identify relationships between the various data entries. However, this is done by allowing the user to select which entries are to be pseudonymised and providing a facility whereby the content of recurring entries can be replaced by repeating pseudonyms and not by any technical means associated with the program or the computer on which it is running. It is important to note that users with restricted access to the data can still extract useful information from the pseudonymised dataset and collate related data entries using the repeat pseudonyms without the need to restore the original data which has the effect of reducing the computational burden on the system. This apparent improvement in the computational power of the system is achieved as a result of not having to process as much data and not by any specific technical means. I have already indicated that merely substituting the content of data entries with computer generated pseudonyms using known techniques available in existing Microsoft software and giving the user the opportunity to select from one of a number of known algorithms does not seem to suggest that the contribution includes a new method of

[5] AT&T Knowledge Ventures/Cvon Innovations v Comptroller of Patents [2009] EWHC 343 (Pat).
[6] Halliburton Energy Services Inc [2011] EWHC 2508 (Pat).

encryption per se nor does it convey the necessary technical contribution required to save the invention from exclusion.

Business method

22 The examiner has also argued that the selection by the user of data entries to be pseudonymised and the subsequent replacement of the content of those entries with computer generated pseudonyms in an attempt to obscure the personal nature of the data is considered to be a purely administrative act constituting a business method and is therefore excluded as such. However, having found the invention to be excluded as a computer program, I have no need to decide this issue.

Conclusion

23 In the light of my findings above, I conclude that the invention as claimed is excluded under section 1(2) because it relates to a computer program as such. Having read the specification I do not think that any saving amendment is possible. I therefore refuse the application under section 18(3).

Appeal

24 Any appeal must be lodged within 28 days

PETER SLATER

Deputy Director, acting for the Comptroller