Adopting a pseudonym can preserve privacy

Sensitive data can be protected at the same time as allowing users access to less critical elements by means of a technique called pseudonymisation

Alan Lawson

Few would argue with the idea that corporate data, typically relating to customers, business partners, and internal operations, has become one of the mostly highly prized assets any business can possess. The potential downside is that this data is equally valuable to the individuals and organisations that it relates to, and they are increasingly inclined to protect it.

It is now the case that any misuse or compromise of the privacy of this information carries severe penalties. Legislation, including the UK's Data Protection Act (and similarly stringent laws throughout the European Union), and the Health Information Portability and Accountability Act (HIPAA) in the US, now force data-holding organisations to manage their responsibilities in an accountable and auditable manner.

Because such laws are now being enforced, businesses are also increasingly finding themselves trapped between the extremes of protecting their data stores for commercial advantage and to minimise the risk of exposing themselves to legislation from consumers; and making those same stores accessible to dynamic business processes and audit investigations.

These conflicting demands seem to prove difficult to define in terms of policy, and even more so in terms of how best to deploy technology to serve corporate interests. Although the problem is clear – each organisation must utilise its data whilst demonstrably preserving its sanctity – the resolution is not, and solutions capable of meeting one set of demands often conflict with the other set of demands.

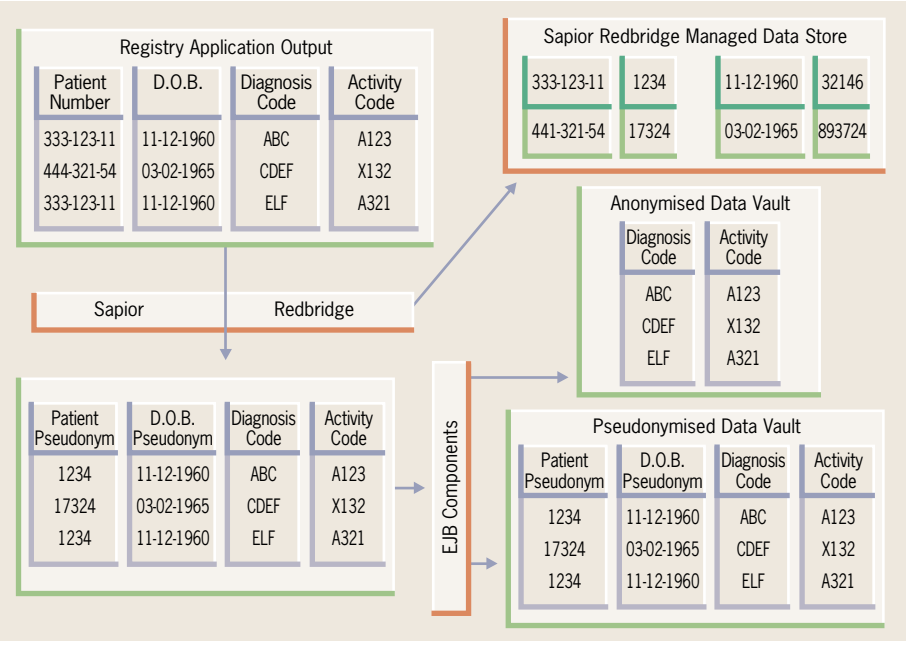

Figure 1: Medical records pseudonymised in a Sapior Redbridge system

Data protection is usually seen as a subset of security, and will usually be managed via technologies such as encryption and password protection. In one respect this is logical, as these approaches are designed to limit access to specific resources, and to validate the identities of users, which would seem to be ideal attributes for the protection of data.

However, these are all or nothing measures, in the sense that an individual either has full access to data or none at all – and therefore, there is nothing to actually protect the data from misuse by a trusted user, or one that has stolen access rights.

If we accept that limiting access to data does not fully answer the problem, then we must consider alternatives that satisfy the need to protect data, whilst still enabling value to be extracted from this resource. An interesting option for a privacy-oriented means by which data could be managed was suggested by the former UK Information Commissioner, Elizabeth France, who proposed the concept of pseudonymisation of data.

In simple terms, this process handles sensitive data by substituting critical data elements with pseudonyms. Information workers accessing data do not see the data itself, which is not directly accessible – only the specific elements of data relevant to an enquiry are returned to the user.

The result of the pseudonymisation process is that, although a user can still search data for relationships, that user cannot capture all the value of the data outside the exact context of the interaction, and cannot amend it in an unauthorised manner at all.

Copying pseudonymised data is similarly pointless, as the keys connecting the valuable links between the accessible pseudonym and the actual data itself are held elsewhere.

Pseudonymisation is also a good fit with a clearly defined security policy. Butler Group has consistently maintained the importance of beginning security with an ongoing risk management process that identifies the weak points that will inevitably exist in every organisational infrastructure, before applying appropriate technology at each point in turn.

Blanket protections will sometimes prove unnecessary where policy has been sufficiently well drafted, which will reduce the cost of deployment and simplify subsequent management. Password protection is a good example of this point. The majority of employees do not require full database access, but, as noted above, access is often an all or nothing affair. A security administrator is typically forced to maintain a password system for all an organisation's employees, instead of the

relative handful that can actually use the database in a truly valuable and strategic manner.

However, if the majority of employees can only access data that has been passed to them through a pseudonymisation process, rather than accessing the core data itself, the administrator has a far simpler task. The act of making the data anonymous and secured from amendment simplifies the demands of protecting the data; although every employee will be able to search for data relationships and capture valid results, only a strictly limited number of high-level users who genuinely require full access will need password administration on an ongoing basis.

This efficiency is only one aspect of an internal benefit to pseudonymisation. Although harder to quantify, the element of protecting data from internal misuse is also becoming increasingly important. Disgruntled or opportunistic employees can steal or amend corporate data for their own purposes or profit. For example, harvesting e-mail addresses to sell to mass marketers – this is a surprisingly lucrative trade, and it seems likely that spammers make as much profit from the sale of addresses as they do from customers selling through their services.

As noted earlier, companies need to be able to demonstrate (in a court of law if necessary) that they are taking precautions to protect data relating to their customers and trading partners, and pseudonymisation has much to recommend from this point of view. The concept of pseudonymisation can be applied to database operations by sufficiently skilled administrators, and research suggests that some of those aware of the advantages of the approach have opted for the build, rather than buy, approach.

The risk in so doing is that home grown solutions tend to possess too narrow a focus, and do not always scale well. Scalability is always a fundamental requirement if data is to be accessed at all times. It seems likely that some early experiments with data anonymity could have been hampered by poor scalability.

There are still only a handful of solutions that apply pseudonymisation to database operations. To a large extent, this is due to the still comparatively low profile of the approach – if end-users are not demanding the feature, few vendors will go to the trouble of developing and promoting it, after all. An interesting exception to this rule is Sapior Redbridge, which was originally designed as an Extract Transform and Load (ETL) tool. Sapior Redbridge is now sold as a security tool, with an ability to pseudonymise data in a manner that satisfies the legal issues referred to throughout this article. Butler Group believes that this is a shrewd positioning of the tool.

The product illustrates the concept of pseudonymisation in actual business usage very clearly indeed. Sapior Redbridge replaces sensitive fields, typically identifying data such as name and address, with pseudonyms, and maintains a separate secure data store of relationships between critical identifiers (name, address, date of birth) and pseudonyms. One of the clearest market opportunities for this product (and indeed any similar pseudonymisation tool) is in the medical and pharmaceutical arena, where legislative protections are being zealously enforced. Figure 1 (page 9) shows how Sapior Redbridge works.

The figure illustrates the use of patient number and date of birth as identifiers. These are likely to be core elements in medical records – common sense dictates that date of birth will often be referenced in searches for relationships with other factors. A medical company might choose to output the pseudonymised results into two data vaults, as illustrated, one containing fully anonymised data consisting of limited field results that are relevant to the enquiry being made.

The second vault contains fully pseudonymised field results. The diagram presumes that the source data is patient information being held by a medical company. This could be a pharmaceuticals company, or even a hospital, and in the latter case there will be strong legislative imperatives to protect the privacy of each patient's data. A registry application outputs three records of data, and in this instance the first and third records belong to one individual.

As can be seen, as a result of the pseudonymisation process, medical research can be carried out on these patient records without the researchers involved having direct access to confidential information at all – for example, pseudonymised results in a data vault can be used to accurately identify the exact level of incidence of any given disease, but will do so without also identifying the individuals involved to the researchers. The ability to correlate data remains, which is invaluable in identifying trends and relationships, but at no time is privacy or anonymity compromised.

Similar conditions apply to a financials company attempting to extract value from a Customer Relationship Management (CRM) system without running the risk of violating customer privacy – using pseudonymisation, data exercises can be carried out without an employee having access to any sensitive consumer information at all.

Although the financial and medical industries are the most obvious examples of sectors that could benefit from the deployment of pseudonymisation strategies and products, the approach should be considered by any customer-centric business. The fact of the matter is that legislation applies to every business acting as a data controller, and many organisations have found to their cost that they are poorly prepared to live up to that responsibility.

Pseudonymisation represents an interesting additional layer of security, and is one that can be added to an existing infrastructure with a minimum of difficulty and expense. The financial and medical industries are faced by steadily increasing legislative demands, and the implementation of pseudonymisation goes a long way towards enabling these shifting expectations to be met – it is therefore likely that in time this approach will find considerable favour with these industries, which tend to act as proving grounds for advanced technology.

At a glance

- Data relationships can be extracted from corporate databases without infringing upon legislative requirements to protect privacy using a process called pseudonymisation.
- Pseudonymisation can be carried out in-house with sufficient expertise, although it seems likely that home grown solutions may pose scalability problems when faced by extremely high volume exchanges of data.
- Products based on the pseudonymisation concept are beginning to appear, and make a useful extra layer of data security.
- Use of data pseudonymisation can reduce the maintenance costs of other security solutions, for example password systems, by reducing the numbers of users that need full access.
- Key areas where pseudonymisation should be considered include medicine/healthcare, the financials industry, and any customer-centric business.

Alan Lawson
Research Analyst
alan.lawson@butlergroup.com