Written by Alan Lawson, June 2003

TA000252SEC

Technology Infrastructure

Butler Group Subscription Services

Security
TECHNOLOGY AUDIT

Sapior
Sapior Redbridge 1.07

Abstract *The Sapior Redbridge product enables an interesting, although currently little-known, form of internal data protection to be undertaken. The process of pseudonymisation separates sensitive data elements (such as names, and similar identifying characteristics) to be replaced in output responses from a database by a pseudonym, or string of abstract characters. The result is that although the data store can be searched for valuable data relationships, the user involved will be unable to capture the whole of the data records being searched – only the specific points required for a task are presented, and these cannot be associated with any individuals. Personal data is therefore effectively protected using pseudonymisation, which has important legal connotations, whilst the general simplification of security maintenance leads to reductions in operational costs. Although undeniably a new and essentially unproven market, this approach is extremely promising and bears closer examination by data-centric businesses.*

KEY FINDINGS

✓ Innovative approach and technology that closely fits both security and data management needs.

✓ Pragmatic developmental emphasis upon design for enterprise operations.

i Data pseudonymisation was first defined by the former Information Commissioner, Elizabeth France.

✗ Sapior is a small company in a market that has yet to be fully developed.

Key: ✓ Product Strength ✗ Product Weakness *i* Point of Information

LOOK AHEAD

Sapior Redbridge will feature a fully standard compliant Java 2 Enterprise Edition (J2EE) Connector Architecture (JCA) from Q4, 2003. This will enable leading J2EE application servers to integrate pseudonymisation services with business logic, which will be a major advance for the scope of the product.

Butler Group

FUNCTIONALITY

Wherever corporate security has been deployed in an ineffective manner, opportunistic individuals can subvert a host of functions to capture unauthorised control. Data can be copied, amended, or even stolen outright, and, depending upon the skill of the attacker, it may even be the case that the company being attacked will remain unaware that anything has happened. When considering this scenario, it is normally the case that defences will be deployed with the expectation that an attack will originate outside the company's boundaries, somewhere beyond the firewall, but it is also possible that the attack will be launched inside the defensive perimeter, by a disgruntled or greedy employee.

The damage and cost of such an internal attack is typically far higher than corresponding external assaults, although the actual reporting of such incidents is relatively low. An employee with a malicious agenda will usually be far better informed about which items of proprietary information are valuable than a hacker on the outside of the firewall, and so the attack becomes a surgical strike, rather than the hit and miss approach that an external user must adopt in 'feeling' his or her way around an unfamiliar network.

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 context of the interaction, and cannot amend it in an unauthorised manner at all.

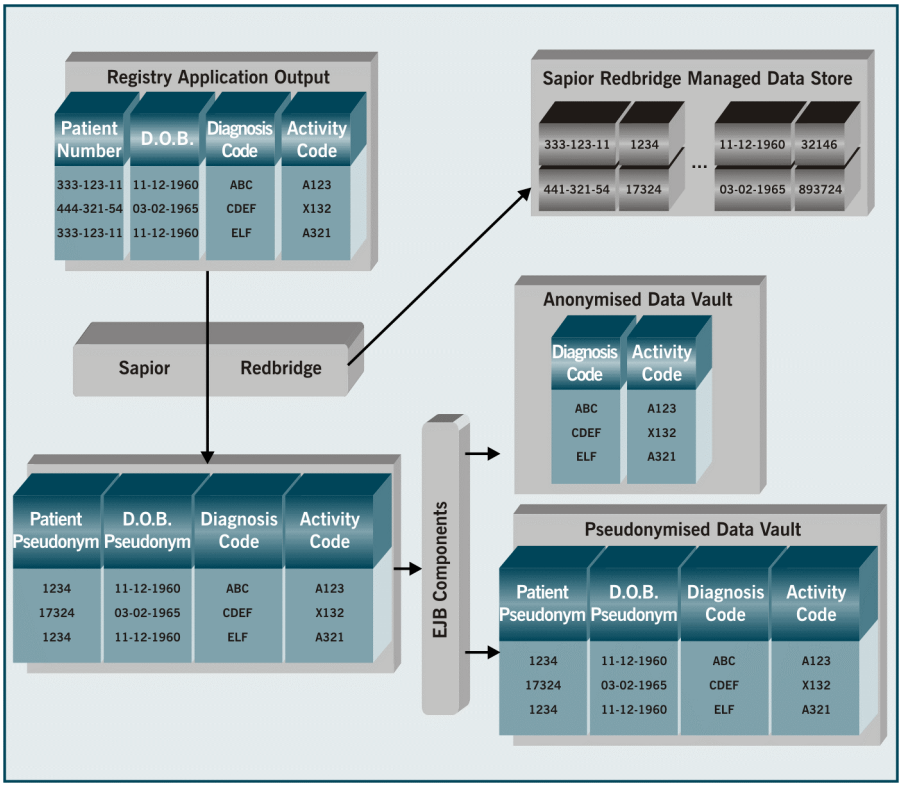
There is therefore a strong case to separate users from direct contact with the valuable corporate data stores, providing them with only as much access as they need to perform their work-related tasks. In most cases, the success of this approach depends upon a clearly understood and closely supported security policy, which defines and enforces the rights of various users and roles – and realistically speaking, the majority of organisations fall at this first hurdle, either through poor definition of the chosen policy, or inadequate enforcement.

Given that compliance to set policy can fail, it may prove preferable to automate the separation of users from data. One means by which this can be done is by substituting critical data elements with pseudonyms. 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 context of the interaction, and cannot amend it in an unauthorised manner at all. Copying pseudonymised data is similarly pointless, as the key connecting the valuable link between the accessible pseudonym and the actual data itself is held elsewhere.

The diagram overleaf illustrates the process of pseudonymising data using the subject of this Technology Audit, Sapior Redbridge. In this instance, only three items of data are being used, but in a realistic enterprise environment the number will be vastly greater than this – accordingly, Sapior has built scalability into the processing capability of Sapior Redbridge as a core requirement.

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.

Sapior Redbridge replaces sensitive fields, typically identifying data such as name and address, with pseudonyms, and maintains a separate data store of relationships between critical identifiers (name, address, date of birth) and pseudonyms. The diagram below uses patient number and date of birth as identifiers. The 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 (relevant to the enquiry being made), whilst the second contains fully pseudonymised field results. In this case, the results would be processed using Enterprise JavaBean (EJB) components.



Pseudonymisation of Medical Records Using Sapior Redbridge

It is worth considering the implications of treating data in this manner. In this case, medical research can be carried out on patient records without the researchers involved having direct access to confidential information at all – the anonymised 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. Correlation of data remains valuable in identifying trends and relationships, but at no time is privacy or anonymity compromised, a strong legislative requirement in this field.

Pseudonymised results can add richer levels of analysis, by enabling extra fields, such as treatment types, geographical location, age, etc, to be cross-referenced for relationships. Again, the key identifiers, in this case the identities of the individuals involved, remain unavailable to the user, but this in no way impacts upon the ability of the operator to extract value from the data itself.

The field of medicine does represent a key opportunity for the promotion of pseudonymisation, but the value of the approach as represented by Sapior's product extends into a wide range of data-intensive companies. Customer-centric companies, such as financial institutions and telecommunications providers, face enormous difficulties in extracting value from their Customer Relationship Management (CRM) investments whilst still operating within increasingly strict legislative boundaries concerning their use of customer data.

Sapior Redbridge is positioned as an additional layer of security against internal threats, through the transformation of sensitive data into a format that is still usable by the majority of employees, but one that is not directly comprehensible, as illustrated above. Although being primarily promoted as an internal tool, Sapior Redbridge can also be used in situations where a business must exchange data for analysis or sharing with partners. Because security in a Business-to-Business (B2B) context is often little more than an afterthought, the use of pseudonymisation also has value beyond the boundaries of the enterprise itself, if the company adopting the technology takes the time to incorporate it into data transactions.

The pseudonymised data acts as a mid-level between 'all or nothing' access, enabling an organisation to derive more information from its data with less risk and reduced impact of theft or inappropriate exposure of that sensitive data. Organisations often use sensitive identifiers (account, medical record, and social security numbers) to identify an entity (customers, patients) and its activity. Using data pseudonymised by Sapior Redbridge permits this identification without giving access to those sensitive identifiers.

Functionality

The main functionality of Sapior Redbridge is the ability to replace sensitive data with pseudonyms, or meaningless values, on a one-to-one basis, using a pseudonymisation engine. This process hides the sensitive data, whilst still providing access to the valuable data relationships – customers with similar attributes, for example. Sapior Redbridge would normally be positioned in an organisation's data flow prior to loading data into a database for access by other applications. Optimally, pseudonymisation would be part of the default preparation of all sensitive data prior to use by any and all of the organisation's activities involving the data.

The data transformation is done in the flow of data, a step that effectively minimises the exposure of the sensitive data in the various applications downstream. This minimised exposure results in reduced risk and impact of the theft of proprietary information. The Sapior Redbridge product features the pseudonymisation engine and a single front-end interface as standard, with additional front-end interfaces being available as optional extras.

The product is a Back-end or Server-side application, with its interface (or interfaces) being capable of reading/writing data and metadata, depending on specific client requirements. An instance of the application is loaded into memory for the duration of each job, and is then unloaded. Interfaces currently supported include UNIX shell and Informatica PowerCenter. Based on mature usage experience, Sapior has also incorporated the ability to accommodate business driven changes in formats of sensitive data, whilst retaining previous pseudonyms for continuity of analysis.

Initially designed for financial services, Sapior Redbridge handles very high volumes and arrival rates of data and automatically scales to meet the changing needs of the environment. User scalability is dependent upon available CPU and memory resource, and because Sapior Redbridge has no common server components, it is these hardware elements that will determine any limit upon the number of concurrent users.

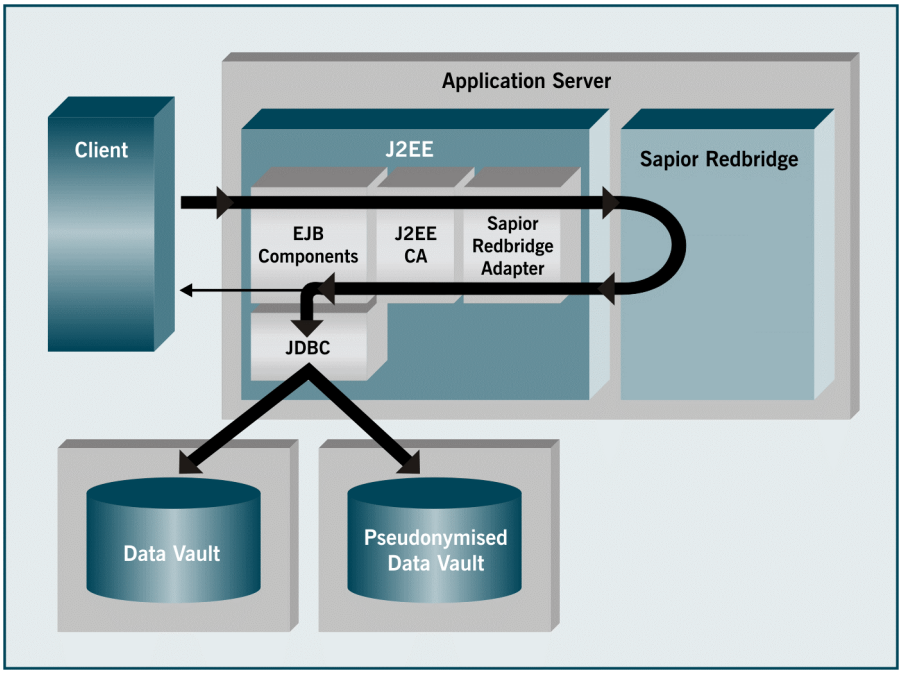
The modular design of Sapior Redbridge supports good integration with pre-existing infrastructure, and the product has been designed to show 'good behaviour' when operating within an enterprise network environment – the application will only utilise as much bandwidth as has been allocated to it, and co-exists well with other database accessing applications.

An additional aspect to the deployment of Sapior Redbridge is that as a result of reducing the threat of information theft or misuse using pseudonymisation, security processes and policies can be streamlined. Complex password matrices can be reduced to an individual, rather than company-wide, level, as only a handful of users in the organisation will actually require full access to data, and the secure management of these users will require far lower levels of resource.

DEPLOYMENT

Sapior Redbridge requires some implementation and subsequent fine-tuning in order to optimise performance levels. The extent to which this customisation will be required will be entirely dependent upon the extent to which Sapior Redbridge is expected to be used, and upon the type of environment involved. A database or systems administrator will usually have to spend approximately one week on the implementation of the product, supported by a Sapior Redbridge expert – either a Sapior consultant, or a Systems Integrator (SI) that is familiar with the product. The product can be deployed with just the pseudonymisation engine and the initial interface, and the subsequent deployment of additional interfaces will not disturb existing functionality.

The operation of the product after implementation will require knowledge of the data and record formats in use, along with familiarity in the insertion of new data transformation functionality into existing Extract, Transform, and Load (ETL) processes. At present, options include the use of UNIX scripts, Informatica PowerCenter, or EJBs within a Java 2 Enterprise Edition (J2EE) compliant application server environment. J2EE development is a key element of the immediate future of Sapior Redbridge, and this commitment will enable J2EE application servers, such as market leaders WebLogic and WebSphere, to manage pseudonymisation services.



Interaction Between Sapior Redbridge and a J2EE Application Server

In the diagram above, note that the thicker lines indicate high volumes of data flow. The response to the requesting client uses simple code in much lower volumes, which aids in bandwidth management.

When Sapior's J2EE resource adapter is released later this year, customers will also gain the option of building their own EJBs to encapsulate business logic for use with Sapior Redbridge.

Ongoing management of the product is carried out at the UNIX shell level, and the product operates on UNIX (Sun Solaris and IBM AIX, with support being developed for HP-UX) and Linux.

In some instances, customers may have an existing pseudonym implementation (and this may be a 'hand-built' pseudonymisation function that utilises database features) or use surrogate keys to replace business keys in an ETL environment. Sapior Redbridge can retrofit to such systems, by loading existing pseudonyms from the database into its own data store, in a proprietary format.

For a retrofit, Sapior offers a utility to simplify the loading of data from IBM DB2/UDB databases. A utility for Oracle databases is scheduled for the Q4, 2003 release of Sapior Redbridge. Database connections are made using the X/Open Call Level Interface (CLI) Application Programming Interface (API).

Fault tolerance features include the inclusion of Atomicity, Isolation, and Durability features usually associated with a Relational DataBase Management System (RDBMS). These are included to simplify the process of restarting the application (regardless of why it has failed – user error, power interruption, or system error).

When Sapior's J2EE resource adapter is released later this year, customers will also gain the option of building their own EJBs to encapsulate business logic for use with Sapior Redbridge.

It is worth noting that Sapior Redbridge is positioned between source data systems and a corporate data repository, which will be utilised by reporting, analytical, and similar applications. Feeds through to these applications will normally take the form of ETL products, or application servers; as data records are sourced, pseudonymisation occurs along with transformation processes. Sapior Redbridge is not designed to cleanse the data being utilised (for example, standardising address fields) and a company must undertake its own data cleansing activities prior to an implementation.

Initial licence fees are based upon the number of server installations, the size of the server being used, the number of CPUs involved, and upon the number of interfaces required. A typical installation is cited as ranging between GB £30-50 thousand, with an associated annual fee of 15-25% of this cost for software support, maintenance, and updates.

▶ PRODUCT STRATEGY

Although its origins lie in ETL requirements, Sapior Redbridge is being marketed as a security tool, and in this role it serves to place a potentially vital layer of defence between corporate data and those who would misuse it from within the firewall. Given the greater levels of damage and cost that are typically associated with internal instances of data theft and misuse, this aspect of Sapior Redbridge alone creates a good business case for its deployment. It should also be noted, however, that additional and equally compelling benefits arise from the use of pseudonymisation in general terms.

The effects of legislation vary from industry to industry, but specific verticals, including the financials sector and healthcare providers, are subject to tight regulation on their day-to-day management of data. There are instances where even the legitimate usage of customer/patient data contravenes legislation, and yet these businesses must make use of their data resources in order to function at all. Pseudonymisation enables data to be handled at one step removed, as it were, separating the user from the actual protected data, but still allowing valuable results to be extracted and used as a basis for informed decision making.

As noted earlier, pseudonymisation enables rich analysis of available data whilst still ensuring tight control over personal and private data, which greatly simplifies the legal operations of data-intensive and customer-centric organisations.

The topic of legislation also illustrates the issue that in business, requirements will frequently change due to reasons beyond the direct control of the organisation itself. New functionality may be required in order to meet governmental, as well as operational, expectations.

Pseudonymisation is a relatively new concept in the effective use and management of data, and accordingly there are very few products currently in the market designed to enable an actual implementation. This is a small advantage for Sapior, as it faces very little competition in the provision of its functionality, but also a problem, as only a handful of businesses are even aware of the potential advantages of the approach at all.

In Butler Group's view, the development of Sapior Redbridge has followed practical and pragmatic lines – the emphasis upon scalability in processing data is absolutely correct at an enterprise level, for example, whilst the commitment to J2EE connectivity ensures that Sapior Redbridge will function in the widest possible range of environments. This strong developmental awareness seems set to continue, and must now be matched by raising the awareness of end-users as to the potential of the pseudonymisation process itself. Sapior and its competitors must educate potential clients before the model can fully claim what we see as a valid place within network security as a whole, but as this awareness spreads, the returns will justify the effort required.

COMPANY PROFILE

Sapior was established in 2000, and comes from a background in data warehousing and business intelligence consulting, coupled with scalable data management design. Sapior's founders were consultants, who identified a need for an ETL tool capable of integration with pre-existing infrastructure in order to manage very high volumes of data. Although not the original focus of Sapior Redbridge, the security aspect of the tool has been correctly identified as a strong market opportunity, and the product is now being marketed accordingly.

Sapior is privately owned, with business development partners in the US and Belgium and technology associates in the UK and California. Its flagship product, Sapior Redbridge, has been implemented by UBS AG, and Sapior is working closely with potential customers, SIs, and consultancies to further the awareness of its product and of the pseudonymisation process in general.

SUMMARY

Sapior Redbridge enables organisations to minimise security liabilities concerning proprietary information, whilst simultaneously reducing security management costs and encouraging data manipulation activities with the goal of gaining competitive advantage. Consumers benefit from enhanced data privacy whilst reaping the benefits of business intelligence, audits, and other data analysis activities.

▶ CONTACT DETAILS

Sapior Ltd.
87 Grove Road
South Woodford
London
E18 2JY
UK

Tel: +44 (0)20 8530 4015
Fax: +44 (0)20 8530 5476

www.sapior.com