CASE STUDY:

Sapior Safeguards Patient Data Privacy in NHS SUS

Executive Summary

The English National Health Service (NHS) faced a critical dilemma with its programme to centralise patient information for better management and improved healthcare: how to balance data sharing with data privacy. Surpassing extensive privacy laws and guidelines, the NHS raised the bar by requiring that data for "Secondary Uses" be provided in de-identified form to guard against internal risks.

In bidding on the project, BT sought out Sapior’s expertise in specifying, designing and implementing De-Identification solutions to hide patient details but expose patterns for analysis. Sapior had worked with major, global companies implementing the critical elements of scalability, performance and accuracy. Sapior’s “Lookup” table-based Pseudonymisation engine met onerous performance and scalability demands. It addressed the Change Management issues, which are unique to long-life databases, and the accuracy requirement, which is key to trustworthy De-Identification. Sapior’s proprietary product fulfilled extensive NHS requirements by:

- avoiding any mathematical relationship in the assignment of pseudonyms;
- guarding against a single compromised IT staff member from accessing the original sensitive values; and
- preventing departments from colluding to uncover the original sensitive values by combining their respective pseudonym sets.

For more information, contact:

Sapior Ltd
16 Byron Avenue
London
E18 2HQ

T: +44 (0)20 7060 2965
F: +44 (0)20 7748 0970
www.sapior.com

SAPIOR
ENABLING ETHICAL DATA SHARING

Introduction

When the English National Health Service (NHS) launched a program to centralise patient information to support initiatives for better management and improved healthcare, it faced a critical dilemma: how to balance data sharing with data privacy. The NHS required that patient data should be used in de-identified form except where specific justification could be made for "clear data" and approvals provided. In bidding on the project, BT sought out Sapior's technical expertise in de-identifying data, aware of Sapior's considerable experience in Business Intelligence and ethical data sharing.

Said to be "the world's biggest civil information technology programme", the National Programme for IT (NPfIT) was based on the recognition that access to and analysis of comprehensive patient information is critical for good healthcare. "Secondary Uses", which include healthcare commissioning and planning, improvement, preventive care, research and a variety of other Business Intelligence initiatives, are the target purpose for the patient data provided by the Secondary Uses Service (SUS). However, extensive patient confidentiality laws and guidance require patient records to be kept secure and strict security standards maintained to prevent any unauthorised access. There was deep concern at the prospect of this gargantuan government program being launched without sufficient safeguards for sensitive patient information. It was vital to the success of the SUS program to maintain public confidence by showing that a balance could be found between data sharing and data privacy.

The stakes could not have been higher.

That's when BT turned to Sapior, De-Identification experts, to help specify, design and implement the Pseudonymisation requirement which met the NHS *Care Record Guarantee* and *Confidentiality: NHS Code of Practice*. Ultimately, Sapior was able to meet these stringent guidelines and solve the SUS technical challenge by providing a proven, scalable, forward-looking data privacy solution which resolved the issues of such a demanding project. The Sapior solution assisted the NHS in allaying privacy advocates' concerns over SUS and scored a big win for patients who would ultimately benefit from improved health care without foregoing privacy.

The NHS and SUS Challenge

The threshold issue for the SUS project was that of the privacy and security of patient records. Not only did it have to comply with a host of privacy and confidentiality laws and guidelines, but there was a critical need to maintain public confidence. The public and the media were extremely sceptical of the government's commitment and ability to maintain privacy of patient records. There was intense scrutiny of the process which made it imperative that the NHS "get it right".

In fact, the Department of Health (DH) had taken extraordinary steps to raise the bar beyond what is generally expected by the UK Data Protection Act by implementing its *Care Record Guarantee* and *Confidentiality* frameworks. This principled and comprehensive approach led to extensive requirements designed to best protect patient privacy. To guard against internal risks for what is quite possibly the largest database of sensitive patient data, the DH mandated that only de-identified data be available for all Secondary Uses. There would be special provision for re-identifying data in the few instances that require it. This sea change required considerable innovation and extensive business process modification. In fact the initial Pseudonymisation impact assessments estimated a one year delay in implementation due to the large number of existing business processes that needed modifying to no longer work with patient identifiable data. The Department of Health would be perhaps the first organisation, public or private, to take the extra step of regularly de-identifying its own data for internal use.

De-identified data can be loosely defined as data with varying degrees of identifiability to an individual. This ranges from aggregated data, where only totals of individuals relating to a given characteristic are provided, to Distinguishable or Pseudonymised data, where information can be traced to a particular individual but does not divulge the individual's actual identity. The key point is that Pseudonymised data can be used for analysis work in this protected state as the process hides patient details but exposes the patterns in the data. In comparison, simple encryption only protects data in transit or storage. It must be unencrypted and thus unprotected during use.

In order to pseudonymise the SUS data, the NHS set out extensive and rigorous requirements. It was imperative that BT, as prime contractor, could rely on Sapior's expertise and ability to execute each requirement:

- Pseudonymisation with no mathematical relationship between the original sensitive values and the replacing pseudonyms
- A way to guard against a single compromised IT staff member having the ability to access the original sensitive values

- A way to keep departments from colluding to uncover the original sensitive values by combining their respective pseudonym sets
- Ability to cope with high data volumes, throughput and concurrency

Sapior’s Experience

Sapior was an innovative software developer in the security industry whose groundbreaking products were changing the way organizations protect their customer's sensitive data. Its founders had an extensive background in high performance data warehousing, business intelligence and scalable computing and had already developed a proprietary De-Identification product.

The experience and skill sets Sapior offered were a perfect match for the demanding requirements of the NHS project. Members of Sapior’s development team had worked as researchers in advanced scalable computing, as scalable product developers and as international consultants in scalable Data Warehousing with particular expertise in large volume, high concurrency environments. Part of their previous work experience had focused on implementing the critical elements of scalability, performance and accuracy on behalf of large, complex companies such as Travelers Insurance, a multi-billion dollar risk management company, and Swiss giant UBS AG, one of the world’s leading financial firms.

Sapior’s Proprietary De-Identification Solution

Sapior’s experience in the field had led it to identify a need for and develop critical solutions to overcome the key hurdles for high volume, privacy-sensitive projects. Sapior's base Pseudonymisation engine was developed for UBS and had been running in production at UBS for over two years with complete accuracy in a highly contended environment. In addition, while bidding on a project for the Singapore Health Authority, working with NEC, a leading global technology provider, Sapior had already developed a specialised, proprietary De-Identification product which included the security and privacy features needed to meet the NHS requirements. This proven and reliable system was designed to limit the risk of breach to the system by outside hackers and to mitigate internal breach risks. This slotted in well with BT’s technical model surrounding the SUS.

Pitfalls and Challenges

The following are some of the issues Sapior’s solution addressed, many of which would be invisible to developers without deep experience in high performance, long-term data management and mastery of scalable and concurrent processing:

- **How to Pseudonymise** - In the de-identification process, reliable privacy results are dependent on what method is used to create the Pseudonyms.

  If Pseudonyms are created by encrypting the original sensitive data on a one-to-one basis, there will be a resulting mathematical relationship between the original patient information and the de-identified data. Working with a set of the Pseudonyms, an outside hacker could work out the mathematical relationship being used to generate Pseudonyms and de-code the patient data. Worse yet, there would be no way to know that the data had been breached until the damage was done.

  If Pseudonyms are generated arbitrarily, with lists of the paired sensitive data and Pseudonyms kept in “Lookup” tables residing on a protected server, the information is fundamentally more secure. A hacker can not sit at home and work at his leisure to crack the code: physical access to the Lookup table (i.e. by copying it off the server where it lives) is necessary to link the Pseudonyms with the “Sensitives” (i.e. name, NHS number, postcode, etc) and thus expose the patient information. Protecting local servers is a solved problem. By using monitors it will be clear if a breach has occurred.

- **Performance and Scalability** – Designing a highly concurrent solution is key to addressing the performance challenges that come with a Lookup-based Pseudonymisation solution. Anywhere from one to ten million records pass through the SUS system each day. Each record must be compared against a Lookup table of sensitive data to check if each Sensitive has already been linked with a Pseudonym. The SUS system accumulated 300 million Sensitives (and their related Pseudonyms) over its first 6 months of use; all of these potentially need to be checked when a new entry to a patient's record arrives in the system. As each Lookup table grows over time, the challenge is that the time required to check through these Lookups slows the system down. The Sapior solution was designed to share these Lookups extremely efficiently.

In addition, scalable software is essential to take advantage of cheaper computing boxes which can be upgraded easily. Again scalability depends on concurrency. It's difficult to make custom code, built around a general tool such as a database, as efficient as a finely-tuned purpose-built solution. Drawing on its expertise in concurrent processing, Sapior created a hand crafted, specialised engine which splits work evenly among resources to maximise speed and efficiency.

- **Change Management issue** – Sapior taught BT about the unique data management requirements for a long-life database (i.e. 10-plus year data store). In particular, this included the need for Change Management or the ability to accommodate infrequent changes in data which may occur over a long period of time. For example, assume patient X is admitted into hospital without identification and gets assigned a temporary NHS number. These records will appear to be for a separate patient. Once patient X's actual NHS number is learned simply mapping the temporary number alongside is all that is required to fully connect that patient's history henceforth. Sapior's system accommodated these changes and kept valid patient history intact.

- **Correctness** – Accuracy is vital for trustworthy de-identification. Data Processors and Users must be able to rely on the accuracy of de-identified data since it can't be visually spot checked: the sensitive data is no longer accessible. The slightest error (2+ Pseudonyms mapped to 1 Sensitive or 1 Pseudonym mapped to 2+ Sensitives) will result in the "nightmare scenario" which requires purging all relevant databases, fixing the problem without knowing how far back the first error occurred, reloading all the data and re-running every report produced during the period in question. To ensure perfect accuracy, Sapior paid stringent attention to correctness. Beyond the usual best practices for quality assurance, Sapior also provided the client with a means for verifying results in production. This "belt and braces" extra caution, coupled with Sapior's long experience with Pseudonymisation, minimised any risk of such an expensive error.

NHS Success

Supported by Sapior's specialised expertise, the NHS achieved its goal of enabling privacy-enhanced access and analysis of pseudonymised patient data, effectively hiding patient detail but exposing patterns. This significant investment in ethical data sharing demonstrates the NHS' commitment to ensure and safeguard patient trust, the preservation of which is critical to the future success of the NHS. Key initiatives such as World Class Commissioning depend on the use of patient information; so the success of this project will have not only immediate, but also long range and far reaching effects.

> **"Sapior offers a mature, forward-looking data privacy solution that integrates easily and already meets significant future requirements," explains Rob Story, NHS Care Records Service programme director, BT. "Sapior has been extremely responsive to the demands of this ambitious project."**

Setting an Example by Raising the Privacy Bar

The SUS system clearly demonstrates some of the benefits of data sharing, but data sharing is vital not just to the Health sector. Both the Public and Private sectors can learn from SUS: a working example of balancing data sharing and data protection. In our Information Society, it is increasingly clear that organisations must master the use of information to function efficiently, let alone gain competitive advantage. Increased data integration between business partners is the next opportunity in Business Intelligence. But privacy is the keystone to sharing success and must be addressed unequivocally to reduce the barriers to that vital data sharing.

Despite frequent criticisms of Department of Health projects, the SUS system should be considered a vanguard in the arena of data privacy. The NHS has set an example by elevating the standard above that required by the Data Protection Act (UK). Raising the privacy bar further reduces the risk of a breach, creates a safer environment for data sharing, and makes possible the use of tools like de-identification to facilitate even further exchange of data.

Sapior's De-Identification Expertise - an Asset for Both the Public and Private Sector

Sapior is an acknowledged De-Identification expert with successful solutions and a proven track record. The Sapior De-Identification solution is the de-facto standard for the English NHS Spine/SUS. Sapior can provide the specialised technical support and guidance your organisation requires to balance its critical data sharing with its equally critical data protection needs.