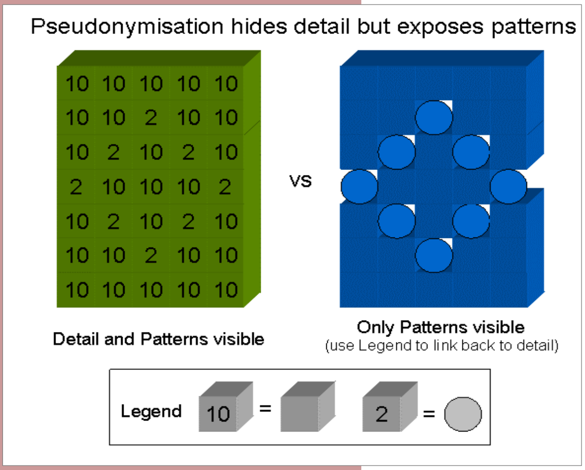
Sapior Redbridge

Pseudonymisation Solutions for Data Privacy



Sapior Redbridge de-sensitises data for analysis, data sharing and other instances where “distinguishable” but not “identifiable” data can be used to protect the privacy of individuals.

The permanent nature of anonymisation may not always be acceptable as it makes it impossible to link patterns to particular individuals. Pseudonymisation or “reversible anonymisation” offers a solution.

The process of replacing identifying fields with pseudonyms can vary both in how many fields are replaced in one go and how that replacement process is implemented. All that matters in order to ensure pseudonymisation is that the original identifying values be perfectly retrievable in some fashion.

Sapior Redbridge is a software application that plugs into a variety of data flow architectures, de-sensitising the fields passing through it that are marked as identifying whilst leaving the other fields alone. It does this by maintaining an internal database called the Managed Data Store (MDS).

Examples of Pseudonymisation

Example One: The first four identifying fields are individually replaced with four separate pseudonyms. The last two fields pass through unaltered.

Original record with six fields:

```
        name, address, DOB, SSN, DiagCode, DiagDate
```

Replacement record:

```
        name_cd, address_cd, DOB_cd, SSN_cd, DiageCode, DiagDate
```

Example Two: All four identifying fields have been replaced with a single pseudonym. The last two fields pass through unaltered.

Original record with six fields:

```
        name, address, DOB, SSN, DiagCode, DiagDate
```

Replacement record:

```
        patient_cd, DiageCode, DiagDate
```

Example Three: Single identifying field is replaced using either:

a) ECB mode of your favourite encryption algorithm (Blowfish, AES, 3DES)

b) A hash algorithm with collision avoidance logic

c) A table lookup facility

d) Some combination of the above

Each of these approaches creates a perfectly reversible process and ensures the same input always produces the same pseudonym for output.

Enabling Ethical Data Sharing

Definitions used by the UK and European Information Commissioners

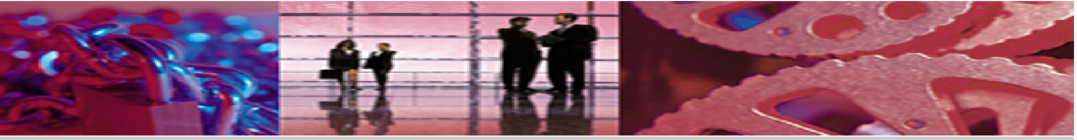
Anonymisation: The permanent removal of all personal identifiers (what defines a personal identifying column varies however), i.e. once anonymised all data linkage is lost forever.

Pseudonymisation: The process of substituting one or more true patient identifiers with pseudonyms. The true identities are not, however, discarded but securely retained allowing the original data to be reconstituted as and when this is required.

Sapior Ltd

16 Byron Avenue
London
E18 2HQ

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



SAPIOR

Product Feature	Key Benefit
Flexible & secure pseudonymisation Sapior pseudonymisation comprises up to 3 stages. A deployment must use Stage-2, but both Stages 1 & 2 are optional.	
Stage-1 (optional) ensures sensitive fields are encrypted on a remote server.	Increase the security of the MDS by only storing encrypted sensitive values therein. Protects against MDS Administrator compromise.
Stage-2 (mandatory) places the sensitive fields into a lookup table and returns a “meaningless but unique number” (MBUN).	No "mathematical" relationship between incoming sensitive and outgoing number, output value is simply the address of the next free table slot.
Stage-3 (optional) processes the Stage-2 output number making different "views" of the Stage-2 output available.	Different departments get different pseudonyms, preventing unauthorised collusion across analysis teams.
Pluggable encryption modules Sapior Redbridge ships with Blowfish, but AES, 3DES or future algorithms can be swapped in.	Future proofing against encryption compromise
Flexible sensitive field replacement Sapior Redbridge can be configured to pseudonymise individual fields, sets of those fields, or expressions thereof. There is no limit to how many pseudonyms are simulteneously calculated per record.	Ultimate flexibility in sensitive field replacement to varying suit application needs.
Ability to track changing sensitives against constant pseudonym Within data management circles it is recognised that maintaining data over years presents new problems. Identifying fields that appear constant over a year or two, actually vary over longer periods. Names change, so do addresses, even telephone area codes. Sapior Redbridge can track different sensitive field values and relate them to the same pseudonym.	Multi-year history of individuals is maintained and linkable.
Full reversibility Should there be a need to contact or identify the person behind the transaction, all Sapior Redbridge pseudonyms are fully reversible. Because Sapior pseudonyms are replacements for either single fields or sets of fields, the amount of sensitive information disclosed during reversal can be tailored for the task at hand.	Full reversibility with field level control to minimise sensitive disclosure.
Fault tolerant (UNDO and REDO logging) Full persistence has been implemented within the Sapior MDS to provide RDBMS quality recoverability. All transactions are logged to ensure MDS consistency via automatic cleanup in the event an unexpected termination occurs. Protection against disk loss is also provided via REDO logging enabling arbitrary roll-forward to chosen point-in-time. All this at 20-50 times the performance within a general RDBMS.	Consistency and recoverability of the data are ensured.
Scalable & multi-threaded Sapior Redbridge has been designed from the ground up as a multi-threaded application to make full use of all available compute resources.	Predictable performance and ability to complete tasks within specified time window.
Multiple interfaces and modes of operation Sapior has identified two common usage modes: Batch and Interactive. Batch modes efficiently push large volumes of records through the pseudonymisation process. Interactive modes need quick response, with usually lower volumes of data and often for "Self-Service" offerings. Sapior Redbridge includes support for both within a variety of data integration platforms: Unix/Linux, Remote Java Library, J2EE bean (web service), Informatica PC.	Relatively painless integration into existing infrastructure and support for multiple modes of use as needs grow.
Rich data support Both fixed and variable length data are supported. Each field can be labelled as NULL and will be processed appropriately throughout the entire system.	Simple connection to existing RDBMS data sources.

SAPIOR