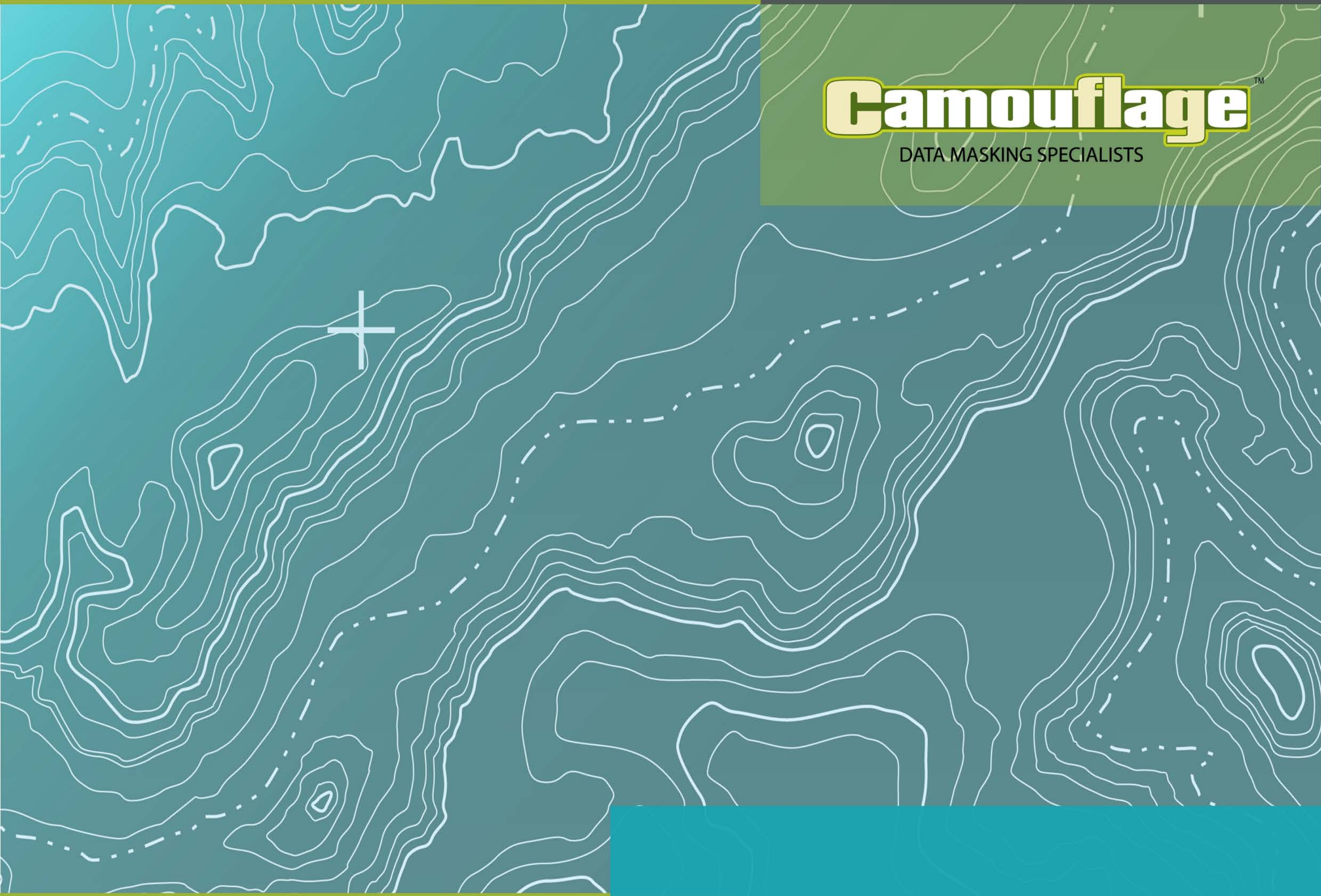


Camouflage™
DATA MASKING SPECIALISTS

WHITE PAPER

DATA MASKING BEST PRACTICE

Four Steps to Making Data Masking a Reality

July 2013

CONTENTS

EXECUTIVE OVERVIEW 2

INTRODUCTION: DATA MASKING 2

STEP ONE: DISCOVER 3

STEP TWO: DESIGN 5

STEP THREE: DEVELOP 8

STEP FOUR: DEPLOY 9

SUMMARY 11

NEXT STEPS 11

EXECUTIVE OVERVIEW

Stories consistently appear in the media reporting the negative impacts of data breaches. The impact of breaches can be detrimental to an organization and can include loss of customer confidence, poor corporate image, a drop in stock price and long-term repercussions resulting from exposed trade secrets. Additionally, non-compliance with security and privacy regulations can come with a hefty price tag. Corporations and their officers have been known to face jail time in addition to associated costs to the organization, estimated at an average cost of US$5.5 million per breach.[1]

Despite heightened awareness of these threats, enterprises continue to fall victim to data breaches. According to the 2012 Data Breach Investigations Report from Verizon RISK, in 2011 there were 855 data breach incidents reported consisting of 174 million compromised records. The overall findings from this report confirmed data breach incidents in 36 countries worldwide, illustrating the global nature of the issue.[2]

While external threats are often sensationalized, the real threat to sensitive data is from insiders. According to a 2013 report from Risk Based Security Inc., insiders accounted for 19.5% of incidents and 66.7% of exposed records, and insider wrong-doing accounted for 56.8% of exposed records.[3] The threat of insider data theft is a very real one with employees stealing trade secrets, business information such as price lists, customer information source code or business plans. The harsh reality is that in most cases users had authorized access to the data they stole. Data masking offers a way to ensure fewer individuals can access sensitive information within the organization and protect it from falling into the wrong hands.

INTRODUCTION : DATA MASKING

Masked data retains the statistical properties, integrity and realism of original data, allowing for effective and efficient testing, development research, and eliminating the risk of disclosure of sensitive data.

An increasing number of enterprises are relying on data masking to proactively secure corporate data, improve data security compliance mandates and lower costs associated with data breaches. Data masking protects data by de-identifying sensitive information contained in non-production environments and enables enterprises to extend their traditional security platforms using a proven technology.

When compared to homegrown data security techniques, data masking represents a paradigm shift in how sensitive data is secured. Masked data retains the statistical

[1] Report "2011 Cost of Data Breach Study: United States", Ponemon Institute, March 2012

[2] Report "2012 Data Breach Investigations Report" Published 2012 available at: http://www.verizonenterprise.com/resources/reports/rp_data-breach-investigations-report-2012-ebk_en_xg.pdf

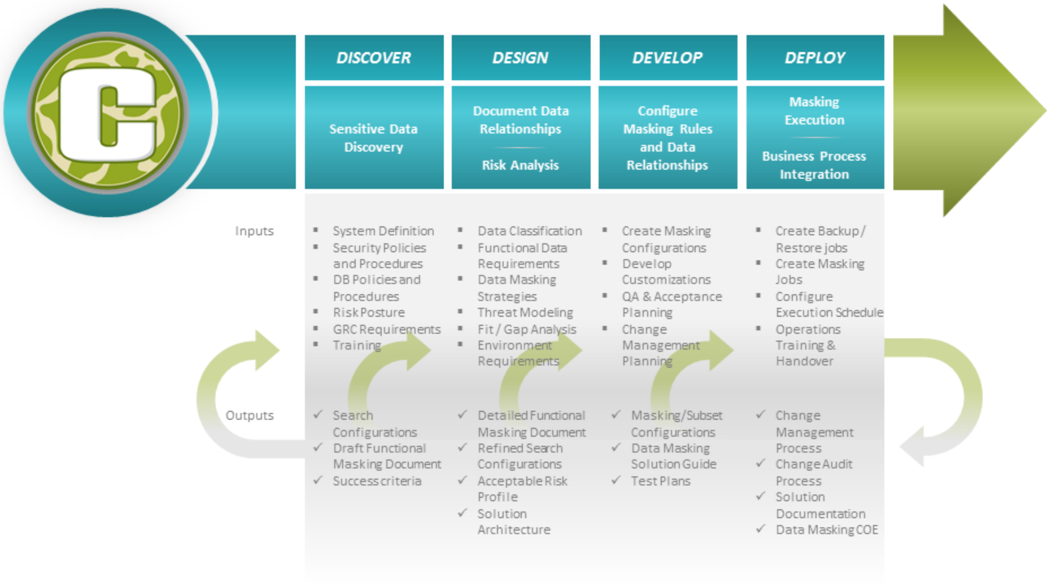
[3] Report "2013 Data Breach Quick View: United States", Risk Based Security Inc., February 2013

properties, integrity and realism of original data, allowing for effective and efficient testing, development research, and eliminating the risk of disclosure of sensitive data.

The Camouflage Data Masking Best Practice is designed to ensure data masking success from the outset, and this four-step best practice enables enterprises to create a comprehensive and practical approach to data masking in their organization.

The Camouflage Data Masking Best Practice includes 4 steps:

1. **Discover:** Identifies data that must be masked without compromising data utility.
2. **Design:** Establishes the criteria that will be used to mask the data and create context around the information found in the Discover step.
3. **Develop:** Creates data masking configurations based upon customer-specific masking requirements.
4. **Deploy:** Puts data masking into action with a plan for integrating data masking into the production-to-non-production data transition process.



	DISCOVER	DESIGN	DEVELOP	DEPLOY
	Sensitive Data Discovery	Document Data Relationships Risk Analysis	Configure Masking Rules and Data Relationships	Masking Execution Business Process Integration
Inputs	▪ System Definition ▪ Security Policies and Procedures ▪ DB Policies and Procedures ▪ Risk Posture ▪ GRC Requirements ▪ Training	▪ Data Classification ▪ Functional Data Requirements ▪ Data Masking Strategies ▪ Threat Modeling ▪ Fit / Gap Analysis ▪ Environment Requirements	▪ Create Masking Configurations ▪ Develop Customizations ▪ QA & Acceptance Planning ▪ Change Management Planning	▪ Create Backup / Restore jobs ▪ Create Masking Jobs ▪ Configure Execution Schedule ▪ Operations Training & Handover
Outputs	✓ Search Configurations ✓ Draft Functional Masking Document ✓ Success criteria	✓ Detailed Functional Masking Document ✓ Refined Search Configurations ✓ Acceptable Risk Profile ✓ Solution Architecture	✓ Masking/Subset Configurations ✓ Data Masking Solution Guide ✓ Test Plans	✓ Change Management Process ✓ Change Audit Process ✓ Solution Documentation ✓ Data Masking COE

STEP ONE : DISCOVER

Retrieve and analyze sensitive data.

To get started with data masking, the first step is to identify the data that must be masked to sufficiently protect the data without compromising data utility. The Discover step typically takes 20% of the data masking project.

Data masking must factor in all non-production environments within a given domain. The organization's requirements will vary greatly based on the size of the organization, complexity of data, scope and more. When considering data discovery, it is important to appreciate that data should be considered from a variety of perspectives - cell, record, column, table, database, system, etc. Only by examining data from all perspectives can there be a complete understanding of the effects of masking on security and utility.

With a 'virtual' border clearly delineated around the system and data, the next step is to review each field within each table for inclusion in the data masking requirements. This information is captured on a per table basis, ensuring all attributes of the field are specified (e.g. data type, involvement in keys and indexes, nullability and uniqueness).

Using a sensitive data discovery and analysis tool like Camouflage Discovery™, existing security policies and procedures will be used to drive the search for sensitive information using search engine rules. The information will then be cataloged and documented.

System Definition

A key part of the process is to define the system at the data and application levels and factor in imported and exported data sources ('upstream', 'downstream'), potential referential integrity requirements across platforms and the various requirements for non-production versions of production data.

This exercise will provide a complete picture and context for masking requirements. Tasks typically include a review of system context, cross-system integrity, non-production system requirements and existing non-production data propagation methods.

Security Policies and Procedures

Typically, most organizations not utilizing data masking will secure their data via access control and limit the number of people that have access to sensitive real data. The security and compliance policies and procedures in place within the organization are reviewed at this stage. Areas such as how database copies are handled, how databases are created, who manages or handles this information, the number of copies and how often they need to be refreshed will be examined. This will help identify any loopholes that may exist and be considered in the data masking process.

DATABASE POLICIES & PROCEDURES

The policies and procedures that exist today within the enterprise for managing data are reviewed. Some of the questions to be clarified include:

1. What is done currently with respect to creating database copies?
2. How are databases created for development and testing?
3. Who does it?
4. Who manages the process(es)?
5. How many copies of non-production data exist in the development environment?
6. How often do those copies need to be refreshed?

This is conducted to understand the existing environment and identify any loopholes that may exist and should be considered in the data masking process.

Risk Posture

Based on factors such as the types of sensitive data and security policies, the risks associated with the data and current processes are documented and the data privacy issues and concerns are discussed. All parties should have a firm understanding of the degree of risk mitigation that data masking does and does not afford.

GRC Requirements

For many organizations, governance, risk and compliance requirement challenges are important to how they perform internal tasks. GRC typically includes activities like corporate governance, compliance with government standards and regulation, and enterprise risk management. The client should identify and assess what GRC requirements they are facing in the Discover phase to ensure total compliance measures are being upheld.

Training

Training can be conducted at the client site or remotely via on-line collaboration software if requested. Camouflage offers both prepackaged training courses and custom training based on organization specific requirements. After being trained on the solution, end-users will be proficient enough to fully understand and use the product. Clients typically choose the Camouflage Quick Start Training Program which is delivered as an intensive five-day course to quickly get your organization up and running. The program provides a high level overview of the Camouflage Data Masking Best Practice and uses a standard configuration and implementation training component that focuses on rapid delivery of results.

STEP TWO : DESIGN

Establish Context for Discovered Data.

The second step puts in place the criteria that will be used to mask the data and create context around the information from the Discover step. The Design step is typically 40% of the entire data masking project.

Data Classification

Some types of data are more sensitive or valuable than others and data classification schemes reflect those differences. An appropriately applied data classification scheme can help ensure that only data that actually needs to be subject to rigorous controls is subject to those measures.

Some examples of how sensitive data may be classified include the following:

- By policy – e.g. Public, Sensitive, Private, Confidential.
- By compliance regulation or standard - e.g. PCI, HIPAA, SOX, etc.
- By business unit or line of business - e.g. commercial markets, consumer markets, etc.
- By demographic profile - e.g. all families with at least three children making over $65,000/year.
- By date - e.g. all records for the year 2012.

Functional Data Requirements

The purpose of defining functional data requirements is to ensure that all data relationships have been factored into the masking process and that the application will work after masking. Failing to do so can have serious impacts on application testing and test scripts. The major types of relational integrity to be considered are database, application and enterprise level consistency.

The proposed masking strategies need to be reviewed by the application experts from three different perspectives:

Database Level - The largest consideration is explicit Referential Integrity (RI) in the form of defined foreign keys at the database level. For many database products there are also implications if a masked field is part of an index.

Application Level - The most complex part of the data analysis is identifying what the proposed masking strategies will do to the application's ability to function. Key questions to ask here include:

- Is RI defined at the application level and/or foregoing foreign keys at the database level?
- Are there data parsing routines that could fail if masking is not done properly? For example, is ZIP/postal code compared to the state/province to ensure they are compatible?
- Are SSNs/SINs or National IDs stored with embedded spaces or are the spaces added for display only?
- Are there data type mismatches in the application level RI?

User Level - A key objective of masking data in non-production environments is to maintain a high degree of data usability. For instance, changing all names in an application to 'Xxxx/Yyyy' will ensure that names from production are not vulnerable and the application will work fine, however that masking scheme would provide little or no value in a training database. Also, developers will need accurate test cases to simulate production scenarios. These gaps should be approached from a cost-benefit perspective, weighing the impact of the proposed masking solution against the time and effort of alternative methods that ensure the end user's needs are satisfied.

Once this review is concluded, a **Functional Masking Document** is created listing specific data masking requirements to be handed off to skilled data masking user(s) responsible for building the masking configuration. The document defines specific strategies, architecture, masking requirements, workflow and any phased approaches that may be required.

Eliminating re-engineering risks

A critical step in the threat modeling process should be a risk analysis of the threat of re-engineering of masked data to original data. During threat modeling, the risk associated with de-masking the masked data is analyzed.

Hackers may attempt various statistical and propagational techniques to re-engineer or triangulate the masked data back to the original data, which is a serious threat to organizations. Some of the techniques used by hackers include single record triangulation and multiple record triangulation.

Data Masking Strategies

Having identified the various fields to be masked, the emphasis shifts from 'what to mask' to 'how to mask it'. Any number of algorithms can be applied to this process, but an understanding of the implications of the strategies is imperative. It is also important to remember that different datasets require different degrees of security and have different privacy concerns. Masking options to consider include: modified, generated, algorithmic, custom, statistical and consistent.

Threat Modeling

The objective of threat modeling is to establish a risk profile based on the company's risk tolerance. With this risk profile, the organization will decide what constitutes an acceptable level of masking for their environment and select which rows and columns to mask. For example, most organizations collaborate with suppliers and partners. The risk appetite for an external partner may be lower than an internal employee or vice versa.

Fit / Gap Analysis

Before moving to the next step, two additional considerations must be examined. First, it is important to ensure that the complete data masking solution will fit the requirements of the client. An analysis should be conducted to identify any gaps and to ensure a seamless installation of the solution in the client environment.

Environment Requirements

Additionally, a full examination should be undertaken of the infrastructure that will available during the next step that focuses on development. At this point, all players need to fully understand the uses of the post-masked data for the various environments within the enterprise. Questions should be answered in the following areas: the criteria for users, numbers of copies of data, access, issues of performance and capacity planning.

STEP THREE : DEVELOP

Creating Data Masking Configurations.

The goal of the Develop step is to create data masking configurations based upon customer-specific functional masking requirements defined in prior steps. Generally this step is 30% of the data masking project.

During this step, consideration is given as to how data masking configurations will be integrated into the overall refresh process for non-production environments. This step also provides an opportunity to develop data masking schedules and establish appropriate change management processes.

Create Masking Configurations

The Develop step starts with building data masking configuration suites based on the specific functional masking needs. This step is designed to develop the data masking configurations that will be integrated into the overall production to non-production data transition process. The process provides an opportunity to develop data masking schedules and integrate data masking into existing refresh and change management processes.

The initial masking configurations define table and field level details, database and application relationships, and provide the basis for creating masking rules.

Once these decisions are made, masking algorithms and methodologies are selected that best fit the environment. Some of the more common masking algorithms are generators such as account numbers, mutators such as modification of existing values, algorithmic which take the entire data set into account when masking data elements, custom options, and data loads which are useful when applying fictional lists of values to names, part numbers, etc.

Develop Customizations

Data masking rules are highly and easily customizable, either by modifying them using the Camouflage user interface or by scripting them manually. During this step, a key task is to determine the requirement for specialized masking methods in a particular enterprise and then use the Camouflage Custom Transformer functionality to configure a solution.

Create Database Subsets

Larger databases may need to be broken down into manageable subsets by creating smaller-sized copies of production databases as required by application development and testing teams. Database subsets are created based on user-defined sub-setting rules. For example, a subset database may contain all records from the year 2012 only.

The Big Decision – Your Organization's Masking Configurations

At this stage in the process, it's time to make a decision about your data and how it will be handled so masking can be set up properly.

Questions to answer now include:

- How should data be masked?
- What specific databases will be used for masking?
- What are the parameters for the transformation(s)?

QA & Acceptance Planning

Testing and QA are required to ensure that the scripts developed meet the functional masking needs and produce the correct results. The scope of acceptance testing and the appropriate sign-off criteria should be agreed upon during the Discover step.

Change Management Planning

A data masking project within Camouflage is a set of managed assets and code. As an organization's database structure changes, so will the Camouflage project, as any additional personally identifiable information (PII) stored will also have to be included in the masking strategy. Appropriate change management procedures will have to be implemented to facilitate any updates to the Camouflage assets.

STEP FOUR : DEPLOY

Integrate data masking with business processes.

With all the planning done, the Deploy step is about moving your data masking into action with a plan for integrating it into the overall production-to-non-production business process.

During this step, data masking schedules will be developed and data masking will be integrated into the existing change management process. This step should be about 10% of the complete data masking project as all the key decisions have been made.
Key activities during the Deploy step include:

Create Backup/Restore Jobs

Create Test Databases

After user acceptance testing is complete, test databases are created for development, testing and training environments.

Create Test Subset Databases

Similar to the creation of test databases in the "Create Test Databases" section above, for larger databases subset databases are created for development, testing and training environments. Customers with larger databases and/or customers who are concerned with database performance should consider database subsetting before data masking (i.e. creating a subset of an original database to be masked). Some data masking solutions offer selective masking functionality that allows for masking of select data within the larger database. Customers can pick and choose particular tables and columns that must be masked. This feature is also helpful for improved performance or security compliance.

Create Masking Jobs

Move Masking Scripts to Source Code Implementation Libraries

Adhering to the existing governance change management processes within the organization, the masking scripts are moved to source code implementation libraries for both version control and leveraging configurations across the enterprise.

Write Job Scheduling Scripts

Scripts facilitate the masking process by automating it for the purposes of creating development, testing QA and training environments unattended. The job scheduling process calls the data masking software to run the appropriate script during the creation of the masked database.

Configure Execution Schedule

The execution schedule is the logical order of operation to ensure database/application level integrity and improve overall performance. For example, if masking values like first, last, and full name, you need to ensure first and last name values have been masked before full name to ensure the full name accurately reflects the masked versions of first and last name.

If masking multiple data sources, determining a proper execution schedule will allow you to maximize hardware by running specific tasks in parallel.

Operations Training & Handover

Prior to handoff, in-depth training is conducted with development and testing teams. Camouflage can offer a variety of professional service offerings as well as onsite or online training for a host of data masking training classes.

A wrap-up information session is held to answer any questions the customer may have and to clarify anything that is not clear to any stakeholders within the organization

Success of the project will be closely linked to the timely review of deliverables by reviewers and approvers in accordance with the project schedule. At this point, the project is transferred to the data masking owners within the organization. Camouflage can provide post implementation coaching and support after our consultants leave the customer premises.

SUMMARY

Amid growing compliance legislation and an ever-increasing call from consumers to protect sensitive data, organizations need solutions that enable them to protect information without sacrificing the productivity of their employees.

Data masking goes beyond access control to protect data from internal users by masking copies of the most current data used in production databases, data masking enables the creation of fully functional and realistic data. Once masked, the data retains its representation without disclosing the original 'real' information.

NEXT STEPS

Make Data Masking a Reality

At the foundation of the Camouflage Data Masking Best Practice are the products and services that allow organizations to rapidly realize successful and repeatable data masking projects.

The Camouflage Data Masking Best Practice along with its complete enterprise-grade Data Masking Solution enables organizations to protect against data breaches from the inside. Some of the world's most security conscious organizations currently rely on Camouflage's deep domain expertise and products to deliver open standards based data masking approach with a very reasonable total cost of ownership.

To learn more about the software applications that form the basis of the Camouflage Data Masking Best Practice, visit our product demonstration center at http://online.datamasking.com/demo or contact a Camouflage representative today at info@datamasking.com or by phone at 1.866.345.8888.

Camouflage™
DATA MASKING SPECIALISTS

Data Masking Best Practice
White Paper
July 2013

Corporate Headquarters
130 Southside Road
St. John's, Canada
A1B 0A2

About Camouflage Software Inc.

As pioneers in data masking, Camouflage Software Inc. offers best-of-breed data masking technologies designed to ensure sensitive information is properly protected for use in application development, testing, outsourcing and training. Combined with our team data masking experts and Best Practice guidance, Camouflage offers a proven and comprehensive approach to achieving compliance.

Learn more about how Camouflage's integrated solutions help ensure information security and regulatory compliance. Visit us at www.datamasking.com, or contact a Camouflage representative at info@datamasking.com or call 1-866-345-8888.

There are no warranties, express or implied, with respect to the contents of this documentation. Camouflage shall have no liability for errors, omissions or inadequacies in the information contained herein or for interpretations thereof.