Information Obfuscation (Data Masking)

Protecting Corporate Data-Assets

Presented by Michael Jay Freer

Michael Jay Freer - Presenter Bio

Michael Jay Freer - Information Management professional providing thought leadership to fortune 500 companies including MetLife Bank, Tyco Safety Products, Capital One, Brinks Home Security, and Zales.

Over his 25+ years experience he has worked with business executives providing solutions in financial management, manufacturing, supply chain management, retail, marketing, and hospitality industries.

As an Enterprise Architect at MetLife Bank, Michael Jay specialized in Information Obfuscation facilitating project solutions for protecting business Confidential and Restricted data.

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Agenda

- Outlining the Problem
- Data Masking Golden Rule
- Defining Information Obfuscation
- Information Classification
- Who is Responsible
- Defining a Common Language
- Data-Centric Development
- Governance
- Summary
- Appendix

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Outlining the Problem

Problem Statement

Corporate Data breaches are occurring at an alarming rate.

1) It is incumbent on organizations to protect the customer, partner, and employee data with which they are entrusted.
2) Using unmasked confidential and restricted data in non-production environments exposes risks to company reputation.

Business Rationale for Obfuscating Data

- Reduce Data Breach Risks
- Heightened Legal and Regulatory Scrutiny of Data Protection Services (i.e.: SOX, HIPPA, GLBA, NPPI, FFIEC, PCI)
- Company Policies and Standards
- Fundamental assumption on the part of customers that their data is already de-identified in non-production systems

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Data Masking Golden Rule

To put Information-obfuscation (Data-masking) into perspective simply think about yourself:

How many vendors or service-providers have your personal information (banks, mortgage holders physicians, pharmacies, retailers, schools you applied to, utilities, cellular carriers, internet providers, etc.)?

Michael Jay's Data Masking Golden Rule

"Do unto your company's corporate data assets as you would have your banker, healthcare provider, or retailer do onto your personal information."

(Use this as your compass to navigate)

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Defining Information Obfuscation

Definition

Information Obfuscation is the effort in both business operations and non-production systems to protect business confidential and restricted data from easy access or visibility by unauthorized parties.

Framework

For our purposes, obfuscation includes access management, data masking, encryption of data-at-rest (DAR) and encryption of data-in-transit including principles for protecting business communications.

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Information Classification

Sensitive Data

"Sensitive" is a broad term for information considered to be a business trade-secret; or consider "private" by regulatory rule, legal act, or trade association (i.e.: GLBA, HIPPA, FFIEC, PCI).

Information Classification Levels

- **Public** – non-sensitive data, disclosure will not violate privacy rights
- **Internal Use Only** – generally available to employees and approved non-employees. May require a non-disclosure agreement.
- **Confidential** – intended for use only by specified employee groups. Disclosure may compromise an organization, customer, or employee.
- **Restricted** – extremely sensitive, intended for use only by named individuals.

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Information Classification

Sensitive Data

"Sensitive" is a broad term for information considered to be a business trade-secret; or consider "[illegible]le, lega[illegible]PCI).

PII (Personally Identifiable Information) will vary based on your company, your industry, government regulations, and jurisdiction.

- **Pu**[illegible]rights
- **In**[illegible]ilable to employees and approved non-employees. May require a non-disclosure agreement.
- **Confidential** – intended for use only by specified employee groups. Disclosure may compromise an organization, customer, or employee.
- **Restricted** – extremely sensitive, intended for use only by named-individuals.

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Who is Responsible

You are!

No matter your role in the organization, you are responsible for protecting the "Corporate Data-Assets."

Everyone else is also Responsible

All of your peers are also responsible for protecting the Corporate Data-Assets.

However, you don't have control over your peers, only over your own vigilance and how you make your management aware of any concerns, risk, or issues with the security of the Corporate Data-Assets.

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Defining a Common Language

Information Obfuscation

Information-Obfuscation (or Data-Masking) is the practice of concealing, restricting, fabricating, encrypting, or otherwise obscuring sensitive data.

This is usually thought of in the context of non-production systems but it really encompasses the full information management lifecycle from on boarding of data to developing new functionality to archiving and purging historical data.

Common Language

The Business-Information Owner, Project Stakeholders, Development Teams, and Support Teams need to use a common language when discussing the various obfuscation methods and where in the environment lifecycle an action will occur.

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Defining a Common Language

Information Obfuscation

Information-Obfuscation (or Data-Masking) is the practices of concealing, restricting, fabricating, encrypti
obs uring sensitive data.

This
but
fron
arch

The simple phrase 'just mask the data' does not address what to mask, how to mask, where to mask, or who is responsible for understanding the impact masking will have on functionality.

nguage

The Busi er, Project Stakeholders, Development Teams, and Support Teams need to use a common language when discussing the various obfuscation methods and where in the environment lifecycle an action will occur.

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Common Language – Environment Lifecycle

Common Environments

1. **Development** – Code is created, modified and unit tested
2. **Testing / QA** – System, integration, & regression testing
3. **User Acceptance (UAT)** – Business-user validation
   Test new business requirements and regression test existing functionality
4. **Business Operations** – Day-to-day business environment
5. **Business Support** – Replicate and troubleshoot business issues

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Common Language – Environment Lifecycle

Common Environments

1. **Development** – Code is created, modified and unit tested
2. **Testing / QA** – System, inte[illegible]sting
3. **U**
   T
   e
4. **E**[illegible]-day business
   environment
5. **Business Support** – Replicate and troubleshoot business issues

For Another Time

"Which of these environments will hold some level of "sensitive-data" and which are maintained as "Production Environments?"

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Common Language – Environment Lifecycle

Other Possible Environments

- **Isolated On-boarding** – When data from 3rd party partners are transitioned in, there may requirements for a secured environment to cleanse and prepare data for integration into the business operations environments.
- **Isolated Data-Masking** – Unmasked Confidential and Restricted Data should not be transferred to non-production environments. A separate secure environment allows for standardized data masking in-place

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Common Language – Environment Lifecycle

Other Possible Environments

- **Isolated On-boarding** – When data from 3rd party partners are transitioned in, there may requirements for a secured environment to cleanse and prepare data for integration into the business operations environments.
- **Isolated Data-Masking** – Unm

Example

A mortgage service provider staging loans when the servicing responsibility has not officially transferred.

Do you consider this to be "production-data?"

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Common Language – Environment Lifecycle

Other Possible Environments

- **Isolated On-boarding** – When data from 3rd party partners are transitioned in, there may requirements for a secured environment to cleanse and prepare data for integration into the business operations environments.
- **Isolated Data-Masking** – Unmasked Confidential and Restricted Data should not be transferred to non-production environments. A separate secure environment allows for standardized data masking in-place

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Common Language – Environment Lifecycle

Other Possible Environments

- **Isolated On-boarding** – When data from 3rd party partners are transitioned in, there may requirements for a secured environment to cleanse and prepare data for integration into the business operations environments.
- **Isolated Data-Masking** – Unmasked [illegible] Data should not be transferred [illegible]

Example

Company policies often state sensitive data may not be stored in non-production environments.

Moving data to "development" or "test" environments before masking would violate such company-policy.

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Common Language – Environment Lifecycle

Other Possible Environments

- **Isolated On-boarding** – When data from 3rd party partners are transitioned in, there may requirements for a secured environment to cleanse and prepare data for integration into the business operations environments.
- **Isolated Data-Masking** – Unmasked Confidential and Restricted Data should not be transferred to non-production environments. A separate secure environment allows for standardized data masking in-place

Discussion of these environments and defining what constitutes "Production" vs. "Non-production" would be a full presentation of its own.

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Common Language – Masking Taxonomy

Methods of Obfuscating Information

- **Pruning Data**
- **Concealing Data**
- **Fabricating Data**
- **Trimming Data**
- **Encrypting Data**

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Common Language – Masking Taxonomy

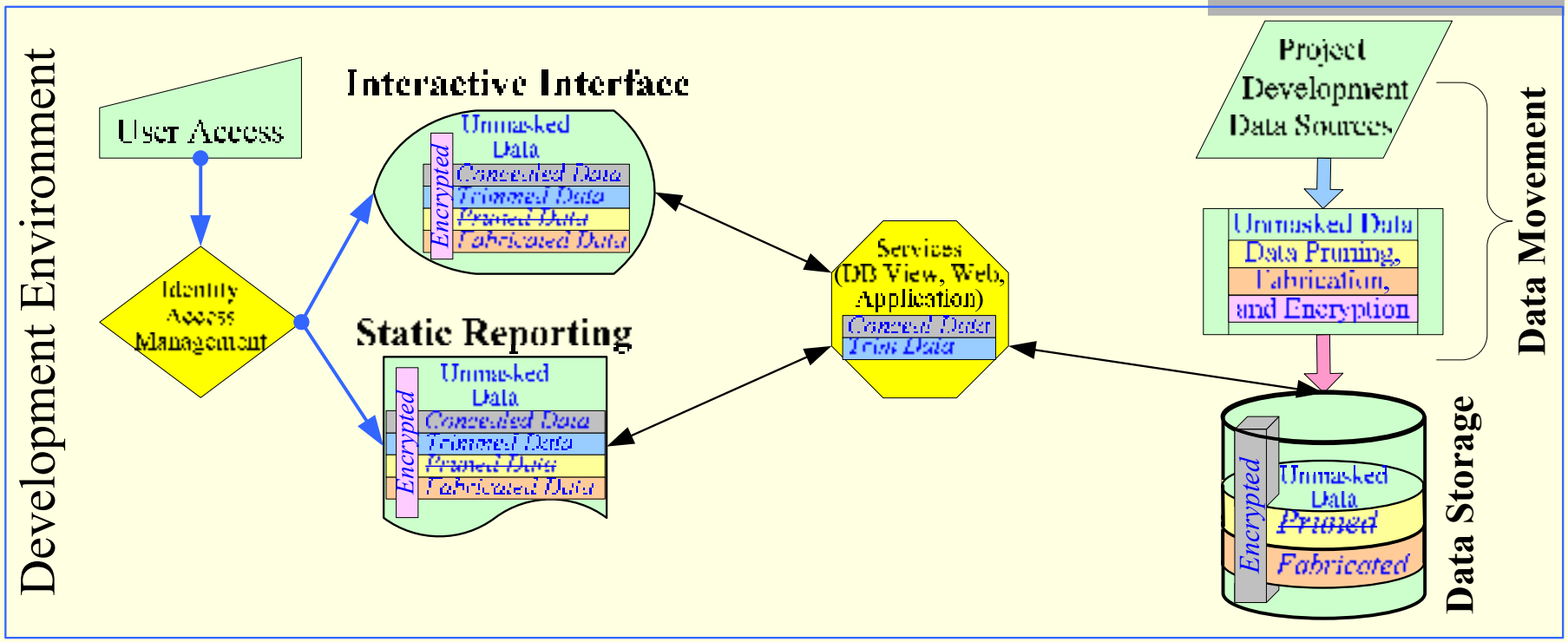
Methods of Obfuscating Information

- **Pruning Data**
- **Concealing Data**
- **Fab**
- **Tri**
- **Enc**

Before we discuss Obfuscation Methods we need to discuss where obfuscation occurs.

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

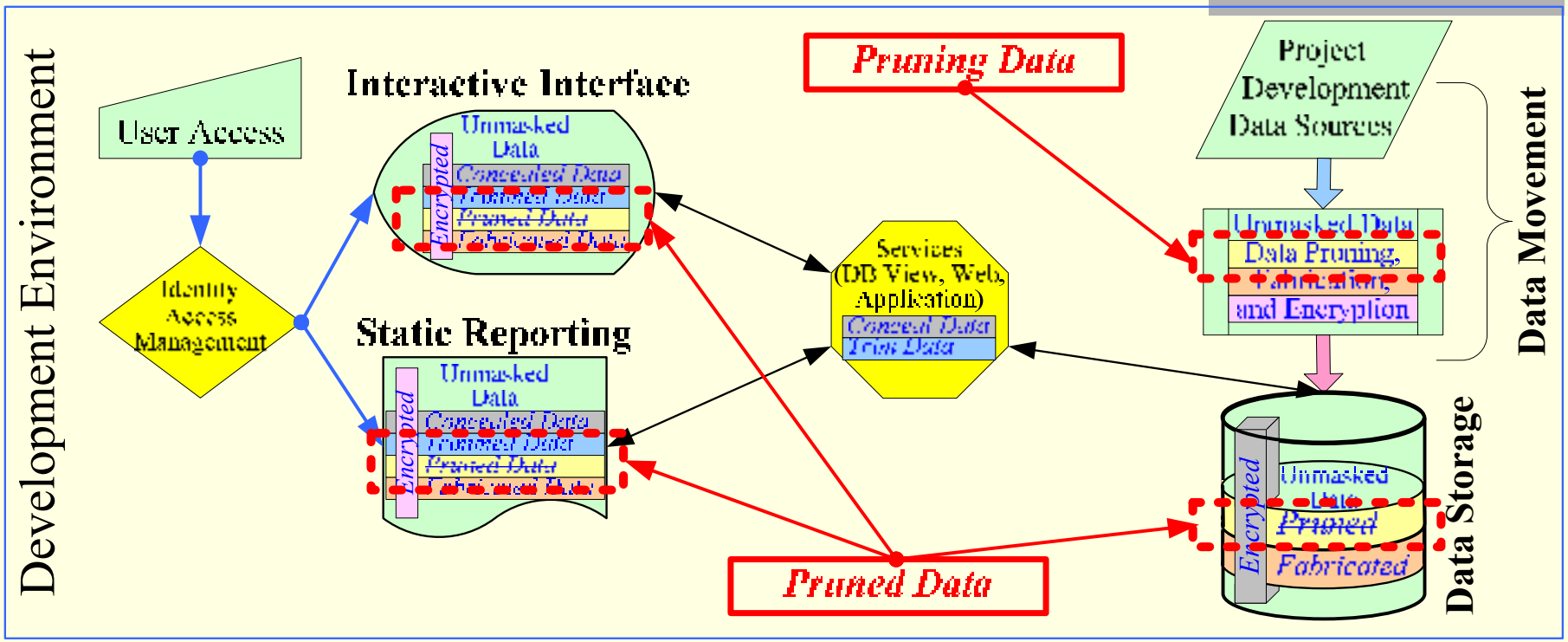
Common Language – Masking Taxonomy



- **Data Movement** – Data can be removed, shorten, or encrypted
- **Data Stores** – Data can be encrypted , data-at-rest (DAR)
- **Interactive User Interfaces** – Only show required data or portions of attributes for identification (i.e. account#, license#, SS#)
- **Static Reporting** – More restrictive than Interactive User Interfaces

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

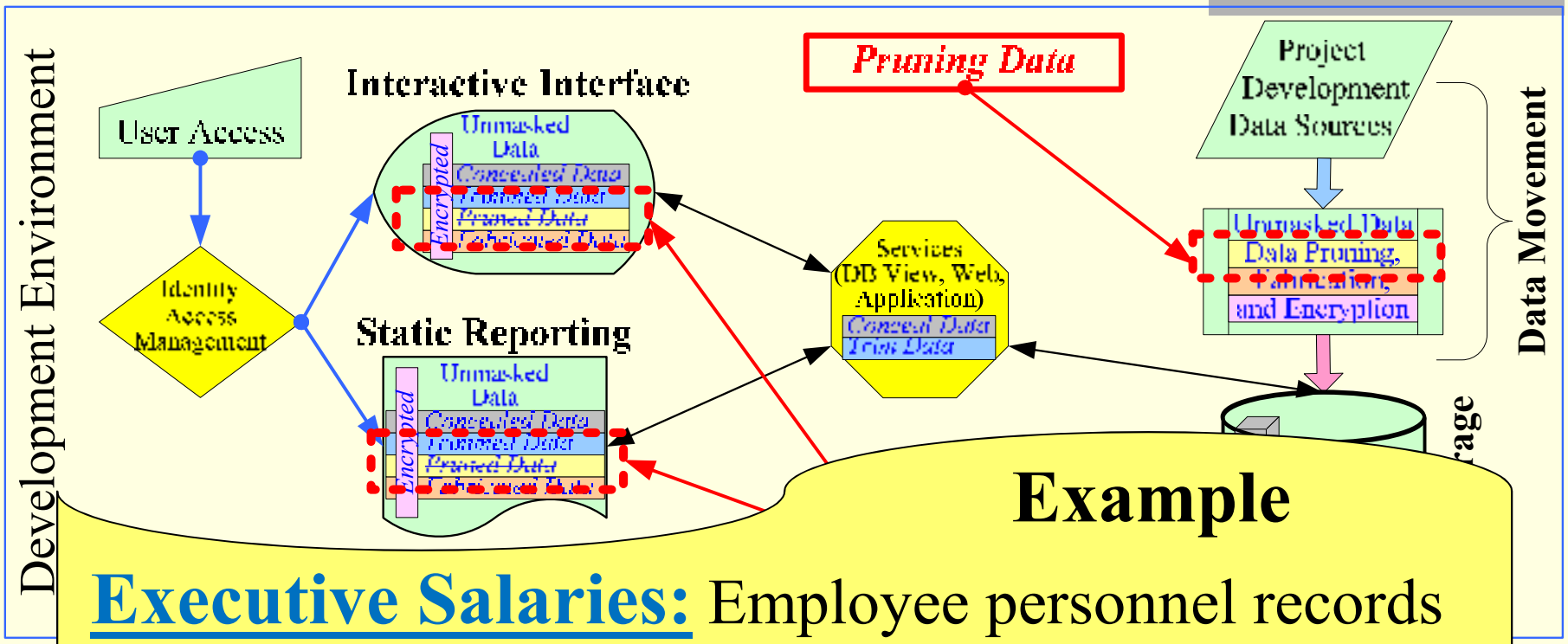
Common Language – Masking Taxonomy



Pruning Data: Removes sensitive data from attributes in non-production environments. The attribute will still appear on data entry screens and reporting but be left blank.

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Common Language – Masking Taxonomy

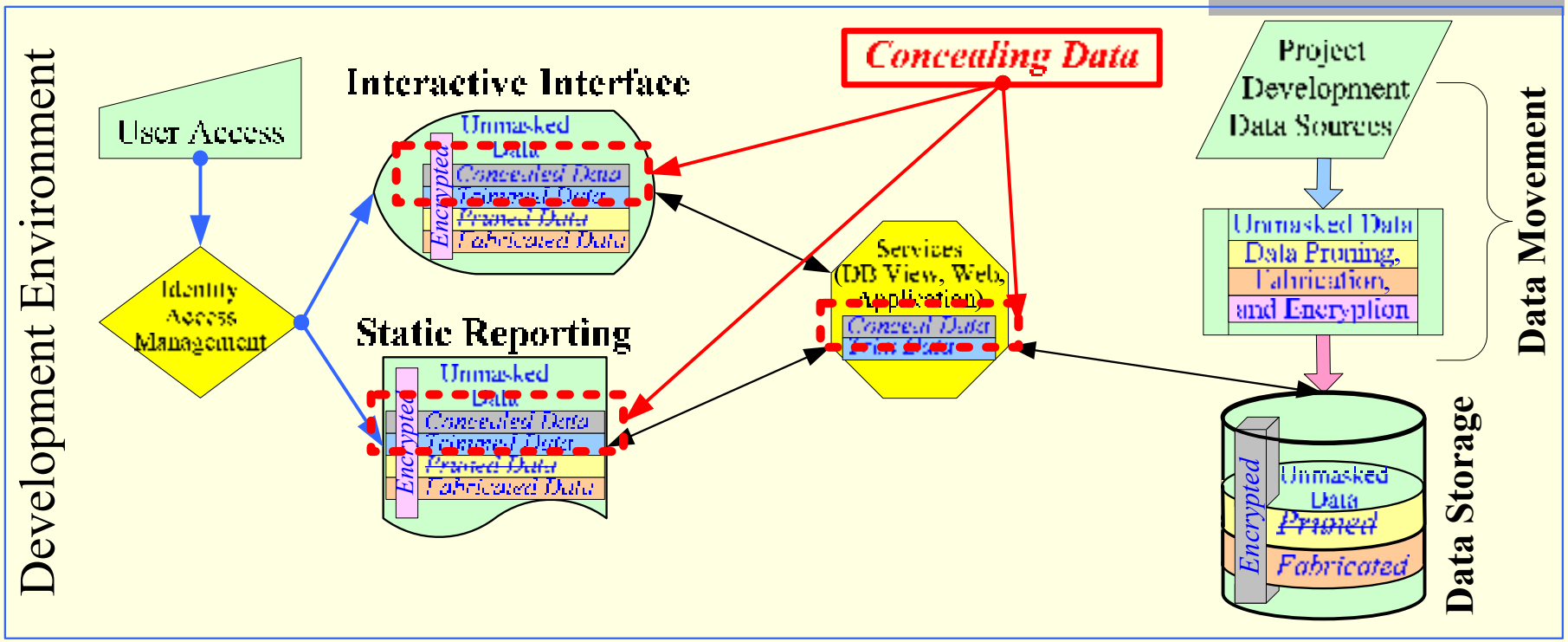


Example

Executive Salaries: Employee personnel records can de-identify by changing Emp#, SS#, & names but executive management records are easily tied back to the organizational hierarchy (top 10 salaries).

Michael Jay Freer MJFreer@Comcast.net / (954) 249-1530

Common Language – Masking Taxonomy

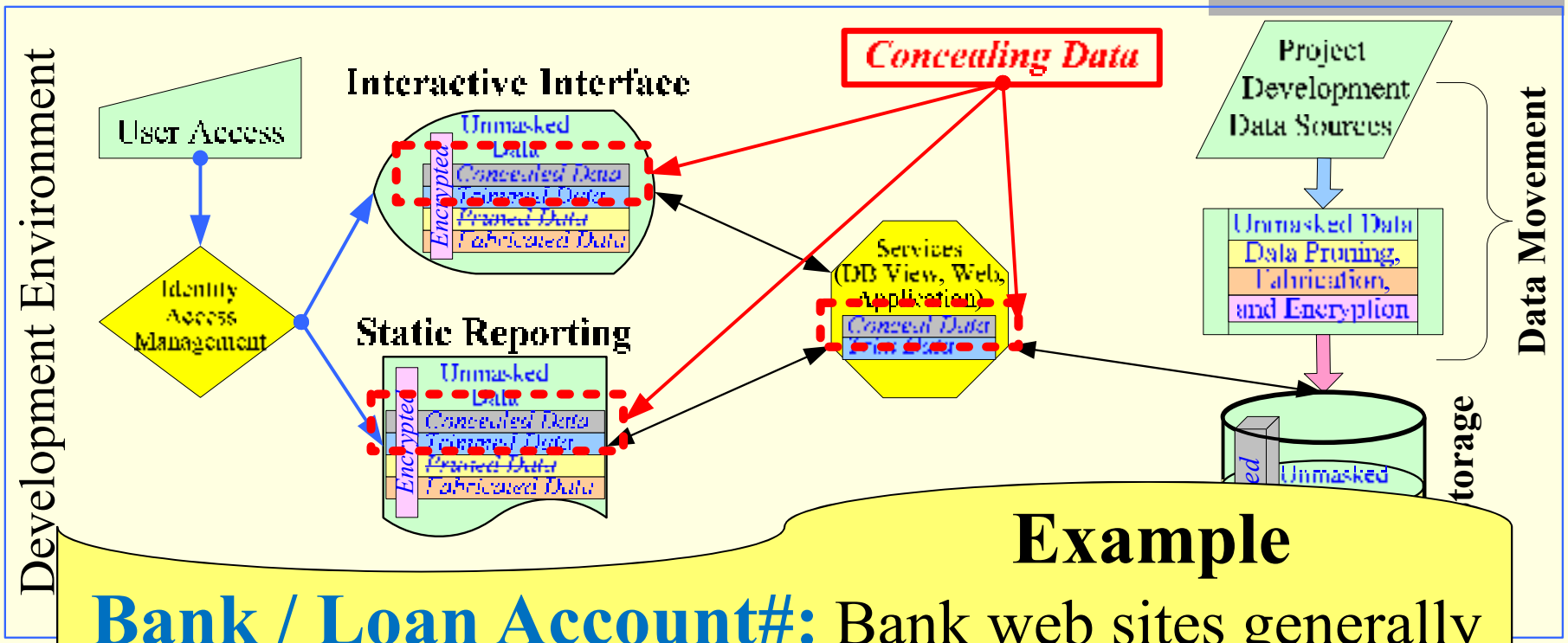


Concealing Data: Removes sensitive data from user access and visibility. For data entry screens and reports, the attribute does not appear at all versus being ***Pruned*** (blank).

Concealing data depends on clear rules for Access, Authentication, and Accountability.

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Common Language – Masking Taxonomy



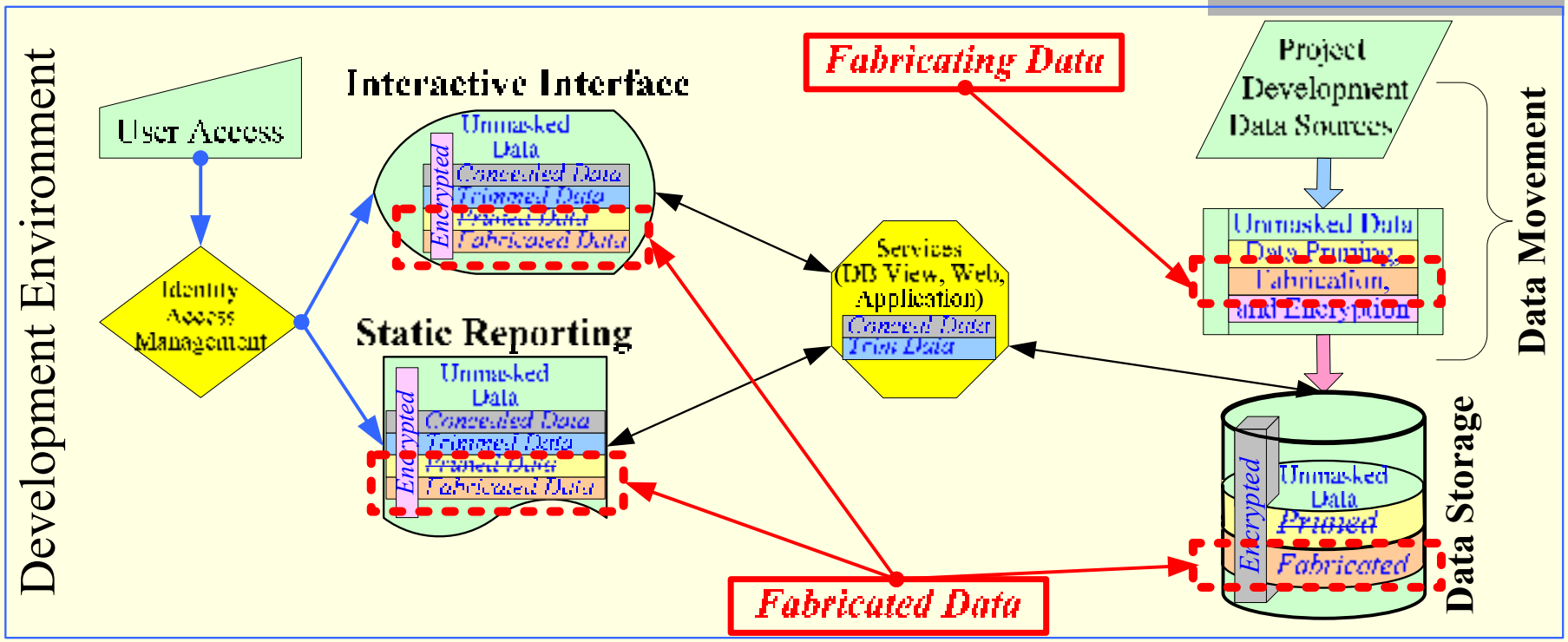
Example

Bank / Loan Account#: Bank web sites generally do not display account numbers even to the account holder.

Concealing data depends on clear rules for Access, Authentication, and Accountability.

Michael Jay Freer MJFreer@Comcast.net / (954) 249-1530

Common Language – Masking Taxonomy

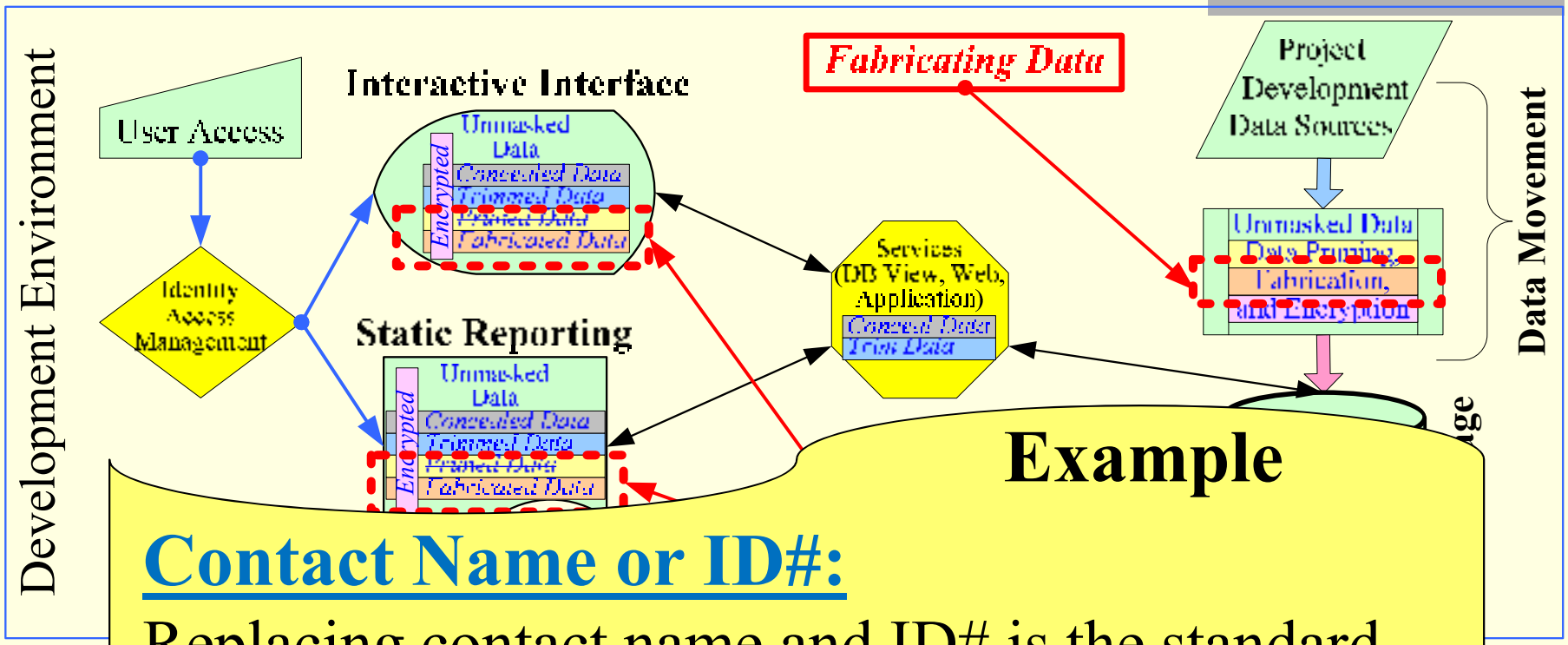


Fabricating Data:

1) Creating data to replace sensitive data
2) Creating data to facilitate full functional testing

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Common Language – Masking Taxonomy



Example

Contact Name or ID#:

Replacing contact name and ID# is the standard method for de-identifying customer and employee records.

Fab

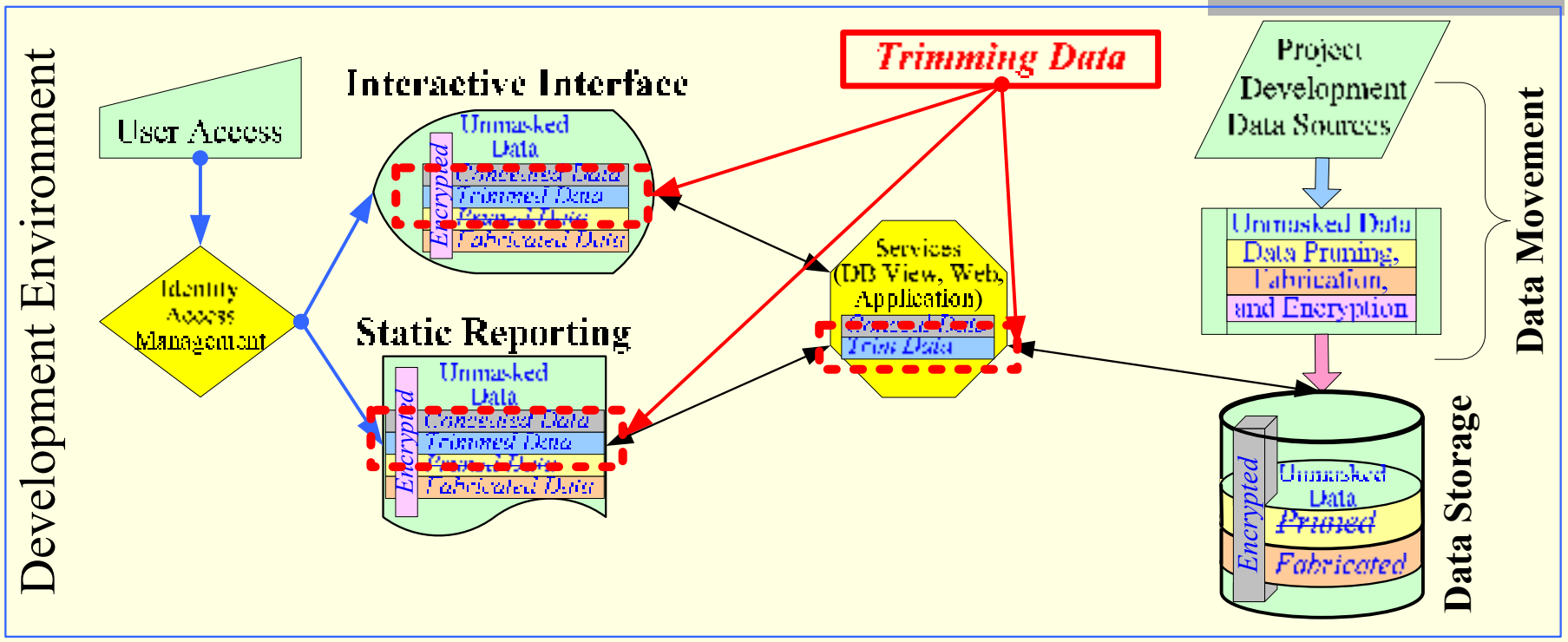
1) C

2) C

…n functional testing

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

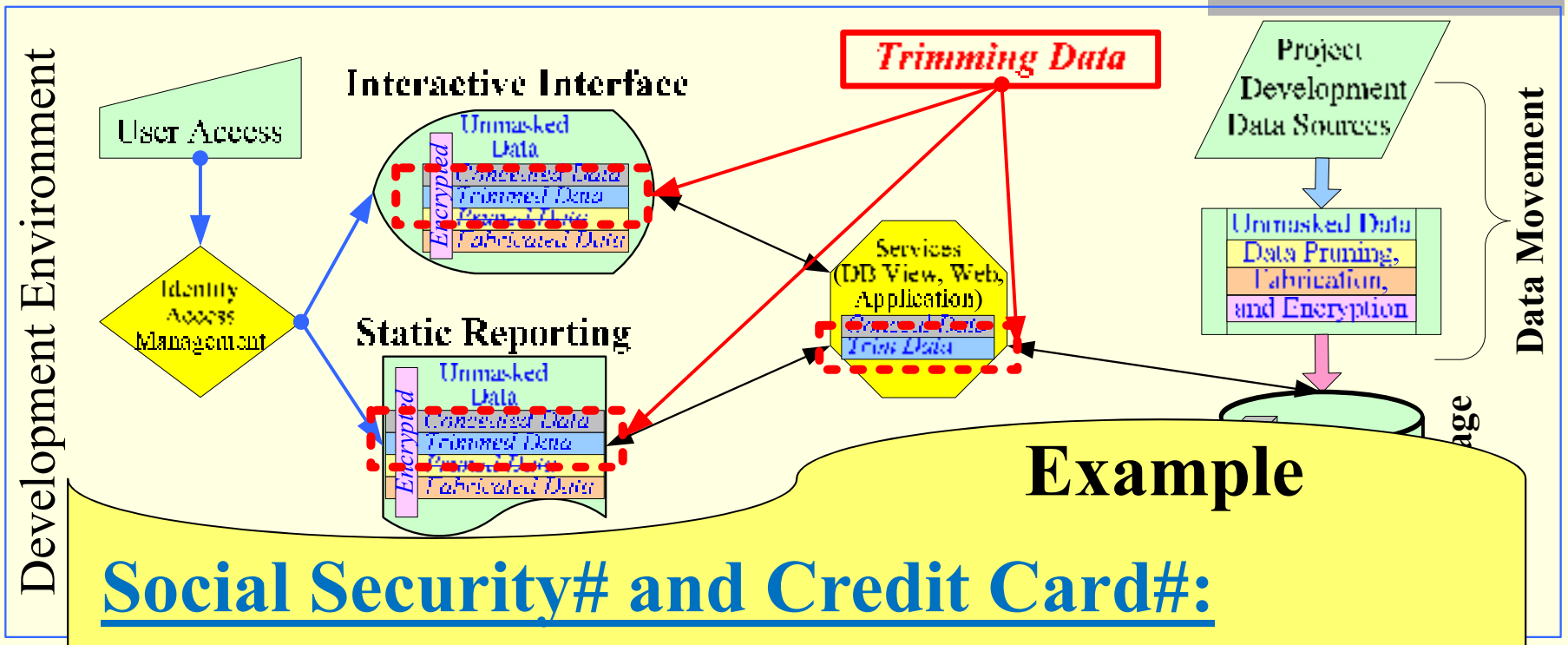
Common Language – Masking Taxonomy



Trimming Data: Removes part of an attribute's value versus ***Pruning*** which removes the entire attribute value.

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Common Language – Masking Taxonomy



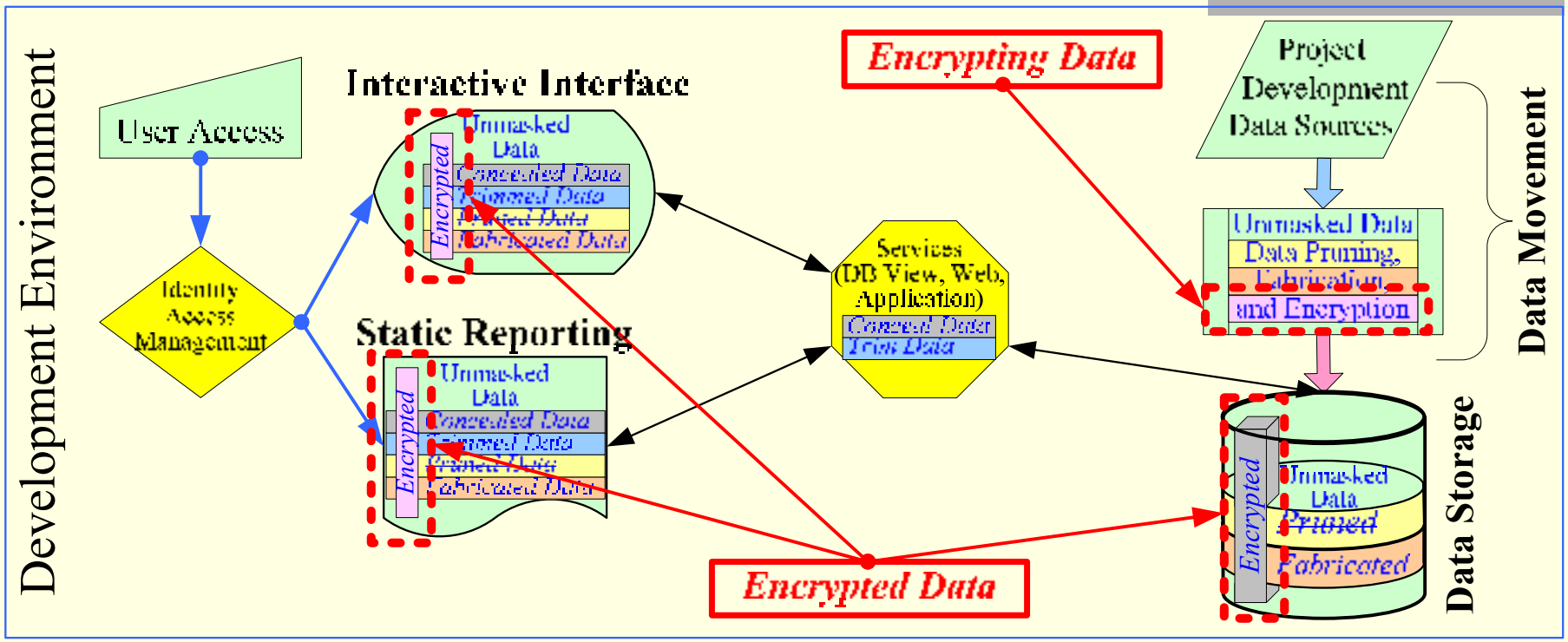
Example

Social Security# and Credit Card#:

Changing SSN# from 123-45-6789 to XXX-XX-6789 (or a new attribute = 6789) so that only part of the information is available, usually for identification.

Michael Jay Freer MJFreer@Comcast.net / (954) 249-1530

Common Language – Masking Taxonomy

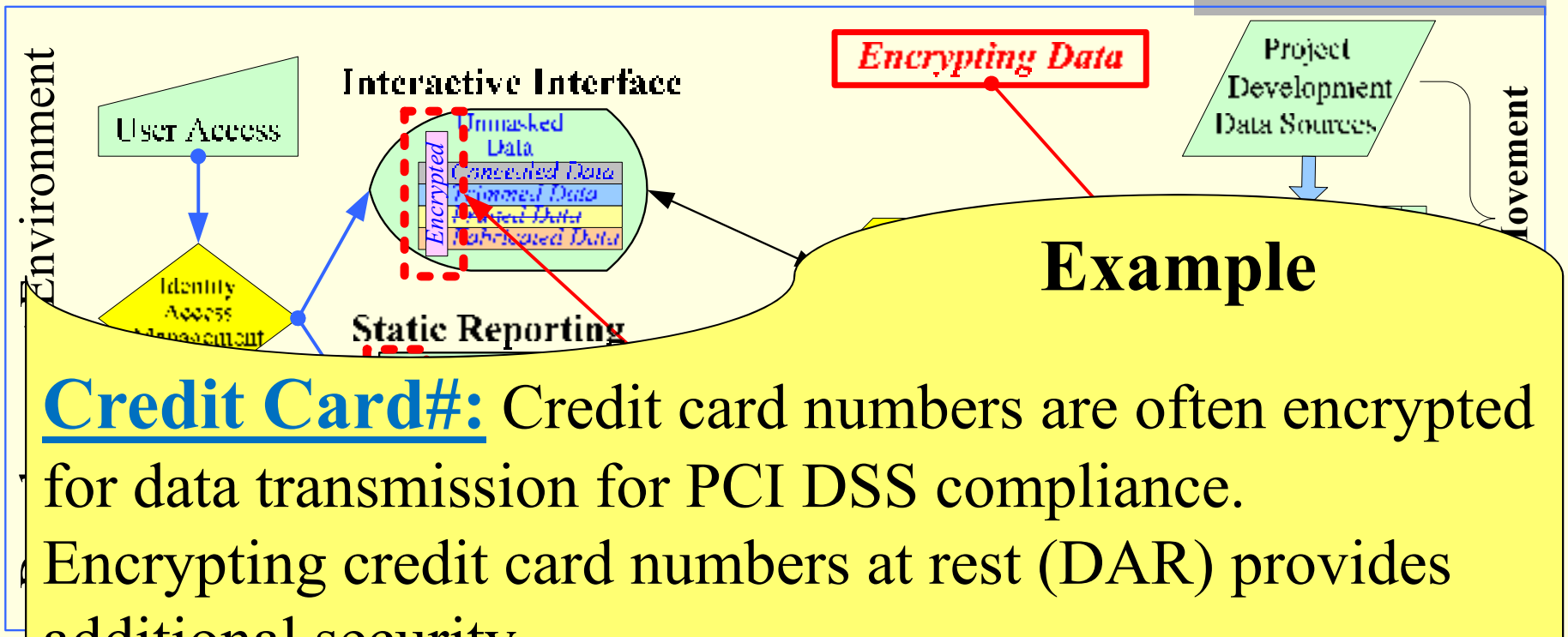


Encrypting Data: Encryption can be done at the attribute, table, or database levels

(Encrypted data can be decrypted back to the original value)

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Common Language – Masking Taxonomy



Common Language – Masking Taxonomy

- **Prune** – Pruning data removes values from non-production systems. Attribute appears on data entry screens and reporting but are blank.
- **Conceal** – Removes sensitive data from user access or visibility. For data entry screens and reports, the attribute may not appear at all or be obscured verus being *Pruned* (blank).
- **Fabricate** – Creating data to replace sensitive data and facilitate proper application testing.
- **Trim** – Removes part of a data attribute's value (*Pruning* removes the entire attribute value)
- **Encrypt** – Unlike Fabricated Data, encrypted data can be decrypted back to the original value.

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Data-Centric Development

Projects that center around data analysis (i.e.: dashboards, BI, data-marts / warehouses, etc.) often claim that they must have "production data" to develop the solution.

It is true that the business will need production data for user acceptance testing (UAT) but let's consider a few other facts:

1) Negative testing will require fabricated data
2) New functionality will also likely require fabricated data
3) Existing production data may not contain all possible values or permutations of data so full positive testing will also require some level of fabricated data
4) Full regression testing will require a standardized test set including the items above and is likely to be a combination of fabricated and masked data

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Data-Centric Development

Projects that center around data analysis (i.e.: dashboards, BI, data-marts / warehouses, etc.) often claim that they must have "production data" to develop the solution.

It is true that the business will need production data for user acceptance testing (UAT) but let's consider a few other facts:

1) Negative testing will require fabricated data
2) New functionality will also likely require fabricated data
3) Existing production data may not contain all possible values or permutations of data so full positive testing will also require some level of fabricated data
4) Full regression testing will require a standardized test set including the items above and is likely to be a combination of fabricated and masked data

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Data-Centric Development

Projects that center around data analysis (i.e.: dashboards, BI, data-marts / warehouses, etc.) often claim that they must have "production data" to develop the solution.

It is true that the business will need production data for user acceptance testing (UAT) but let's consider a few other facts:

1) Negative testing will require fa

2

3

4

If your source data has already been cleansed how would you test for exceptions (negative testing)?

Based on functional-requirements, negative tests should be created for everything outside expected ranges.

s likely to be a combination of fabricated and masked data

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Data-Centric Development

Projects that center around data analysis (i.e.: dashboards, BI, data-marts / warehouses, etc.) often claim that they must have "production data" to develop the solution.

It is true that the business will need production data for user acceptance testing (UAT) but let's consider a few other facts:

1) Negative testing will require fabricated data
2) New functionality will also likely require fabricated data
3) Existing production data may not contain all possible values or permutations of data so full positive testing will also require some level of fabricated data
4) Full regression testing will require a standardized test set including the items above and is likely to be a combination of fabricated and masked data

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Data-Centric Development

Projects that center around data analysis (i.e.: dashboards, BI, data-marts / warehouses, etc.) often claim that they must have "production data" to develop the solution.

It is true that the business will need production data for user acceptance testing (UAT) but let's consider a few other facts:

1) Negative testing will require fabricated data
2) New functionality will also likely require fabricated data
3) Existing production data may not contain all possible values or permutations of data so full positive testing will also require some level of fabricated data
4) Full regression testing will require a standardized test set including the items above and is likely to be a combination of fabricated and masked data

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Data-Centric Development

Projects that center around data analysis (i.e.: dashboards, BI, data-marts / warehouses, etc.) often claim that they must have "production data" to develop the solution.

It is true that the business will need production data for user acceptance testing (UAT) but let's consider a few other facts:

1) Negative testing will require fabricated data
2) New functionality will also likely require fabricated data
3) Existing production data may n
4) [illegible] likely to be a combination of fabricated and masked data

As systems become more complex and as regulations increase, functional tests require data sets for situations not yet present within the production data.

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Data-Centric Development

Projects that center around data analysis (i.e.: dashboards, BI, data-marts / warehouses, etc.) often claim that they must have "production data" to develop the solution.

It is true that the business will need production data for user acceptance testing (UAT) but let's consider a few other facts:

1) Negative testing will require fabricated data
2) New functionality will also likely require fabricated data
3) Existing production data may not contain all possible values or permutations of data so full positive testing will also require some level of fabricated data
4) Full regression testing will require a standardized test set, including the items above, and is likely to be a combination of fabricated and masked data (de-identified records)

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Data-Centric Development

Projects that center around data analysis (i.e.: dashboards, BI, data-marts / warehouses, etc.) often claim that they must have "production data" to develop the solution.

It is true that the business will need production data for user acceptance testing (UAT) but let's consider a few other facts:

1) Negative testing will require fabricated data
2) New functionality will also li
3) E
   or
   re
4) Full regression testing will require a standardized test set, including the items above, and is likely to be a combination of fabricated and masked data (de-identified records)

Leverage test-datasets already created for source systems.

Governance

Data stewardship is a key success factor for good data governance and in this case for good information obfuscation.

No one person will be aware of every government regulation, trade association guideline, business functional requirement, or company policy.

Include representatives from data stewardship, security, internal audit, and quality assurance teams in your solution planning and project development teams.

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Information Obfuscation Summary

1. Obfuscation occurs throughout the information lifecycle not just in non-production environments
2. Everyone is responsible for protecting the corporate data assets and the best data security tool is vigilance
3. Use a defined language to communicate who, what, where, when, and how obfuscation will occur
4. Make Information Obfuscation part of your organization's business-as-usual (BAU) processes
5. Follow Michael Jay's Data Masking Golden Rule

   "Do unto your company's corporate data assets as you would have your banker, healthcare provider, or retailer do unto your personal information."

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Questions?

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Appendix

Reference Material

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Legal & Regulatory Alphabet Soup (Sampling)

- **GLBA** – The Gramm–Leach–Bliley Act allowed consolidation of commercial & investment banks, securities, & insurance co.
- **NPI** – Nonpublic Personal Information - Financial consumer's personally identifiable financial information (see GLBA)
- **OCC** – Office of the Controller of Currency regulates banks.
- **PCI** – Payment Card Industry; defines Data Security Standard (PCI DSS) processing, storage, or transmitting credit card info.
- **PHI** – Patient Health Information - Dept of Health & Human Services ("HHS") Privacy Rule (see HIPAA).
- **PII** – Personally Identifiable Information; used to uniquely identify an individual. (Legal definitions vary by jurisdiction.)

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530

Sample Cross Reference Chart

Data Point	PII		
	PCI	NPPI	PHI
Customer - The Fact That an Individual is a Customer **		X	
First, or Last Name *; Mother's Maiden Name		X	X
Country, State, Or City Of Residence *		X	X
Telephone# (Home, Cell, Fax)		X	X
Birthday, Birthplace, Age, Gender, or Race *			
Social Security#, Account#, Driver's License#, National ID ++		X	X
Passport#, Issuing Country			
Credit Card Numbers, Expiration Date, Credit Card Security Code	X	X	
Credit Card Purchase		X	
Grades, Salary, or Job Position *			
Vehicle Identifiers, Serial Numbers, License Plate Numbers			X
Email - Electronic Mail Addresses; IP Address, Web URLs			X
Biometric Identifiers, Face, Fingerprints, or Handwriting			
Dates - All Elements of Dates (Except Year +)			X
Medical Record#, Genetic Information, Health Plan Beneficiary#			X

* More likely used in combination with other personal data
** GLBA regulation to fall into the "Restricted" classification
\+ All elements of dates (including year) if age 90 or older
++ Varies by Jurisdiction

Michael Jay Freer MJFreer@Comcast.net
/ (954) 249-1530