LTI – Data Privacy

Anonymization
Need for Data Anonymization

Use Cases

- Secondary use of data
- Third-party data use
- Data minimization
- Cross-border data movement
- Data Subject Rights Management (GDPR, CCPA and other regulations)
- Data security by design
- Reporting – Attribute-based access control
- Data movement to non-production systems

Data Sharing

- Across different stakeholders (internal/external) & geographies

Consumer Rights Management

- Anonymizing data for DSAR requests

Data Retention & Archival Management

- Anonymizing data that is no longer needed

Test Data Provisioning & Management

- Anonymizing fields, which are business-sensitive and contain personal information

Test Data Provisioning & Management

- Anonymizing data that is no longer needed

Private & Confidential
Data Anonymization - Lifecycle

Discover personal Data
Identify direct and indirect identifiers
Define Anonymization rules
Determine reidentification risks
Review risks
Apply Anonymization
Transformed data for use

Implementation Risks
- Applying anonymization on inaccurate fields
- Confusing pseudonymization techniques with anonymization
- Incompatibility with Data Quality Framework
- Irregular application of rules
- Improper evaluation of anonymized data

Business Risks
- Data privacy non-compliance
- Breaking of data chain across downstream systems
- Impact on existing reporting and analytics
- Coverage across legal agreements with third parties
LTI Anonymization SmartHub

Data Sharing across different stakeholders + Data Subject Rights Management + Data Retention & Archival Management + Test Data Provisioning & Management = LTI Anonymization SmartHub

Discover: Identify and classify different categories of personal data

Protect: Apply different masking techniques on personal data for protection

Control: Define roles and control access to personal data

Monitor: Track use of data across different systems and notify stakeholders

Scale: Integrate different systems covering internal and external data flow movement

Privacy by Design: Meet privacy requirements by curating different rules for anonymization
LTI Anonymization SmartHub – Building blocks

IBM Cloud Pak for Data

<table>
<thead>
<tr>
<th>Watson Discovery</th>
<th>Watson Knowledge Catalog</th>
<th>Red Hat OpenShift</th>
<th>Privacy Readiness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anonymization Engine</td>
<td>Data Profiling</td>
<td>Data Security Policies</td>
<td>AI-based data classification</td>
</tr>
</tbody>
</table>

End Users

- Web Portal
- SmartHub Configurator
- Web Services

<table>
<thead>
<tr>
<th>End Users</th>
<th>Watson Discovery</th>
<th>Watson Knowledge Catalog</th>
<th>Red Hat OpenShift</th>
<th>Privacy Readiness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Java Integration</td>
<td>Security</td>
<td>ERP Adaptors</td>
<td>WS/REST Integration</td>
<td>Legacy Applications</td>
</tr>
</tbody>
</table>
LTI capabilities and accelerators

**Privacy Assessment**
- Data Impact Assessment
- Sensitive Data Discovery
- Data Classification
- Data Scanning
- Target Data Search
- Structured & Unstructured
- Multilingual Support
  - LTI PrivateEye

**Data Mgmt. & Governance**
- Data Lineage
- Data Retention
- Report Automation
- Data Deletion
- Third-Party Data Sharing
- Test Data Management
- Data Inventory
  - LTI Data Lineage Framework

**Data Protection & Security**
- Privacy by Design
- Data Breach Management
- Incident Management
- Anonymization
- Data Security Controls
- I & A Mgmt.
  - LTI Data Protection Framework

**Consumer Rights Management**
- Cookie & Consent
- Request Intake Portal
- Request Broadcasting
- Workflow Automation
- Audit Trail & Logging
- SLA Management
- Privacy Notice
  - LTI EPRMS

**Privacy Reporting & Analytics**
- Subject Right Status
- DPIA Results
- Breach Mgmt. Status
- Control Testing Results
- Risk Status
- eLearning
- Vendor Risk Management
  - LTI Privacy Insights

- LTI tools leveraged
- Custom build/leverage partners
- IBM-powered
Anonymization SmartHub Capabilities

- AI/ML Services
- Dynamic Anonymization Engine
- Highly Secure (authentication, Audit trail, logging)
- Bots for Task Automation
- Manage Data Movement
- Industry Accelerators
- Active Directory integration
- Pre-built Integrators & Connectors
- Smart Reports/Dashboards
- Deploy On-Prem/On-Cloud
- Single Sign On
- Extensible APIs & Ecosystem
- Simple and easy-to-use interface
Transforming data security using Anonymization SmartHub

Data Catalogs
Identify and create data catalogs across entire technology landscape to browse under one unified platform.

Classes
Create or Edit from a comprehensive library of classes to identify different types of sensitive data.

Choose Masking technique
Pick and choose from a variety of different anonymization techniques that suit the business.

Profile Data
Perform profiling of data to identify and locate different types personal and business.

Apply Masking
Automatically apply masking based on results of profiling using classes identified and anonymization techniques selected.

Original Data
Name – John Smith
Email: John.Smith@abc.com
Phone: 888-111-222

Transformed Data
Name – JXXXX SXXXX
Email: XXXXXXX@abc.com
Phone: XXX-XXX-222
Customer Benefits

**Discover personal information & business critical data**
Identify, classify and locate data across entire technology landscape under one unified platform.

**Secure Data**
Protect data using a variety of rule-based anonymization techniques built on the platform.

**Minimize data leakage risks**
Anonymization of personal information for data sharing, data archival & secondary data usage for lawful use and processing of data.

**Accelerate Data Governance**
Extensible APIs to integrate with all services on the platform for implementing anonymization across entire ecosystem.
LTI SmartHub
Data Anonymization Solution

Highly scalable, high-performance solution that can accelerate variety of data privacy compliance use cases

Powered by IBM’s CloudPak, can be deployed anywhere: on-premise, on private clouds or on public clouds

Leverages IBM’s Watson Engine for cognitive analysis to identify business-critical and personal information

Discover personal information and business-critical data

Identify, classify and locate data across entire technology

Secure Data

Protect data using a variety of rule-based anonymization techniques

Accelerate
Data Governance

Extensible APIs to integrate with all services on the platform

Minimize data leakage risks

Anonymization of personal information for data sharing, data archival and secondary data usage
Anonymization - Preferred way of meeting data privacy obligations

- Anonymization is the destruction of the identifiable data
- Pseudonymization is a method to substitute identifiable data with a reversible, consistent value

<table>
<thead>
<tr>
<th>Privacy Obligations</th>
<th>Pseudonymisation</th>
<th>Anonymization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provide notice to data subject</td>
<td>Not covered</td>
<td>Covered</td>
</tr>
<tr>
<td>Obtain consent or have another legal basis</td>
<td>Partially covered</td>
<td>Covered</td>
</tr>
<tr>
<td>Give right to erasure / right to be forgotten</td>
<td>Not covered</td>
<td>Covered</td>
</tr>
<tr>
<td>Other data subject rights (access, portability)</td>
<td>Not covered</td>
<td>Covered</td>
</tr>
<tr>
<td>Basis for cross-border transfers</td>
<td>Not covered</td>
<td>Covered</td>
</tr>
<tr>
<td>Data protection by design</td>
<td>Partially covered</td>
<td>Covered</td>
</tr>
<tr>
<td>Data security</td>
<td>Partially covered</td>
<td>Covered</td>
</tr>
<tr>
<td>Data breach notification</td>
<td>Not covered</td>
<td>Covered</td>
</tr>
<tr>
<td>Data retention limitations</td>
<td>Not covered</td>
<td>Covered</td>
</tr>
<tr>
<td>Documentation / recordkeeping</td>
<td>Not covered</td>
<td>Covered</td>
</tr>
<tr>
<td>Vendor Agreements</td>
<td>Not covered</td>
<td>Covered</td>
</tr>
</tbody>
</table>

Data de-identification reduces data privacy obligations, additionally Anonymization technique proves to mitigate the risk and helps organization remain compliant.

### Transforming Privacy Operations using Anonymization SmartHub

<table>
<thead>
<tr>
<th>Maximum Coverage</th>
<th>Discover, catalogue and profile personal data</th>
<th>Infuse AI and Machine Learning</th>
<th>Accelerate data privacy governance</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Connect to variety of <strong>structured and unstructured data</strong> sources</td>
<td>• Automatically <strong>identify and classify relevant data</strong> assets using recommendations powered by <strong>Watson</strong></td>
<td>• <strong>Leverage AI</strong> powered by Watson to automatically identify data assets and classify business critical and personal data</td>
<td>• Get deep views into data lineage, quality, distribution through data-profile visualizations, built-in charts and statistics</td>
</tr>
<tr>
<td>• <strong>Single platform</strong> to cover complex technology landscape</td>
<td>• Create <strong>analytical data</strong> assets across systems using data catalog features</td>
<td>• <strong>Build and deploy machine learning models</strong> to accelerate data correlation and classification</td>
<td>• Manage <strong>privacy compliance and governance</strong> covering every step of data lifecycle by infusing variety of policies covering global privacy regulations</td>
</tr>
<tr>
<td></td>
<td>• <strong>Profile data</strong> to understand roots and build lineage</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Techniques- What is to know

- **Anonymization**
  - **Data Suppression**: Removal of the entire part of data (field) in the dataset.
  - **Character Masking**: Changing characters in the data. For example, John Smith will be masked as J*** M***.
  - **Data Generalisation**: Deliberate reduction in the precision of data, also referred to as recoding. For example, Data of Birth will say 1900-01-01 instead of actual date.
  - **Swapping**: Shuffling the data in the dataset such that they do not correspond to original values, but can be reidentified by using the technique used.

- **Pseudonymization**
  - **Hash / Look-up**: Values from the original dataset are replaced by synthetic data created. These can be reidentified using the hash/lookup table used to change the values.
  - **Data Perturbation**: Values from original values are modified using data distortion techniques by randomly modifying the data values often using additive noise.
Role-based Masking using Anonymization SmartHub

Anonymization SmartHub

Name – John Smith
Add – 101, 7th Avenue, 23rd St, NY
Email: John.Smith@abc.com
Phone: 888-111-222

Role based access control

Identify and classify different categories of personal data

Discover

Define roles and control access to personal data

Control

Apply different masking techniques on personal data for protection

Protect

Track use of data across different systems and notify stakeholders

Monitor

Integrate different systems covering internal and external data flow movement

Scale

Meet privacy requirements by curating different rules for anonymization

Privacy by Design

Logistics

Name – JXXXX SXXXX
Add – 101, 7th Avenue, 23rd St, NY
Email: XXXXXXXX@abc.com
Phone: XXX-XXX-222

Marketing

Name – JXXXX SXXXX
Add – XXXXXXXXX
Email: John.Smith@abc.com
Phone: XXX-XXX-222

Marketing

Anonymization

SMART HUB

Marketing

Scale

Privacy by Design

Logistics

Discover

Control

Protect

Monitor

Identify and classify different categories of personal data

Define roles and control access to personal data

Apply different masking techniques on personal data for protection

Track use of data across different systems and notify stakeholders

Integrate different systems covering internal and external data flow movement

Meet privacy requirements by curating different rules for anonymization
Thank You