OPEN DATA CENTER ALLIANCE

OPEN DATA CENTER ALLIANCE℠ USAGE: DATA SECURITY FRAMEWORK REV 1.0

TABLE OF CONTENTS

Legal Notice **3**

Executive Summary **4**

An Introduction to Data as an Asset **5**
Objective 6
Out of Scope 6
Challenges to Data Security 6

Access Control **7**
Levels of Access Control 7

Information Classification **9**
Confidentiality, Integrity, and Availability 9

Data Encryption and Digital Certificates **11**
Encryption Overview 11
Key Management 12
Certificate Management 12
Applying Confidentiality, Integrity, and Availability to Data Encryption 13

Data Masking Techniques **14**
Data Anonymization and Tokenization Methods 15

Security Information and Event Management **16**
Using SIEM to Monitor Multiple Requirements 16

Backup, Archiving, and Deletion **17**
Backups and Archives 17
Secure Deletion 17

RFP Requirements **18**
Cloud Subscriber Requirements 18
Cloud Provider Requirements 18

Summary of Industry Actions Required **19**
Industry-wide Actions 19
Cloud Provider Actions 19
Cloud Subscriber Actions 19

Appendix A: Examples for Data Anonymization and Tokenization **20**

CONTRIBUTORS

Albert Caballero, Trapezoid
Avi Shvartz, Bank Leumi
Ben Li, Deutsche Bank
Christophe Gévaudan, UBS
Ian Lamont, BMW
José Souza, UBS
Manjunath Mahabhaleshwar, Intel
Matt Lowth, National Australia Bank
Robert Rounsavall, Trapezoid
Tino Hirschmann, T-Systems, Deutsche Telekom Group

LEGAL NOTICE

This "**Data Security Framework Rev 1.0**" document is proprietary to the Open Data Center Alliance (the "**Alliance**") and/or its successors and assigns.

NOTICE TO USERS WHO ARE NOT OPEN DATA CENTER ALLIANCE PARTICIPANTS: Non-Alliance Participants are only granted the right to review, and make reference to or cite this document. Any such references or citations to this document must give the Alliance full attribution and must acknowledge the Alliance's copyright in this document. The proper copyright notice is as follows: "**© 2013 Open Data Center Alliance, Inc. ALL RIGHTS RESERVED.**" Such users are not permitted to revise, alter, modify, make any derivatives of, or otherwise amend this document in any way without the prior express written permission of the Alliance.

NOTICE TO USERS WHO ARE OPEN DATA CENTER ALLIANCE PARTICIPANTS: Use of this document by Alliance Participants is subject to the Alliance's bylaws and its other policies and procedures.

NOTICE TO USERS GENERALLY: Users of this document should not reference any initial or recommended methodology, metric, requirements, criteria, or other content that may be contained in this document or in any other document distributed by the Alliance ("**Initial Models**") in any way that implies the user and/or its products or services are in compliance with, or have undergone any testing or certification to demonstrate compliance with, any of these Initial Models.

The contents of this document are intended for informational purposes only. Any proposals, recommendations or other content contained in this document, including, without limitation, the scope or content of any methodology, metric, requirements, or other criteria disclosed in this document (collectively, "**Criteria**"), does not constitute an endorsement or recommendation by Alliance of such Criteria and does not mean that the Alliance will in the future develop any certification or compliance or testing programs to verify any future implementation or compliance with any of the Criteria.

LEGAL DISCLAIMER: THIS DOCUMENT AND THE INFORMATION CONTAINED HEREIN IS PROVIDED ON AN "**AS IS**" BASIS. TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, THE ALLIANCE (ALONG WITH THE CONTRIBUTORS TO THIS DOCUMENT) HEREBY DISCLAIM ALL REPRESENTATIONS, WARRANTIES AND/OR COVENANTS, EITHER EXPRESS OR IMPLIED, STATUTORY OR AT COMMON LAW, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE, VALIDITY, AND/OR NONINFRINGEMENT. THE INFORMATION CONTAINED IN THIS DOCUMENT IS FOR INFORMATIONAL PURPOSES ONLY AND THE ALLIANCE MAKES NO REPRESENTATIONS, WARRANTIES AND/OR COVENANTS AS TO THE RESULTS THAT MAY BE OBTAINED FROM THE USE OF, OR RELIANCE ON, ANY INFORMATION SET FORTH IN THIS DOCUMENT, OR AS TO THE ACCURACY OR RELIABILITY OF SUCH INFORMATION. EXCEPT AS OTHERWISE EXPRESSLY SET FORTH HEREIN, NOTHING CONTAINED IN THIS DOCUMENT SHALL BE DEEMED AS GRANTING YOU ANY KIND OF LICENSE IN THE DOCUMENT, OR ANY OF ITS CONTENTS, EITHER EXPRESSLY OR IMPLIEDLY, OR TO ANY INTELLECTUAL PROPERTY OWNED OR CONTROLLED BY THE ALLIANCE, INCLUDING, WITHOUT LIMITATION, ANY TRADEMARKS OF THE ALLIANCE.

TRADEMARKS: OPEN CENTER DATA ALLIANCE^{SM}, ODCA^{SM}, and the OPEN DATA CENTER ALLIANCE logo® are trade names, trademarks, and/or service marks (collectively "**Marks**") owned by Open Data Center Alliance, Inc. and all rights are reserved therein. Unauthorized use is strictly prohibited. This document does not grant any user of this document any rights to use any of the ODCA's Marks. All other service marks, trademarks and trade names reference herein are those of their respective owners.

OPEN DATA CENTER ALLIANCE[SM] USAGE: DATA SECURITY FRAMEWORK REV 1.0

EXECUTIVE SUMMARY

In many organizations today, there is a significant demand to introducing cloud computing into the enterprise. The hope is that the cloud's multi-tenant, shared infrastructure will enable greater computing efficiency, flexibility, and cost efficiency. At the same time, organizations require that compute platforms are secure and comply with all relevant rules, regulations, and laws. These requirements must be met whether using a dedicated service available through a private cloud or a service shared with other subscribers through a public cloud.

In addition to topics covered to date, the Open Data Center Alliance[SM] (ODCA) recognizes that the more organizations look to leverage the benefits of cloud, the more data they will be sending out of their environment. Therefore, ensuring that data stays secure in a cloud environment is critical to the on-going success of cloud services.

Moving highly sensitive or mission-critical data to a cloud provider is not a decision an organization takes lightly; cloud subscribers should thoroughly understand the data life cycle and the controls that can provide the appropriate level of data protection. Cloud service providers, too, need to understand these controls. Threats of tampering or theft of data when in transit means that most sensitive information is encrypted in transit. However, recent data theft (such as Sony[1]) has occurred while data is at rest—underscoring the need for cloud-based data security.

This Framework defines requirements associated with increasing data security in the cloud, and documents the following data security controls:

- **Access control.** Controlling who or what can access which data when, and in what context.
- **Information classification.** Identifying the sensitivity of the data and the impact of unauthorized access, as well as the organization's need for data integrity and data availability.
- **Data encryption.** Applying the appropriate encryption techniques to enforce data confidentiality requirements.
- **Data masking techniques.** Further increasing data security in the cloud through anonymization and tokenization.
- **Security information and event management.** Tracking and responding to data security triggers, to log unauthorized access to data and send alerts where necessary.
- **Backup, archiving, and deletion.** Identifying backup requirements and how those relate to secure storage and secure destruction of data when it is no longer needed.

This document serves a variety of audiences. Business decision makers looking for specific information around data security and enterprise IT groups involved in planning and operations will find this document useful. Solution providers and technology vendors will benefit from its content to better understand customer needs and tailor service and product offerings. Standards organizations will find the information helpful in defining standards that are open and relevant to end users.

[1] www.guardian.co.uk/technology/2011/apr/27/playstation-users-identity-theft-data-leak

AN INTRODUCTION TO DATA AS AN ASSET

Data—and the information to be gleaned from that data—is one of the most important assets a company owns. Companies protect physical assets; it is just as critical that, especially in a cloud environment, both cloud subscribers and cloud service providers understand the risks associated with data security and how to mitigate them across all layers of the cloud stack—infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS).

Table 1 provides some definitions of terms used throughout this document.

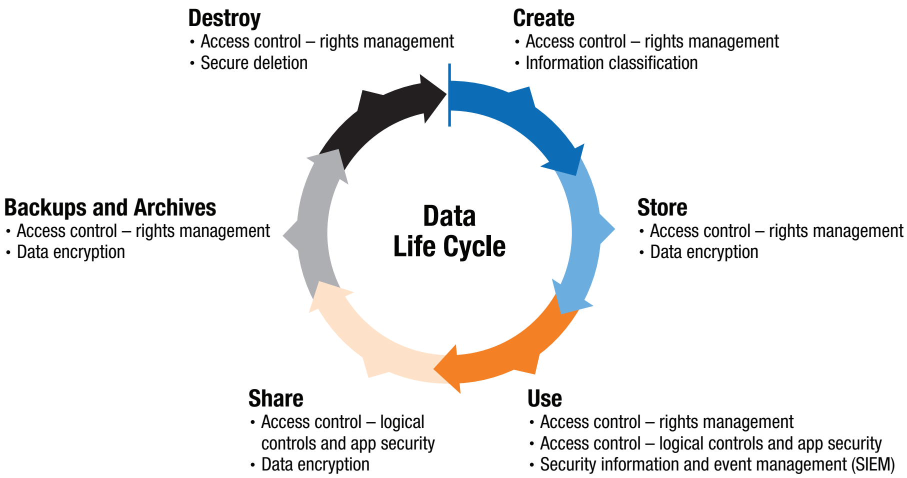
Table 1. Terms and Definitions

Term	Definition
Data	The representation of unorganized facts and figures, usually in electronic or other concrete form. Examples may include data stored on a hard drive, printed on paper, burned onto a CD, and stored/archived voice conversations.
Data protection	The right of a cloud subscriber to decide and control who–at any time and for whatever reason–is authorized to access data and under what conditions data may be used and disclosed to third parties.
Data in transit	Data that is moving through a network, including wireless transmission, whether by e-mail or structured electronic interchange.
Data at rest	Data that is not "moving," such as data that resides in databases, file systems, flash drives, memory, and any other structured storage method.
Data in use	Data in the process of being created, retrieved, updated, or deleted.
Information	A collection of data that may be processed, structured, or interpreted to add value to the data.[2] For example, a series of numbers (data) that has been added to a list of names (also data) can provide the age of each person (information).
Personally Identifiable Information (PII)	Information which can be used to identify an individual, this may include their name, address or any other information which may be unique to an individual.

Although applicable to all sorts of data, data protection is particularly relevant in the context of personally identifiable information (PII), as strict rules and regulations pertain to this type of data in almost every geographical region.

As shown in Figure 1, data protection–often thought of as information asset management–spans the entire data life cycle, from creation through use and sharing, to eventual deletion. At every stage, certain controls must be applied to protect data from unauthorized access and use.

Figure 1. Protecting Data throughout the Data Life Cycle


Destroy
• Access control – rights management
• Secure deletion
Create
• Access control – rights management
• Information classification
Backups and Archives
• Access control – rights management
• Data encryption
Data Life Cycle
Store
• Access control – rights management
• Data encryption
Share
• Access control – logical controls and app security
• Data encryption
Use
• Access control – rights management
• Access control – logical controls and app security
• Security information and event management (SIEM)

[2] www.uoc.edu/in3/dt/20388/20388.pdf

Objective

This Framework discusses the core requirements relating to data security in a cloud environment. It is expected that this document will serve as the foundation for a series of focused usage models that will cover specific elements of data security.

The core requirements discussed in this document are as follows:

- Access control
- Information classification
- Data encryption
- Data masking techniques
- Security information and event management
- Backup, archiving, and deletion

These requirements form the foundation for of a data security strategy for companies that are considering moving to a cloud environment. They also serve as a guide for cloud service providers as they strive to provide a secure environment for cloud subscribers.

Out of Scope

The following topics are out of scope for this white paper:

- Data leakage prevention (DLP), as defined in "Open Data Center Alliance Usage: Security Provider Assurance."[3]
- Physical access controls, such as building security, cabinet security, and physical host security
- Usage models to cover specific implementations of these controls. These will be covered in subsequent documents.

Challenges to Data Security

A number of challenges are associated with providing data security in the cloud. Two of the most prominent and pervasive challenges are the legal requirements governing cloud providers, and achieving acceptable processing speeds while meeting encryption requirements.

Legal Considerations

Several legal considerations apply to data security in the cloud.

- **Encryption requirements.** In most cases, cloud service providers are not covered by legal contracts that allow them to "process" data in the same secure manner as the cloud subscriber; therefore, cloud service providers are usually considered to be within the security boundary of the subscriber. This means that to be secure, data should be in an encrypted state at all points outside of the subscriber's location.
- **Right to be forgotten.** Legal frameworks governing corporate use of cloud computing are poised to change over the next few years as governments and regulatory bodies gain a greater understanding of the inherent risks that the cloud poses. One of the primary areas of focus is data privacy.[4] One of the key elements of data privacy laws is the "right to be forgotten." The interpretation of this element is currently unclear. However, in the strictest interpretation, any customers or employees who end their relationship with a cloud subscriber can expect that all related data and information pertaining to them be completely deleted—including references in all backup and archive copies).
- **Jurisdiction.** A cloud subscriber is required to comply with the laws in all countries in which the subscriber operates. Therefore, if a cloud subscriber is operating in Europe, the subscriber must comply with the local PII laws. However, in a number of other countries where the cloud provider may also wish to operate, laws that require disclosure of information may be in direct conflict with laws in another jurisdiction.
- **Exporting of encryption keys.** Some countries regulate the exporting of encryption keys. If the cloud subscriber maintains the key management system in one jurisdiction, the subscriber may violate laws in another country where the cloud service provider processes or stores data.

[3] www.opendatacenteralliance.org/library

[4] The European Union (EU) is taking the lead in the area of data privacy, and has established strict rules and regulations governing the privacy of data.

Processing Speed

The requirement to deliver real-time results can often be difficult to meet while at the same time meeting encryption requirements. This is because in many cases, encrypting and decrypting data can increase processing time significantly. For example, if a cloud subscriber is using a cloud-based storage location in conjunction with an onsite processing system, the data flowing between the processing engine and the storage location will need to be encrypted and decrypted as it transits the boundary of the subscriber location. This will add significantly to the cost of and complexity of any solution.

ACCESS CONTROL

When data leaves the boundaries of an organization's internal environment, the risk of unauthorized access to that data increases. Because legitimate access to an organization's data is key to the continued operation of a cloud service, providing strong data access controls is critical.

To minimize the chance of data being tampered with or stolen, data access controls should embody the principal of "least privilege," as defined by the National Institute of Standards and Technology (NIST).[5] This principle states that users are given access only to information they need to perform their role or task, thereby minimizing the chance that even authorized users can inadvertently gain access to or modify information that they shouldn't have access to. Furthermore, a formal request process should restrict and govern privileged access, such as system administrator or database administrator.

The Open Data Center Alliance (ODCA) has released a number of usage models covering identity management and access to cloud services, which are relevant to protect a cloud subscriber's data. These usage models, available from the ODCA content library[6], discuss in detail how to provision access, provide access, and revoke access to data in a cloud environment.

- Identity Management Interoperability Guide
- Cloud-based Identity Governance and Auditing
- Infrastructure as a Service (IaaS) Privileged User Access
- Cloud-based Identity Provisioning
- Single Sign On Authentication

Levels of Access Control

This document focuses on data access control. Physical access to buildings and data centers, as well as logical access to premises through remote and local connectivity, is not within the scope of this discussion. Figure 2 shows some of the multiple methods for legitimate parties accessing data, including customers accessing a cloud service, authenticated staff managing accounts, and system administrators managing the service. In addition, it also shows other methods of access including application servers performing data analytics or other forms of data enrichment on the platform.

In each case, access controls to protect against illegitimate use of the data differ, depending on the user type or access type. Table 2 discusses some examples of the common types of threats against each different access type and controls that may provide protection or compensating controls against these threats. For a more detailed list of access types, refer to the ODCA's Identity Management Interoperability Guide[7].

[5] hissa.ncsl.nist.gov/rbac/paper/node5.html

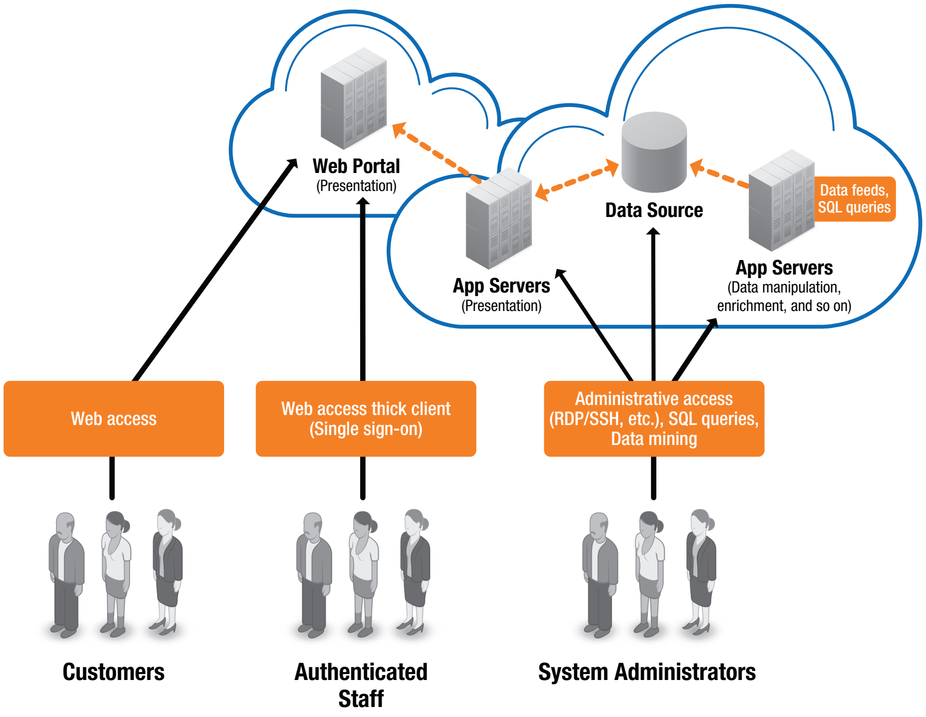
[6] www.opendatacenteralliance.org/library

[7] www.opendatacenteralliance.org/docs/Identity_Management_Interoperability_Guide_Rev1.0_b.pdf

Table 2. Examples of User Access Types and Threats[8]

User Access Type	Common Threats	**Control Recommendations** (See "Open Data Center Alliance Usage: Security Provider Assurance"[8] for a full list)
Customer (Cloud Subscriber)	• Privilege escalation • Unauthorized access (theft of credentials) • Data leakage (theft of individuals' data)	• Encryption of data in transit • Security information and event management (SIEM) • Secure software development standards • Penetration testing and vulnerability scanning of services
Staff (Cloud Subscriber)	In addition to **Customer**: • Data leakage (theft of company data)	In addition to **Customer**: • Multi-factor authentication
System Administrator (Cloud Subscriber or Cloud Provider)	In addition to **Staff**: • Theft of encryption keys • Host compromise	In addition to **Staff**: • Strong encryption of data at rest • "Four-eyes" principle for key administrator changes • Network intrusion prevention systems
Application Access	• Unauthorized access (account theft) • Data leakage (theft of company data)	• SIEM • Encryption of data at rest • Secure software development standards • Vulnerability scanning of services

Figure 2. Examples of Types of Access to Data



[8] www.opendatacenteralliance.org/library

Access controls are closely related to other aspects of the data. For example, the data classification, such as confidential or strictly confidential (discussed in detail in “Information Classification” section below), defines the assurance level a cloud service provider must support to adequately protect the data it holds or processes. Also, the access controls for a data set must include the appropriate authentication strength.

In general, access controls should be implemented at two levels: the system level and the application level. This two-pronged approach delivers protection against attacks from both administrators and users, and also helps determine who is responsible for each access control component. At both levels, access permissions should follow the “need to know, need to do” principle.

- **System level (system administrators).** At this level, access permissions to classified data are given on explicit request, not as part of a default user profile. Elevated access permissions are activated through explicit commands, logged, and regularly reviewed (such as once per year). If the data owner does not reaffirm the elevated access permission, it is revoked.
- **Application level (customers, users, and analytics).** At this level, access permissions need approval of the data owner or a dedicated custodian. Application logic is responsible for enforcing minimal disclosure of data, using techniques such as encryption and data anonymization.

INFORMATION CLASSIFICATION

Not all data is created equal—some data is more confidential or sensitive than other data. In the same way, for some data—such as financial transactions—the integrity of data is critical; for other data, a less stringent level of integrity may be acceptable. And, in certain cases data must always be available, whereas in others, a certain amount of downtime or unavailability may not be important. The measures which a cloud service provider must take—or a cloud subscriber can expect—to protect data are defined by a combination of these confidentiality, integrity, and availability (CIA) factors, in accordance with business needs and any legal, contractual, or regulatory requirements and constraints.

Confidentiality, Integrity, and Availability

To facilitate the implementation of consistent and effective information security measures, the CIA of data should be used as the foundation for classifying information assets.

Confidentiality

To protect against unauthorized or accidental disclosure, information must be classified according to the impact of such disclosure, as described in Table 3.

Table 3. Confidentiality Levels

Confidentiality	Impact Description	Minimum Protection Summary
PUBLIC	Public information that may be disclosed or released to the public as appropriate without concern as to its effect on the firm, its employees, or its business partners. Examples: Press releases, general marketing materials.	• Public information requires no specific confidentiality protection measures. • Choose appropriate external distribution according to intended recipients and time of release.
INTERNAL (default)	Information that is not intended for public disclosure; access is limited to the firm's staff and external persons with appropriate contracts and according to the "need to know, need to do" principle. Examples: Internal memos and other internal documents, internal product and process descriptions, the firm's organization structure data, policies and standards.	• Prevent unauthorized external access. • Apply appropriate protection (such as encryption) when information is transferred, transmitted, or stored outside the firm. • Protect against unauthorized access, misuse, or corruption during storage and while saving, such as by using the "clear screen" option. • Apply appropriate access control mechanisms designed to achieve the "need to know, need to do" principle. • Use the appropriate authentication mechanism, such as National Institute of Standards and Technology (NIST) level 2.[9] • Handle the data using only (a) approved, authorized hardware, software, and services on the firm's premises; and (b) approved software and services for remote users or users of portable devices. • Do not distribute externally without approval by the data owner or assigned custodian.
CONFIDENTIAL	Information that is intended for use within the organization or subject to authorized disclosure with external parties. Its unauthorized disclosure could adversely impact the organization, its clients, employees, or business partners. The information is intended for a restricted group of persons, requiring enhanced security mechanisms to enforce adequate protection. Examples: Product strategy, unpublished financial information, non-public financial risk and operational risk information, unpublished group communication, sensitive internal communication.	In addition to INTERNAL protections: • Establish staff vetting processes that meet the data owner's requirements; allow data owner to assess access authorizations for sensitive data. • Implement enhanced access control security measures, including restricted and periodically reviewed access rights. • Protect with enhanced authentication mechanism, such as NIST level 3. • Limit physical access; keep physical access to the minimum.
STRICTLY CONFIDENTIAL	The most sensitive business information, intended strictly for use within the organization. Its unauthorized disclosure could seriously and adversely impact the organization, its employees, and its business partners. The highest standard of information security is required because of the high business impact that unauthorized disclosure or processing may cause. Example: client identifying data in restrictive jurisdiction, board papers.	In addition to CONFIDENTIAL protections: • Apply appropriate protection, such as by encryption, masking, or anonymization, while data is at rest and in transmission. • Protect with superior authentication mechanism, such as NIST level 4. • Restrict physical access; keep physical access to the absolute minimum.

[9] An overview of the National Institute of Standards and Technology (NIST) electronic authentication guidelines is available from the NIST website at www.nist.gov.

Integrity

To achieve the appropriate level of integrity, data must be classified to reflect the risk associated with its accidental or unauthorized modification or destruction, as described in Table 4.

Table 4. Integrity Levels

Integrity	Impact Description	Minimum Protection Summary
BASIC (default)	Non-critical data: insignificant damage if data is altered or destroyed.	• Apply IT-inherent completeness and accuracy assurance mechanisms. • Apply appropriate access control mechanisms designed to achieve the "need to know, need to do" principle. • Prevent unauthorized changes.
TRUSTED	Critical data: consistency, accuracy, and completeness of data must be enforced in order to be a reliable business partner.	In addition to BASIC protections: • Implement additional or specific traceable completeness and accuracy assurance or validation functionality. • Establish traceability of authorized data modifications. • Use enhanced access control security measures, including restricted and periodically reviewed access rights.
GUARANTEED	Critical data: consistency, accuracy, and completeness of data or non-repudiation of an activity must be provable. Non-repudiation reflects the need that an action must be verifiable; that is, an action cannot be denied.	In addition to TRUSTED protections: • Provide traceability with tamper-proof evidence.

Availability

Availability requirements apply to three operational areas: capacity management, performance management, and backup and recovery. In each case, a cloud service provider should agree upon information and systems requirements with the cloud subscriber according to the service levels described in the "Open Data Center Alliance Usage: Standard Units of Measure for IaaS"[10] and record these agreements. The cloud service provider should also periodically test and verify the ability to deliver against those objectives.

DATA ENCRYPTION AND DIGITAL CERTIFICATES

Data can be secured (as opposed to desensitized, which is discussed in the "Data Masking Techniques" section) using two methods: encryption and digital certificates. By combining encryption and digital certificates, cloud service providers can protect data through authentication, integrity, encryption, and token verification.

Encryption Overview

Data encryption refers to mathematical calculations and algorithmic schemes that transform plain text into cipher text, a form that is non-readable to unauthorized parties. Encryption is one of the most effective ways to achieve data security. The two components required to encrypt data are an algorithm and a key. The algorithm is generally known and the key is kept secret. The key is a very large number that should be impossible to guess, and of a size that makes an exhaustive search impractical. The key, when known, enables decryption of the file.

Encryption algorithms fall into two general categories. Symmetric algorithms are those in which the encryption and decryption keys are the same. For example, the Advanced Encryption Standard (AES) and triple data encryption algorithm (TDEA) use symmetric key algorithms. Asymmetric algorithms are those in which the encryption and decryption keys differ. Public key encryption methods must be asymmetric, to the extent that the decryption key cannot be easily derived from the encryption key. Asymmetric encryption is said to be somewhat more secure than symmetric encryption, as the private key is not shared.

Encryption strength is measured in bits—that is, how big the key is. The strength of the encryption is also heavily dependent on the strength of the Random Bit Generator (RBG) and on the strength of the key itself. Symmetric encryption algorithms commonly use key sizes of 56, 128, 192, and 256 bits. The NIST recommends a minimum key size of 112 bits. For strictly confidential data, a 192- or 256-bit key is suitable. For asymmetric encryption algorithms, keys can be 2048- or 3072-bits for the logarithm-based group; elliptic curve cryptography-based algorithms can use 224-, 256-, or 384-bit keys.

[10] www.opendatacenteralliance.org/index2.php?option=com_productsearch&view=lightbox&proid=14

Key Management

Key management provides the foundation for the secure generation, storage, distribution, use, and destruction of encryption keys. Poor key management may easily compromise strong algorithms. Ultimately, the security of data protected by cryptography directly depends on the strength of the keys as well as the effectiveness of mechanisms and protocols associated with keys. The following are important considerations relating to key management:

- **Policy definition.** Key management is the most complex part of data encryption. An organization must define its cryptographic security and key life cycle management policies. These policies should grant access to cryptographic keys only to authorized users, and this access should be revoked when those users no longer need access.
- **Service model.** Key management also changes based on the service model. SaaS may use external cloud key management services, separate from the cloud service provider. PaaS and IaaS can use enterprise key management services. Enterprise key management services provide the data owner more control over the keys.
- **Location of key management.** Ideally, key management should be separated from the hosting provider. This approach will help to protect against data breaches from external sources as well as an attack originating from the cloud service provider itself. If the keys are stored by an external cloud provider, the provider should keep the keys secure; key management servers should not be shared in a multi-tenant environment.

Certificate Management

Digital certificates provide an additional method of data protection, through authorization. A digital certificate is information that is encapsulated in a file or media stream that confirms that a specific person or device has originated the media. Certificates are usually created or validated by a trusted third party that guarantees that the information contained within the certificate is valid.

Cloud subscribers wishing to send data using a digital certificate create a key pair (public and private keys) and apply for a digital certificate from a Certificate Authority (CA). The CA issues an encrypted digital certificate containing the applicant’s public key and a variety of other identification information. The CA makes its own public key readily available through printed material or over the Internet.

The sender encrypts the data using the private key. The recipient of the encrypted data uses the CA’s public key to decode the digital certificate attached to the data, verifies it as issued by the CA, and then obtains the sender’s public key and identification information held within the certificate.

Applying Confidentiality, Integrity, and Availability to Data Encryption

The encryption should be defined by the CIA requirements associated with the data, as discussed earlier in "Confidentiality, Integrity, and Availability." The solution should be implemented based on the risk and a combination of the security measures adapted from the three techniques mentioned below (data in transit, data at rest, and data in use).

How Data Confidentiality Relates to Encryption

As shown in Table 5, encryption may or may not be required for various levels of confidentiality. However, when enterprises move their infrastructure, data, and applications to the cloud, protection of confidential and strictly confidential data in transit, at rest, and in use is critical. Inappropriate data disclosure could negatively affect the data owner's reputation, financial standing, and compliance with regulatory and legal requirements.

The following techniques can help protect data in its various states:

- **Data in transit.** Strong encryption methods such as Secure Sockets Layer (SSL), Transport Layer Security (TLS), and Internet Protocol Security (IPsec) can help keep data private during transit, such as between two nodes, between two storage units, or between storage and the processing system.
- **Data at rest.** Protection should be enabled through database, storage, or operating system encryptions.
- **Data in use.** The data cannot be retrieved from any location other than the original at-rest state, and requires re-authorization upon each use. Also, the cloud service provider should process data through logically or physically segmented services based on the confidentiality of the data.

Table 5. Encryption Requirements for Various Levels of Confidentiality

	Encryption Required			
	Public	Internal	Confidential	Strictly Confidential
Data in Transit	No	Yes	Yes	Yes
Data at Rest	No	No	Yes	Yes
Data in Use	No	No	No	Yes

Table 6 provides details about the level of encryption that should be implemented in the cloud.

Table 6. Encryption Requirements for Various Levels of Provider Assurance

		Level of Provider Assurance[11]			
	Type of Data	Bronze	Silver	Gold	Platinum
Data in Transit	Communication to/from service	Encryption optional	Encryption optional	Encryption required	Encryption required
	Storage fiber communication between system	Encryption optional	Encryption optional	Encryption required	Encryption required
Data at Rest	Application data	Encryption optional	Encryption optional	Encryption required	Encryption required
Data in Use	VM images and configuration files	Encryption optional	Encryption optional	Encryption required	Encryption required
	Network and storage devices, configuration files	Encryption optional	Encryption optional	Encryption required	Encryption required
	Secure boot process (starting at hardware level)	Encryption optional	Encryption optional	Encryption recommended	Encryption recommended
	System integrity checking (VM level and hypervisor level)	Encryption optional	Encryption optional	Encryption recommended	Encryption recommended
	Full-disk encryption (OS level or hardware level)	Encryption optional	Encryption optional	Encryption required	Encryption required

[11] As defined in "Open Data Center Alliance Usage: Provider Assurance." www.opendatacenteralliance.org/docs/Security_Provider_Assurance_Rev%201.1_b.pdf

How Data Integrity Relates to Encryption

Data integrity is a key consideration when choosing a method for and level of encryption. For example, if encryption keys are stored by the cloud service provider, it can be difficult to detect manipulation or access to the data stored in the cloud. A good approach is to combine the data protection methods, such as using a system integrity check along with digital certificates, which can verify the authenticity of the request. Using such an approach can help keep data private and authentic.

In cases where integrity is essential, the encryption must be tamper-proof.

How Data Availability Relates to Encryption

Encryption poses two availability challenges: data availability and key availability.

- **Data availability.** If the cloud service provider is providing cloud services through clustering—that is, copying data from one data center to another—it is important to make sure that the data are protected wherever it's copied. Even the respective key management server should also be available in the secondary data center
- **Key availability.** In the event that a cloud encryption service is not available, the keys stored with the cloud service provider will be also unavailable, making access to the data impossible without an alternate key store. However, securing an alternate key store is also critical, as losing the key means that direct access to the data is impossible. If cloud subscribers store the keys themselves, they should keep primary and alternative key stores secure and should have complete control over the keys.

DATA MASKING TECHNIQUES

One of the best techniques to desensitize data is to mask it. Data masking is typically used for Personally Identifiable Information (PII) to meet data privacy and other legislative requirements; it can also be used for other types of sensitive data, such as information about new products or sensitive financial data. Applied correctly, data masking methods can allow the processing of sensitive data on a cloud infrastructure with a lower security classification. This may result in reduced costs, faster processing, and easier-to-use applications, compared to a cloud installation that guarantees very high data security.

Data masking requirements are not identical even across regional entities that share a common governing framework (for example, the EU Data Privacy Directive). Therefore, special care should be taken when implementing these methods in order to comply with local requirements. Data masking is the sole responsibility of the cloud subscriber.

The following data masking methods exist, in descending order of security level:

- **Synthetization.** Data are synthetically generated without originating from a specified source and have no direct connection with reality. This technique is relevant only for test data.
- **Anonymization.** Personal data are changed in such a way that the revised details can no longer—or only with a disproportionate investment of time, cost, and labor—be attributed to the original real data. For example, personal data cannot be associated with an identified individual. A disproportionate investment is defined by it being easier for the responsible party to collect the data again rather than de-anonymize the data.
- **Masking and tokenization.** Also called pseudonymization, these techniques replace relevant identification features, such as name, with different features—tokens—or character strings. Typically, tokenized data uses a rule or table that can translate the pseudonymous data back to the original data. Therefore, pseudonymous data is not the same as anonymous data; pseudonymous data is typically still in the scope of statutory data protection regulations or other regulatory requirements, such as auditing requirements. However, handling of pseudonymous or tokenized data is often much simpler because of relaxed security and legislative requirements.

Each of these data masking methods has its advantages and disadvantages. Special care should be taken to correctly apply the methods so that an adversary is not able to restore the original data, such as by context analysis[12]. This is especially true for very large linked data sets where it may be possible to combine the anonymized or tokenized data with data from other sources to remove the protection offered by these techniques.

[12] Correctly applying anonymization and tokenization methods does not guarantee that re-identification cannot occur, if an adversary possesses additional information about the data. Additional techniques, such as K-Anonymity and L-Diversity, can decrease the likelihood of re-identification in these situations.

The following aspects should be taken into consideration when applying these data anonymization techniques:

- Synthetic, anonymized, or masked data should be specially labelled to quickly differentiate it from real data.
- Methods used for anonymization and tokenization should be trustworthy, documented, implemented securely, and open so that they can be verified by other persons or third parties if necessary.
- In case of tokenization, the tokenization should be applied as early as possible but at least before transferring the data to the cloud.

Anonymization and tokenization methods can be applied to data before it is transferred to the cloud infrastructure of the cloud provider (typically IaaS and PaaS). They can, however, also be used "on-the-fly" using a proxy-type system (especially useful in the case of SaaS). Such a proxy must be application-aware so that the application functions correctly. Commercial tools are already available for different applications.

If correctly applied, these methods are intended to allow the processing of data in cloud environments even when the cloud environment is not on the same level of confidentially protection as required by the original data. As always, cloud subscribers and cloud service providers should check local legislative requirements and consult with relevant authorities to ensure they fully comply with all data security requirements.

Data Anonymization and Tokenization Methods

The following list explains some of the methods that can be used to achieve anonymization or tokenization (samples are provided in Appendix A: Examples for Data Anonymization and Tokenization). When selecting a specific method, the actual content of the data should be taken into consideration so the most secure method is applied to protect the data from adversaries.

- **Aggregation.** Abstracting individuals to larger groups, and especially relevant to anonymizing PII. Aggregation should be implemented in such a way that at least five or more individuals are summarized to a group[13]. This provides additional protection if an adversary tries to re-acquire the original data.
- **Non-disclosure.** Removing all or parts of the data. For example, the street address and postal code can be deleted from an address value, leaving only the city.
- **Masking or substitution.** Replacing the original data value with a constant or changing value, a character, or a constant or changing character string. For example, "Max Mustermann" could be replaced with "John Doe." Tables that map the substitutions should never leave the premises of the cloud subscriber.
- **Mixing.** Also referred to as shuffling. The data is scrambled or swapped across all existing data values in a particular database column. The mixing should be based on a random distribution that attributes the (partial) data from each data value to another data value. If this used rigorously, this procedure removes, for example, the reference to an individual. Mixing can be used only when there are at least six data values[14] and duplicate assignments are ruled out.
- **Variance.** Distorting personal data based on figures or dates such that the numerical values are randomly increased or decreased in defined variance intervals.
- **Cryptographic methods.** Provides strong protection, especially for large data sets where unauthorized re-attribution must be prevented. For best results, use up-to-date, secure algorithms and ciphers. Strong protection also depends on correctly implementing cryptographic methods, such as providing sufficient randomness. The following aspects are important when applying cryptographic methods:
  - **Character type retention.** The field type must be preserved. That is, alphanumeric characters are mapped to alphanumeric characters.
  - **Length retention.** The field size must be preserved. That is, the length of the mapped data is the same as the original data. Special care must be taken because original data may be revealed; for example, certain word lengths may be significantly sporadic.
  - **Absence of collision.** Two individual data values should not be mapped to the same cryptographic mapping (especially important for tokenization).
  - **Uniqueness.** One set of original data is always mapped to the same token/mapped item.

[13] Five or more helps ensure basic protection against disaggregation attacks.

[14] Six or more helps ensure that a minimum set of entropy is guaranteed for the mixing process.

SECURITY INFORMATION AND EVENT MANAGEMENT

There are two distinct uses for security information and event management (SIEM) in the provider assurance (PA) usage model, each with different data protection requirements.

- **Meeting multiple requirements.** SIEM tools can be the audit and reporting focal point that allows a provider to prove that they are meeting different service-level agreements (SLAs) relating to security assurance.
- **Specific SIEM-related requirement.** A SIEM tool can be used to meet the specific SIEM-related PA model requirement.

Using SIEM to Monitor Multiple Requirements

Table 7 shows some of the usage model requirements that can be verified leveraging SIEM tools.

Table 7. Provider Assurance Usage Model Requirements that Could Be Monitored using Security Information and Event Monitoring Tools

Provider Assurance Usage Model Requirement	Provider Assurance Level
Antivirus and malware protection (with definition updates within 24 hours)	Bronze
Secure protocols used for remote administration (e.g., SSL, SSH, RDP, etc.)	Bronze
All default passwords and guest access removed	Bronze
Network intrusion prevention; updates applied within 48 hours	Silver
Event logging for all administration-level events (requires controlled access to logs)	Silver
Four-eye principle for key administrator changes	Silver
No administrative access for cloud provider staff	Platinum

The cloud provider should have a dedicated SIEM system that receives events from its cloud management platforms and back-end systems. The SIEM system should be on separate physical equipment and data storage areas than those used to deliver SIEM services to subscribers. The subscriber should not have access to the back-end systems used by the provider's SIEM system. In addition, the provider should use industry best practices to secure the data on the SIEM platform.

Providers can deliver SIEM data to subscribers for audit purposes in multiple ways:

- E-mailed reports on a scheduled basis as stated by the PA usage model
- Access to data in a security portal as stated by the Security Monitoring usage model
- Access to data using API as stated by the Security Monitoring usage model

In the Bronze, Silver, and Gold levels of assurance, the provider manages the SIEM process and sends data to the cloud subscriber. In the Platinum assurance level, the subscriber has access to and is able to manage their SIEM. In the lower service levels, there may be a SIEM device on a similar security management platform specifically focused on subscriber security events. This helps to prevent the subscriber from having access to back-end security data about the provider's platform.

BACKUP, ARCHIVING, AND DELETION

Backup, archiving, and eventual deletion of data are important aspects of the data life cycle, and are subject to data security requirements just like other stages such as store and use.

Backups and Archives

The archiving of data must fulfil the cloud subscriber's underlying business needs and respective legal requirements. If backup has been requested by the cloud subscriber, the cloud service provider should check that the requested backup meets all the requirements of security and isolation defined in relevant sections of the Provider Assurance Usage Model.[15] The requirement to deliver backup of cloud-based data must be identified during contractual notifications. In particular, key management is very important, because without the keys, archived or backup data is useless.

Business Continuity Management

Business continuity management (BCM) is a management process that identifies potential impacts that threaten an organization and provides a framework for building resilience and the capability for an effective response which safeguards the interests of its key stake holders, reputation, brand and value-creating activities. If data or systems must be reinstantiated, it is imperative that the necessary encryption keys also be backed up and available.

Secure Deletion

What constitutes deletion depends on how the data is stored.

- If the backup is stored in an encrypted form, deleting all copies of the encryption key(s) will constitute a full deletion of the data stored (also referred to as cryptographical shredding).
- If the data is not stored in an encrypted mode, any backup copies of data must be stored and deleted according to the data deletion requirements defined in the contract.
- If the data is stored in a multi-tenant environment, the backups of data must be maintained in a way that allows deleting individual subscriber data on either a timed basis (such as after six months) or on a by-request basis.

[15] www.opendatacenteralliance.org/docs/ODCA_ProviderAssurance_Rev.%201.1_Final.pdf

RFP REQUIREMENTS

The ODCA believes the following requirements should be included in requests for proposal (RFP) to cloud providers.

Cloud Subscriber Requirements

- **Classification.** Understand information classification and include any requirements that need to be incorporated to implement the standard classification described in this framework.
- **Data security life cycle.** Understand how the existing data security life cycle impacts data moving to the cloud and what policy changes are required.
- **Rights assignment.** Understand how to manage the rights for users, document which rights are given to various user types, and define integration requirements.
- **Legal compliance.** Decide on the type of regulatory compliance to be aligned and whether there are data sovereignty issues as discussed in the "Open Data Center Alliance Usage: Regulatory Framework."[16]

Cloud Provider Requirements

- **ODCA principle requirement.** Solution is open, works on multiple virtual and non-virtual infrastructure platforms, and is standards-based. Describe how the solution meets this requirement and any limitations it has.
- **ODCA Security Provider Assurance Usage Model 1.1.** Solution must allow assurance levels to be represented and tracked in the SIEM system.
- **Backup and archiving.** Describe how the service offered deals with the "right to be forgotten" on all levels including backups and archives.
- **Legal compliance.** Demonstrate how the services being offered assist the cloud subscriber with efforts to comply with their legal requirements. Also demonstrate how the services assist with compliance with data encryption laws in the country or countries of operation.
- **Data protection.** Current data protection policies should be compared with the risks of moving data to the cloud. Identify any additional data encryption or protection methods necessary for a particular use case.
  - **Encrypted communication between cloud provider and cloud subscriber:** Implemented as defined earlier in this document, such as by using data transit requirements. Typically implemented for external private clouds independent of the service tier.
  - **Strong encryption mandatory for all data in transit and data at rest:** For VMs and persistent data, encryption is the responsibility of the cloud subscriber. For the cloud management API, refer to the secure protocols discussed earlier in this document, in the "How Data Confidentiality Relates to Encryption" section.
- **Data masking.** Clearly articulate where data masking can be done and how the cloud provider will manage it.
- **Secure archive and deletion.** Provide transparency by making the secure archive and deletion auditing report available. Also enforce the secure transfer of data in case of archival on the cloud subscriber's premises.
- **Data availability.** Establish a clear service-level agreement and regularly publish metrics to verify that data availability requirements are being met.
- **Data integrity.** The requirement should prevent malware or any other process from modifying the data. Establish methods to address any data integrity issues.
- **Physical segmentation of hardware.** Provide isolation from all other systems for hardware, such as servers, storage, and network, as defined in "Open Data Center Alliance Usage: Security Provider Assurance."[17]

[16] www.opendatacenteralliance.org/library

[17] www.opendatacenteralliance.org/library

SUMMARY OF INDUSTRY ACTIONS REQUIRED

To make this data security framework easy to understand, implement, and use within the cloud, the ODCA has identified specific areas where there should be open specifications and formal de facto standards. The ODCA will be working with the industry to evaluate and recommend specifications that will be included in future releases of this document.

The following are industry actions required to refine this framework into a formal usage model. Some of these actions affect both the cloud provider and the cloud subscriber. Others are specific to one role or the other.

Industry-wide Actions

- Data security must comply with country-specific legal requirements. These requirements and their implications need to be clearly comprehended by cloud service providers as well as by cloud subscribers.
- Transparency is very important. Cloud providers should enable ways to provide event reporting and metrics in easier and more accessible ways. This could include publishing the metrics themselves, to illustrate protection of cloud subscriber data and the avoidance of data breaches and data leakage within the cloud environment.

Cloud Provider Actions

- Cloud providers and other interested parties are requested to submit input on the proposed data security criteria for the various assurance levels (bronze, silver, gold, and platinum).
- Cloud providers need to clearly communicate how their offerings align with this data security framework.
- Cloud providers are requested to submit early implementations of this framework for consideration and evaluation.

Cloud Subscriber Actions

- Cloud subscribers should examine their enterprises and understand the data security life cycle; then they should validate their findings by comparing them to the RFP questions.
- Cloud subscribers should further develop the usage of annonymization and data masking for data in the cloud.

APPENDIX A: EXAMPLES FOR DATA ANONYMIZATION AND TOKENIZATION

Non-disclosure

Original Data Set	Data Set after Non-disclosure
John Mustermann	*****
Hippelstraße 9a 81827 München	München
089/2342 4223	089

Masking/Replacement

Original Data Set	Data Set after Masking/Replacement
John Mustermann	Max Doe
Hippelstraße 9a 81827 München	1 Main Street, 12345 Mustertown
089/2342 4223	0123/4567 8901

Mixing

Original Data Set				Data Set after Mixing			
First name	**Last name**	**Age**	**Weighting**	**First name**	**Last name**	**Age**	**Weighting**
John	Mueller	23	55	Sepp	Maierxx	81	65
Lana	Maier	42	66	Lily	Muellerxx	30	70
Sepp	Schulze	81	70	John	Schmittxx	23	66
Lily	Schmitt	30	65	Lana	Schulzexx	62	55

Variance

Original Data Set	Data Set after Variance
Date of birth: July 21, 1969	Date of birth: Oct. 6, 1970
Salary: EUR 23,000	Salary: EUR 24,123
Weight: 80kg	Weight: 70kg

Other Examples

Attribute	Method	Old Value	New Value
Number	Re-assignment of the last four characters	1234567890	1234561111
		212345678	212341234
	Replacement of customer number with random numbers. Since this can give rise to new valid customer numbers, they must be additionally labeled.	1234567890	47110815nn[18]
		212345678	24711815nn[18]
	Use of a variance, e.g., ± ≤10%	EUR 23,000	EUR 25,100
		EUR 24,000	EUR 23,900
		EUR 25,000	EUR 26,800
	Deletion/Non-disclosure	EUR 25,000	(ZERO)
Name	Re-assignment using a mapping table	Mueller	Lang_Muster
		Lang	Sousa_Muster
	Replacement using fixed names	Mueller	Mustermann
		Lang	Mustermann
	Replacement of fixed names with serial numbers	Mueller	Mustermann_01
		Lang	Mustermann_02
Zip code	Re-assignment of last two digits using an implementation table	01129	01111
	Replacement of the last two digits with a fixed value	01129	01199
		39114	39199
Date of birth	Setting day and month to a fixed value	21.06.1969	01.01.1984
	Setting the date of birth to a fixed dummy value	21.06.1969	01.01.1111
E-mail address	Deletion/Non-disclosure	john@example.org	(ZERO)
	Replacement with a fixed dummy value	john@example.org	max@muster.de
Religion	Replacement with a fixed dummy value	Catholic	Dummy religion

[18] nn represents random numbers.