cloud security alliance℠

CLOUD CONSUMER ADVOCACY QUESTIONNAIRE AND INFORMATION SURVEY

CDG™ CLOUD DATA GOVERNANCE PROJECT

INTRODUCTION

The Cloud Data Governance (CDG) Working Group has been constituted to provide more detailed research deliverables based upon "Domain 5: Information Management and Data Security" within the CSA publication "Security Guidance for Critical Areas of Focus in Cloud Computing." The CDG Working Group as currently chartered will focus on two key phases of Data Governance research:

1. **Survey Phase:** Understanding the top requirements and needs of different stakeholders on governing and operating data in the cloud, and

2. **Best Practices Recommendation Phase:** Prioritizing and answering of the key problems and questions identified by cloud stakeholders in Phase 1.

As the leading non-profit organization educating and promoting vendor-neutral best practices for cloud computing Security, the Cloud Security Alliance (CSA) is in the ideal position for this data-oriented research. Our team is led by CSA Singapore and CSA Silicon Valley chapters, in collaboration with CSA Global. All CSA members are welcome to contribute to this working group.

The editorial team and committee consists of a mix of practitioners and researchers in the area of cloud data protection and governance. The Working Group aims to publish a white paper featuring different stakeholder groups' concerns over a quarterly publication cycle. Together, these white papers will contribute to the main CSA Guidance documentation's future versions.

The Cloud Data Governance Working Group is supported by the Cloud Security Alliance. If you are interested in contributing to this vendor-neutral project, please contact us.

The permanent and official location for the Cloud Security Alliance Cloud Data Governance research is:

http://www.cloudsecurityalliance.org/research/working-groups/cdg

FOREWORD

Welcome to the Cloud Security Alliance's Cloud Data Governance Cloud Consumer Advocacy Questionnaire and Information Survey (CCAQIS) Preliminary Survey 1.0. This is one of many research deliverables CSA will release in 2011.

The Cloud Data Governance (CDG) Working Group within the Cloud Security Alliance (CSA) has been designated to provide research and guidance for all aspects of data and information in the cloud. The CDG Working Group Co-Chairs are responsible for governance and oversight of the CDG Working Group. The efforts are jointly executed by CSA Global, CSA Singapore Chapter, CSA Silicon Valley Chapter and relevant working groups responsible for authoring CSA's Guidance reports. The Cloud Data Governance Working Group has been formed to coordinate research and the execution of this work.

We would like to extend our thanks to Trend Micro for their sponsorship of this important research.

Best Regards,

Jerry Archer	Alan Boehme	Dave Cullinane
Nils Puhlmann	Paul Kurtz	Jim Reavis

The Cloud Security Alliance Board of Directors

Sponsored by:

ACKNOWLEDGMENTS *(In alphabetical order)*

Advisors

Aloysius Cheang, CSA Global

Jim Reavis, CSA Global

Luciano "JR" Santos, CSA Global

Co-Chair / Editors-In-Chief

George Goh (CSA Singapore)

Srinivas Jaini (CSA Silicon Valley)

Ryan Ko, PhD (CSA Singapore)

Richard Lim (CSA Singapore)

Tim Mather (CSA Silicon Valley)

Managing Editor / Researcher

John Yeoh, CSA Global

Design / Development

Kendall Cline Scoboria, Shea Media

Evan Scoboria, Shea Media

PURPOSE

Cloud Computing marks the decrease in emphasis on "systems" and the increase in emphasis on "data." With this trend, Cloud Computing stakeholders need to be aware of the best practices for governing and operating data and information in the Cloud, taking into consideration organizational risk, compliance, and IT service level requirements.

This is in line with the concerns highlighted by Section II (Domain 5: Information Lifecycle Management) in the Cloud Security Alliance (CSA) Guidance v2.1.

The purpose of this survey is to capture the current state of data governance and data security capabilities offered by leading cloud service providers in the industry. The results of this survey are intended to be used for guidance and research conducted by CSA and its affiliates.

Standardization of data governance policies for cloud environments will accelerate rapid adoption of cloud computing across the industry verticals.

BENEFITS FOR THE CLOUD COMMUNITY

With the exponential increase in data deposited in cloud environments (both public and private), research in the area of data, information, and knowledge stored and processed in the cloud is timely. Data is stored in many different forms, and processed in a myriad of methods. There is a need for an authoritative voice in making sense of the key concerns with data storage and processing techniques. There is also an urgent requirement to align current practices with governance, risk and compliance regulations.

DATA DISCOVERY

↘ Does the Cloud Service Provider (CSP) provide a capability to locate and search all of a customer's data?

ANSWER OPTIONS	RESPONSE PERCENT
Yes	59%
No	41%

↘ If yes, is this a supervised search capability or an unsupervised search capability?

ANSWER OPTIONS	RESPONSE PERCENT
Supervised search capability	60%
Unsupervised search capability	27%
Both	13%

The responses received for the technical details of the search capabilities are:

- Searches are often guided by 1) key-words, 2) category, 3) country/location, and/or, 4) related content.
- This is delivered through the data domain management and knowledge management systems. Some products allow either supervised or unsupervised searches.
- Full text search feature for content and contacts is available to each authenticated user. All searches have to pass an authorization check.
- The search capability have to comply with statutory and regulatory obligations, and satisfy legal law enforcement access requests.
- Using a simple key-based object store. When you store data, you assign a unique object key that can later be used to retrieve the data. Keys can be any string, and can be constructed to mimic hierarchical attributes.
- Using a cloud-based relational database service built on server technologies. It provides a highly available, scalable, multi-tenant database service that enables easy provisioning and deployment of multiple databases. High availability and fault tolerance are built-in and no physical administration is required.

There were also some responses for the scope of the search results:

- The data should only be for the particular customer. The data of one customer should not be exposed to another customer.
- Optional services that clients can leverage to search through databases, email and other data.
- Search all stored data into non-relational storage to review policies
- All data is stored in a single relational database or the SAN file system

- Some of the respondees mentioned that search is highly contextual and domain-dependent. For example, a respondee mentioned the provision of industry-required reports of the desired data controlled by predefined SQL Procedures. The respondee's company does not allow direct database access from the end-user's point of view.

WORKING GROUP RECOMMENDATION

In this sense, it can be seen that search is a context dependent feature, which heavily depends on policies (both internal and external).

There is also a requirement to consider the scope of the search, the scope of the user's authorized access to information, and the scope of the storage of the data results or data sources.

With the proliferation of data storage technologies in cloud environments, a quintessential need is born to address key concerns such as data discovery. Data discovery helps in managing customer data in the cloud, as well as in allowing other processes such as classification of data to help in governance, risk and compliance.

LOCATION OF DATA

↘ Does the CSP allow a customer to select a specific location for the use and/or storage of the customer's data?

ANSWER OPTIONS	RESPONSE PERCENT
Yes	82%
No	18%

↘ Does the CSP provide any technical enforcement to prevent a customer's data from moving through or to a customer proscribed location?

ANSWER OPTIONS	RESPONSE PERCENT
Yes	73%
No	27%

DEFINITION AND SCOPE OF CLOUD DATA GOVERNANCE

Governance and the management of cloud data is important. In our perspective, Cloud Data Governance is a discipline involving the processes, roles and technologies for managing and governing data in cloud computing environments.

To the best of our knowledge, there is currently no large emphasis on scoping of the governance of data. Typically, the research in the area of Cloud Data Governance revolves around cloud accountability and cloud privacy.

↘ Does the CSP allow a customer to select a separate, specific location for the back-up or replication of data that still meets any customer restrictions on the nation-state level of location restrictions?

ANSWER OPTIONS	RESPONSE PERCENT
Yes	65%
No	35%

The cloud provider's ability to provide information about pin-pointing location of data is not a good-to-have feature, but a need. Data location has become an important differentiating factor in choosing a cloud provider.

DATA AGGREGATION AND INFERENCE

↘ Does the CSP provide customers with controls over its data to ensure that data can or cannot be aggregated according to customer needs and/or restrictions?

ANSWER OPTIONS	RESPONSE PERCENT
Yes	58%
No	42%

The responses include the following controls to ensure control of aggregation of data according to customers' needs and/or restrictions:

- Hypervisor level controls exist and are documented in the hardening guidelines

CLOUD DATA

In the perspective of our working group, cloud data is defined as data that at any point in its life cycle exists within the scope of a cloud computing environment. There are many perspectives to viewing cloud data.

One perspective is to view what the data describes. Examples include:

- Cloud user data
- Data about cloud service provider entities (e.g., roles, assets, etc)
- Data about cloud service provider processes (e.g. computing programs, Web services, logs, etc)
- Metadata describing the above (e.g. rules describing what can and cannot be done to the data, descriptions about the categorisation of data).

- Storing customer data in logical units. Other users then can be invited by users who have the "invite" privilege.
- Authorization schemes with different level roles which controls file-specific privileges to govern access control to data exchanges, data functions and their content.
- Logical access control including change managements while granting or revoking any access.
- All data retrieval is controlled from predefined SQL [stored] procedures – therefore only the data defined for the purpose defined can be collected and it can only be outputted as defined by the application.

WORKING GROUP RECOMMENDATION

Not a large majority of CSP's provide customers with control over data aggregation and inference. This shows a possibility of low awareness or low level of maturity here and merits deeper research into this area.

↘ In cloud databases, what mechanisms are provided for the customer to determine what columns are encrypted and to prevent inference from non-encrypted columns?

We received responses which mentioned the following mechanisms/ techniques:

- Product encryption
- Some products allow data to be encrypted before it is persisted in the cloud. This allows finer control on data encryption.
- All customer content (files uploaded to the exchanges) is encrypted on disk. Database columns are not encrypted.
- Full disk encryption.
- A vendor specified certain encryption functions that need to be employed as a precursor to data-migration to the cloud.
- Only the bucket and object owners originally have access to product resources they create.
- Client responsibility

WORKING GROUP RECOMMENDATION

While there are techniques addressing the need of encryption of the storage or the partial encryption where select portions of data are encrypted, there is an absence of mention on the prevention of inferences of semantics or context of sensitive data from non-encrypted columns. Cloud providers, in this area, do not offer the robustness and complexity that is currently required by the customers. The absence of mechanisms to provide customers with information on columns encrypted and the prevention of inference from non-encrypted columns is a critical challenge that needs to be addressed. This critical need calls for further research in this area.

↘ Does the CSP provide the ability to mask data from selected customer personnel, as determined by a customer, to prevent data aggregation or inference problems for a customer?

ANSWER OPTIONS	RESPONSE PERCENT
Yes	65%
No	35%

As disparate data sets move into the cloud, the robustness of features and frameworks offered by a cloud service provider in data management encourages customers to implement best practices and mitigate risks.

COMMINGLING DATA WITH OTHER CLOUD CUSTOMERS

↘ What technical enforcement mechanisms does a CSP

use to prevent the commingling of data with other cloud users?

The mechanisms include:

- Partitioning, encryption, VLAN segregation, user restrictions, access rights
- Exchanges having unique user identification (ID) numbers and full authentication (login) service – the users see the list of exchanges they are authorized for and tasks awaiting for their attention.
- On the server storage side, all files are encrypted by unique data keys and issued unique file IDs, which are tied back to the exchanges and users who have authorization to view/delete.
- Keeping separate virtual environments, i.e. "containers," from other cloud users.
- Either dedicated storage devices, or we use the logical controls available to ensure data separation
- Data classification or data tagging
- Each customer's data is encrypted with a unique key for that customer. This uses the Advanced Encryption Standard (AES) 256-bit encryption
- Multi-tenancy is accomplished using discrete business units, and roles-based permission models
- Customers can create access log records for all requests made against it. These access log records can be used for audit purposes and contain details about the request, such as the request type, the resources specified in the request, and the time and date the request was processed.
- Multi-tenancy leakage controls in place based on virtualization separation.
- Using domain segregation.

WORKING GROUP RECOMMENDATION

The above responses tells us that the typical mechanisms are:

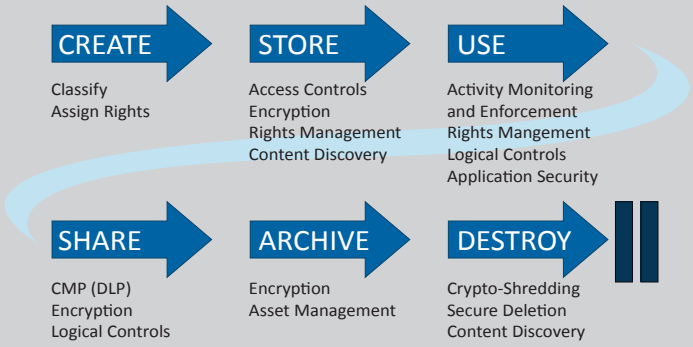
- Encryption of data for different users
- VLAN segregation
- Partitioning
- User/access rights controls
- Data classification and tagging

There is a need to further investigate the effectiveness of these techniques and their limitations.

↘ If the CSP is using data tagging, are those tags cryptographically signed?

CLOUD DATA LIFE CYCLE

Because of the definition above, cloud data evolves around the following cycle:



ANSWER OPTIONS	RESPONSE PERCENT
Yes	39%
No	61%

USE OF DATA SECURITY CONTROLS

↘ Does the CSP adhere to any established governance framework(s) involving data security controls?

ANSWER OPTIONS	RESPONSE PERCENT
Yes	74%
No	26%

↘ If yes, does the CSP undergo any regular (e.g. annual) 3rd party audit(s) for compliance with any established governance framework(s)?

ANSWER OPTIONS	RESPONSE PERCENT
Yes	79%
No	21%

↘ Does the CSP allow customers to audit the CSP's data security controls?

ANSWER OPTIONS	RESPONSE PERCENT
Yes	67%
No	33%

↘ What mechanisms does the CSP provide for customers to define access to their data?

Some of the mechanisms are:

- Security Content Automation Protocol (SCAP) control implementation through a third party cloud security vendor
- Logging mechanisms
- IP ranges controls
- Time stamp access, which is cryptographically signed.
- The use of IAM policies, bucket policies, ACLs and query string authentication
- Access control systems relating to active directory policies, servers, database, administrator access management.
- Mechanisms used are case-by-case depending on the data's level of sensitivity

WORKING GROUP RECOMMENDATION
The abundance in mechanisms from the survey results demonstrate an acceptable level of maturity in considering this aspect of cloud data governance.

CSA's Cloud Controls Matrix and similar governance standards provide a much needed guidance to establish and enforce security controls specific to cloud environments.

ENCRYPTION AND KEY MANAGEMENT PRACTICES

↘ Does the CSP provide end-to-end encryption for data-in-transit?

ANSWER OPTIONS	RESPONSE PERCENT
Yes	84%
No	16%

↘ Does the CSP offer encryption to its customers to use for data-at-rest?

ANSWER OPTIONS	RESPONSE PERCENT
Yes	69%
No	31%

↘ If the CSP does offer encryption to its customers to use for data-at-rest, then does the CSP use formally vetted encryption algorithms (e.g., under NIST's FIPS 140-2)?

ANSWER OPTIONS	RESPONSE PERCENT
Yes	75%
No	25%

↘ If the CSP uses formally vetted encryption algorithms, under what specific program(s) have these encryption algorithms been vetted?

Many services are compliant with various certifications and third-party attestations. These specific programs include:

- NIST SP 800-67
- NIST Special Publication 800-53[1]
- NIST FIPS 140-2[2]
- Advanced Encryption Standard (AES)-256
- DoD Information Assurance Certification and Accreditation Process (DIACAP)
- OpenPGP
- SAS 70 Type II[3]
- PCI Data Security Standard (DSS) Level 1
- ISO 27001

WORKING GROUP RECOMMENDATION
The current level of maturity in the vetting of encryption algorithms are satisfactory, and well covered.

↘ If the CSP does offer encryption to its customers to use for data-at-rest, then how is (cryptographic) key management handled (i.e., by the CSP or by the customer?)

Most of the responses mentioned that key management is typically handled by the CSP. However, with that said, a few of the respondents highlighted the following:

- The final decision on cryptographic architecture is actually being left to the customer.

- Key management is handled in a multi-layer system (e.g. data keys unique to each file and a master key to protect the data keys).
- If using CSP's server-side-encryption feature, key management/rotation is handled by CSP, if using client-side encryption the customer is in charge of managing/rotating the keys.
- The customer controls and manages keys for VPN, storage and archival products.
- Customers provides public key to the CSP and private keys are maintained by the customer.

WORKING GROUP RECOMMENDATION

From the responses, it was obvious that this is an area where encryption is a point of emphasis with CSP's. However, the key management implementations are highly dependent on the provider and need to be vetted carefully according to tenant needs.

DATA BACKUP AND RECOVERY SCHEMES FOR RECOVER AND RESTORATION

↘ Does the CSP offer data back-up and recovery services for customers?

ANSWER OPTIONS	RESPONSE PERCENT
Yes	88%
No	12%

↘ If yes, is the specific location for such selectable by the customer?

ANSWER OPTIONS	RESPONSE PERCENT
Yes	53%
No	47%

DATA REMANENCE AND PERSISTENCE

↘ How does the CSP handle the issue of data remanence or persistence and which method(s) does a CSP utilize to ensure that removed data is indeed removed?

A minority of the responses are unsure of any existing methods. On the other hand, some of the responses mentioned the following interesting options:

- Auditing by a few parties (including 3rd party) to ensure the proper and eventual removal of data.
- As all data at rest is encrypted, it is deleted upon customer request.
- A mature decommission process that involves DoD specified overwrite processes and independent verification of this process by an audit team
- Schedules that control the data lifespan or the length of permanency of data in the storage.
- Data is associated with products. Depending on the definition of how that product is designed to work, there may be schedules that control data lifespan or the data may be designed to be permanent.
- The physical destruction of physical media storing the data.
- Multi-pass disk overwrite under de-provisioning automation (varies according to client requirements).

WORKING GROUP RECOMMENDATION

The results show that the area of eventual and assured deletion of user deleted data in the cloud is maturing. Traditional approaches such as (1) policies determining lifespans of data usability and (2) the physical destruction of data commonly seen in industries handling sensitive data have been adopted.

↘ Does the CSP's method of handling data remanence or persistence meet any identified standard(s)?

ANSWER OPTIONS	RESPONSE PERCENT
Yes	33%
No	67%

For CSP's that answered yes, some of the specific standard(s) adhered to for addressing data remanence were:

- DoD, NIST, ISO 27000
- If required, it can meet government standards
- Depending upon the data, there may be legal requirements to retain data such as call details records and data relate to billing transactions.

WORKING GROUP RECOMMENDATION

The fairly low percentage of CSP's meeting any identified standards, coupled by a lack in clarity of, show us possible

clues that CSP's methods of handling data remanence or persistence mostly do not meet or even consider any identified standards. This gap should be addressed and further researched.

↘ What guarantees does a CSP provide for the timeliness of the removal of data?

From the results of the survey, it was shown that there is generally an absence of clear guarantees of the removal of data. While some mentioned guarantees being a best-effort level of commitment, there were responses which show that the guarantees are based on the following:

- Service-level agreements (SLAs)
- Master service agreements with the customer
- Contract guarantees
- Services organization archive and mark data for deletion upon customer request. An automated process purges data marked for deletion with routines that cleanup if the regular process fails.

WORKING GROUP RECOMMENDATION

This shows that guarantees of removal of data is unclear and non-uniform across the cloud-service providers. There is much room for the industry to self regulate these guarantees using international bodies to standardize and/or the local governments to discuss proposals mandating clear guarantees for consumers on the actual removal of data.

As enterprises move to the cloud to take advantage of the ephemeral nature of cloud computing, one of the key concerns is what happens to data at decommission time. Standardization of data decommissioning best practices will help in building a trusted cloud.

CLOSING REMARKS

The aim of this initial survey is to enable all Cloud stakeholders to understand the current state and maturity of cloud service providers in the area of data in the cloud. While some areas are well developed,

MATURE AREAS

- Control over aggregation of data
- Vetting of encryption algorithms
- Define access to their data
- Technical enforcements of multi-tenancy

some were identified as lacking in capabilities and solutions.

IMMATURE AREAS

- Timeliness of removal of data
- Cryptographic key management scalable to cloud computing
- Methods for handling data remanence or persistence
- Data remanence or persistence and which method(s) does a CSP utilize to ensure that removed data is indeed removed
- Mechanisms provided for customers to determine which columns are encrypted and to prevent inference from non-encrypted column(s)

These issues pave the way for the next phase of the CDG Working Group research.

We are eager to hear your feedback regarding this initial report and look forward to your participation in future CDG Working Group research initiatives. If you found the results helpful or would like to see the report improved, please consider joining the CDG Working Group as a contributor. For more information please contact csa.datagovernance@gmail.com.

On behalf of the Cloud Security Alliance and the Cloud Data Governance (CDG) Working Group we would like to thank each and every Subject Matter Expert for their time and effort that was put into responding to this survey.

FOOTNOTES

[1] As of release, V3 is the current version. V4 is close to release (December).

[2] As of release, FIPS 140-2 is the current version. 140-3 is close to release.

[3] SAS 70 is no longer applicable, as it has been replaced by SSAE 16 as of June 2011.

