Project Title	Trust-aware, Reliable and Distributed Information Security in the Cloud
Project Acronym	TREDISEC
Project No	644412
Instrument	Research and Innovation Action
Thematic Priority	Cybersecurity, Trustworthy ICT
Start Date of Project	01.04.2015
Duration of Project	36 Months
Project Website	www.tredisec.eu

<D1.5: INNOVATION STRATEGY AND PLAN>

Work Package	WP1 Management
Lead Author (Org)	Ghassan Karame (NEC)
Contributing Author(s) (Org)	Beatriz Gallego-Nicasio Crespo (ATOS)
Reviewers	Jose F. Ruiz (ATOS), Melek Önen (EURC), Matthias Kohler (SAP)
Due Date	M3
Date	30.06.2015
Version	0.8 (Final)

Dissemination Level

X	PU: Public
	PP: Restricted to other programme participants (including the Commission)
	RE: Restricted to a group specified by the consortium (including the Commission)
	CO: Confidential, only for members of the consortium (including the Commission)

The work described in this document has been conducted within the project TREDISEC. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 644412. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of its content.

Versioning and contribution history

Version	Date	Author	Notes
0.1	26.05.2015	Ghassan Karame (NEC)	Initial version
0.2	28.05.2015	Ghassan Karame (NEC)	Minor adjustments
0.3	02.06.2015	Beatriz Gallego-Nicasio Crespo (ATOS)	New version of section 4
0.5	15.06.2015	Ghassan Karame (NEC)	Overall revision
0.6	23.06.2015	Review from WP Leaders	Review
0.7	26.06.2015	Ghassan Karame (NEC), Beatriz Gallego-Nicasio Crespo (ATOS)	Addressing review comments
0.8	30.06.2015	Quality Manager (ATOS)	Quality Check

Disclaimer

Table of Contents

List of Tables

List of Figures

Executive Summary

This document establishes the strategy, processes, milestones and role assignments to ensure an innovation-driven research and development in the TREDISEC project.

The document includes an early assessment of the relevant technologies in order to serve as input to the work packages action plan. Based on this assessment, we identified 7 main innovation key aspects that should be monitored in the context of TREDISEC.

We then presented our innovation strategy plan whose main goal is to ensure that the innovation goals of the project do not lose relevance given changing market trends.

Finally, we presented a complete framework that enables the continuous assessment of our identified innovation-related indicators throughout the project's lifetime.

1. Introduction

1.1 Purpose and Scope

One of the main objectives of TREDISEC is to strongly support the exploitation of the research results of the project. This goes beyond the production and dissemination of novel cloud security techniques and concerns mainly the innovation impact of TREDISEC. Indeed, innovation often characterizes an activity which does not only consist of dissemination but especially deals with the exploitation of the R&D results of the project.

Studies [1] have shown that one of the best ways to successfully convert research into innovative success stories lies in ensuring that all cooperating research organizations can prepare themselves for the challenges of market-oriented exploitation. The earlier and more organized is this preparation, the more likely it is to reach a success story.

In TREDISEC, this is ensured through an innovation management scheme which is devised to ensure the successful exploitation of the project results and the conversion of these results into innovative success stories.

In TREDISEC, the main purpose of innovation management is to ensure that the project research activities, technological developments and achievements are kept well connected to outside technology developments. An additional goal of innovation management here is to maintain low risk level for the project and to prevent the project results from losing relevance given the evolving trends in the market.

The innovation management task will be carried out by the Innovation Director (ID) of TREDISEC who will also be supported by the rest of the members of the Executive Board (EB), which includes the Work Package Leaders (WPLs). In what follows, we discuss in greater details the innovative management approach adopted by TREDISEC.

1.2 Structure of the document

This document is structured into 4 main sections and a last **section 5** that concludes the document providing an outlook to the activities and deliverables that will be conducted in the upcoming months, in the context of Task 1.3 Innovation Management.

Section 1 provides an introduction to the Innovation Management process and the approach and activities that will be followed in project TREDISEC.

Section 2 identifies the innovation potential of the project, analysing the current status of the different research topics that the technical work packages (i.e. WP2, WP3, WP4 and WP5) will investigate, and describes the 7 key innovation points of the project.

Section 3 outlines the project Innovation strategy, listing the main activities that will be implemented in the project to ensure meeting the overall objectives of the Innovation Management task.

Section 4 presents an assessment framework that will help TREDISEC project management to monitor and assess the project interim research advances and milestones with regards to three dimensions: Scientific-Technical dimension, Market dimension and Organisational dimension.

1.3 Innovation Management: TREDISEC Approach

The TREDISEC consortium comprises a rich mixture of large industrial players, who have a strong overview of the cloud security market trends. The involvement of these players in the TREDISEC consortium only increases the success rate of project innovation. The enterprises can indeed make a strong difference in successful market-oriented exploitation by ensuring that the research outcomes are in line with their expectations, and are consistent with the current demands of the market. This is further strengthened in TREDISEC by integrating the technology into bundles that can be directly used by service providers.

As mentioned earlier, the communication and management of innovation-related activities will be orchestrated by the ID, who will ensure that all the different entities in the consortium can manage the respective innovation-related risks, and in the event that a risk becomes an actual challenge or threat, the ID will follow an emergency management scheme to alleviate such risks. In Section 3, a management scheme is well-devised according to the expected innovation impact of TREDISEC. It will be continuously updated, and followed up throughout the entire research project lifetime.

Our management scheme includes the following milestones:

- Protecting innovation before disseminating the project results. Innovation management will provide clear and efficient procedures for rapidly protecting new results and agreeing on dissemination, therefore ensuring that no information is published which could be detrimental to the protection of some innovative project results. A formal process is described in D1.1 section 3.3, where the Project Coordinator, the Communication Director, the Scientific Director, the Innovation Director and the WP7 leader, by means of the TREDISEC Press Office, validates any publication or dissemination activity before its actual release to the public. Needless to mention, we ensure that this process does not hinder the publication of the project results at academic venues. Note that academic manuscripts need to contain substantial novelty and originality in order to be eligible for publication. For that purpose, the Press Office will analyse the camera ready version of the publications that any partner wishes to disseminate in academic venues, prior to the actual publication. The Press Office might request the amendment of parts of the document in order to ensure the maturity of the presented results and/or to protect some of the innovative parts of the project. In case an agreement on the amendment was not possible, the ID will call for an urgent meeting of the Steering Board in order to resolve the matter, according to the terms agreed in the Consortium Agreement (section 8.3 Dissemination).
- Specifying “innovation-related activities” such as monitoring, emergency plans, or take-up activities.
- Definition of strategies relating to the granting of licences to third parties or to the identification of potential hurdles for the implementation of the results of the project (e.g., standards or third parties’ patents, etc.)
- Progressively refining the innovation management strategy as market trends evolve. That way, the exploitation of the project results becomes more accurate and aligned with relevant tendencies.
- Identifying and acquiring feedback from different entities and communities (e.g., advisory board body, related projects, cloud and security communities, etc.) to better align the project results with users’ expectations.

1.3 Detailed Activities

Innovation management in TREDISEC is conducted at the WP1 management level.

In this deliverable, we include an assessment of the technologies and available solutions that are relevant to the project and present our **Innovation Strategy and Plan** (ISP). We will use our assessment of the technologies to determine the key innovation points that should be monitored throughout the project. The ISP adapts the innovation process to the project particular context, defines tasks and milestones, assigns roles and responsibilities and provides the guidelines to successfully accomplish the objectives of the project. The ISP will be presented by the ID to the WPLs to count on their engagement and get their approval, since this plan will influence the research and development work during the entire project duration.

By doing so, innovation management will have the ability to influence the initial WP action plans according to the devised ISP.

During the project execution, we foresee the following innovation management activities:

- Continuously monitor market trends to support the definition of the business cases and plan sustainability activities.

- Monitor the WP progresses according to the Innovation Strategy and propose necessary actions to be taken if necessary.
- Regularly assess the innovation level of the project with regards to a set of innovation-related indicators grouped into framework dimensions.

Finally, in the late stages of the project, we plan the following activities:

- Conduct an assessment of the maturity of the project technical results to support the exploitation and sustainability plans.
- Produce a report on the main innovations and achievements of the project and identify candidates for a technology transfer process.
- Provide input to the exploitation and sustainability plan with regards to the procedure to handle potential market opportunities.
- Introduce our solutions to end-users and acquire their feedback on the usability and services offered of the solutions.

2. Innovation potential relevant to TREDISEC

In what follows, we start by assessing the technologies and solutions relevant for the TREDISEC project. We then leverage this analysis to draw conclusions on the expected innovation outputs of TREDISEC. As mentioned earlier, these outputs will be continuously monitored in the devised ISP to ensure that they do not lose relevance in spite of rapidly evolving market trends.

2.1 Assessment of Cloud Security Technologies and Solutions

2.1.1 Storage Security

Within the end-to-end security paradigm, data is encrypted very close to its source at the client side, and the client is the only one in possession of the keys used to encrypt; thus no information is revealed to the cloud provider or other cloud provider tenants. For example, system components such as Microsoft Cluster Shared Volumes [2] and the IBM General Parallel File System (GPFS) provide such a capability. An immediate side-effect of this paradigm is that storage efficiency functions such as deduplication and compression are no longer effective, due to the high-entropy and randomized nature of semantically secure ciphertext. Inverting the order of encryption and data reduction would provide a solution to the problem; however this approach is often not practicable due to the fact that it either (i) requires significant changes to the existing client infrastructure (e.g. in case the storage system on the premises of the customer does not support deduplication) or (ii) because it does not use resources efficiently (e.g. in the case of server-side deduplication, data needs to be transferred to the cloud in any case to check whether a copy has already been uploaded in order to perform deduplication. In this case, the bandwidth cannot be saved).

As encrypted data is pseudorandom, in the presence of end-to-end encryption, most compression techniques underperform at best, and are usually downright ineffective. Compression of encrypted data was investigated by Johnson et al. [3], where a methodology based on distributed source coding theory was proposed to perform compression over encrypted data. However, this method is only efficient for encrypted data in the case of ideal Gaussian source, while the compression rate is reduced for more general and common data distributions such as those encountered in cloud computing scenarios.

Similar arguments can be made for deduplication: semantic security requires that multiple copies of the same datum are indistinguishable, thus avoiding the benefits of deduplication. Trivial solutions, such as forcing users to share encryption keys or using deterministic encryption fall short of providing acceptable levels of security.

Convergent encryption is a cryptographic primitive introduced by Douceur et al. [4], attempting to combine data confidentiality with the possibility of data deduplication. Convergent encryption of a message consists of encrypting the plaintext using a deterministic (symmetric) encryption scheme with a key which is deterministically derived solely from the plaintext. Clearly, when two users independently attempt to encrypt the same file, they will generate the same ciphertext which can be easily deduplicated. Unfortunately, convergent encryption does not provide semantic security as it is vulnerable to content-guessing attacks. Later, Bellare et al. [5] formalized convergent encryption under the name message-locked encryption. As expected, the security analysis presented by the authors highlights that message-locked encryption offers confidentiality for unpredictable messages only, clearly failing to achieve semantic security. Bellare et al. presented DupLESS [6], a server-aided encryption for deduplicated storage. Their solution uses a modified convergent encryption scheme with the aid of a secure component for key generation. Unfortunately, this scheme does not tackle the root weakness of convergent encryption: indeed, in both solutions, a malicious cloud provider is still free to formulate guesses on the ciphertext uploaded by users, and to use the system's interfaces to verify whether the resulting ciphertext matches.

2.1.2 Preventing Abuse of Deduplication

A separate line of work addresses the leakage of information resulting from deduplication; a deduplication-enabled system can effectively be used as an oracle to query whether a particular datum has already been uploaded [7]. For instance, a user can check whether another user has

already uploaded a file by trying to upload it as well and by checking whether the upload actually takes place. This attack was performed in Dropbox[1] and is particularly relevant for rare files that may reveal the identity of the user who performed the upload. Proof of Ownership (PoW) [8] schemes address the root-cause of the aforementioned attacks to deduplication, namely, that the proof that the client owns a given file (or block of data) is solely based on a static, short value (in most cases the hash digest of the file), whose knowledge automatically grants access to the file. PoW schemes are security protocols designed to allow a server to verify (with a given degree of assurance) whether a client owns a file. The probability that a malicious client engages in a successful PoW run must be negligible in the security parameter, even if the malicious client knows a (relevant) portion of the target file. A PoW scheme should be efficient in terms of CPU, bandwidth and I/O for the server and all legitimate clients. In particular, PoW schemes should not require the server to load the file (or large portions of it) from its back-end storage at each execution of PoW. Existing PoW schemes are unfortunately not yet mature enough to be deployed in practice, either because their security cannot be fully analysed or because they are too taxing at the server side or at the client side.

Clearly, storage systems are expected to undergo major restructuring to maintain the current disk/customer ratio in the presence of end-to-end encryption.

2.1.3 Securing Multi-Tenant Environments

Cloud systems are composed of several, often complex software modules: in the presence of vulnerabilities or colluding privileged users, a malicious entity can subvert the correct execution of the system and compromise confidentiality and integrity. Ideally a multi-tenant cloud system serves requests of multiple customers (tenants) in such a way that (i) computing and storage resources are shared among such customers and (ii) this sharing of resources does not weaken system security. In practice, multi-tenancy is a trade-off between security and costs: the wider the subset of resources shared (e.g., same physical machine vs. same OS), the more the cloud system can amortize costs and increase utilization. However, this sharing leads to weaker isolation and consequently higher security risks. A properly designed system can still enjoy a large degree of resource sharing without unduly compromising security. Unfortunately, most cloud storage systems address multi-tenancy by only taking into account application-level security, e.g., by authenticating each request prior to executing e it under a single cross-tenant privilege. While this approach of application-level isolation makes cloud systems flexible and able to serve a large number of tenants, it has the negative impact that if a single malicious user finds and exploits a vulnerability this user may be able to access resources belonging to any user of any tenant. Furthermore, it is often hard to contain attacks on a single component (e.g. management console, storage, computations, etc.).

The use of dedicated security hardware such as Trusted Platform Modules (TPMs) and Hardware Security Modules (HSMs) relaxes some of the constraints of traditional solutions and makes the problem more easily treatable. Indeed, it has been shown that Trusted Execution Environments (TEEs) can be built, effectively allowing the execution of trusted code in untrusted environments. The vision enabled by this approach is that subsets of the cloud may become TEEs and process the data that is otherwise encrypted and thus inaccessible to other tenants or to malicious insiders. Unfortunately, current solutions suffer from bad performance, have scaling problems or require trusting too-large portions of the cloud provider.

The security and usability of multi-user and multi-tenant systems require reliable access control policies and enforcement mechanisms to ensure that legitimate principals always get access to data whilst access is refused to malicious parties. In the cloud, the problem is compounded by its distributed nature and by the more complex trust model: distributed access control mechanisms aim at providing users with advanced access control policies that go beyond the classical notion of centralized ownership (i.e., where each file has a single user, who is the owner of the file and decides the access control policy).

To overcome the limitations that conventional access control models have in multi-tenancy cloud settings, a generalization of RBAC, which is an attribute-based access control (ABAC) developed by

[1] http://www.extremetech.com/computing/179495-how-dropbox-knows-youre-a-dirty-pirate-and-why-you-shouldnt-use-cloud-storage-to-share-copyrighted-files
https://www.dropbox.com/en/help/8

NIST [9], will be used. In ABAC, authorization to perform an action is determined by evaluating attributes associated with the subject, object, and the requested action against a policy that describes the allowable set of actions for a given set of attributes. Jin, et al. [10] have recently proposed a unified formal model of ABAC with the aim of exhibiting the minimal features necessary to accommodate the most prominent access control models, namely, discretionary access control (DAC), mandatory access control (MAC), as well as RBAC.

Perhaps counter-intuitively, when it comes to a storage system, access control rules must include the support for secure data deletion; that is, the rightful owner must be able to instruct the system to destroy any copy of their data, regardless of caching, snapshots, replicated or erasure-coded copies. Traditional solutions (e.g. digital shredding with overwrite patterns) are either widely impractical when we meet the scale of today's cloud storage systems, or are not fine-grained enough, or fail on specific media (e.g. log-structured systems used in modern SSDs). Cryptographic solutions to this problem have been found but are ineffective when combined with storage efficiency functions, and deduplication in particular.

2.1.4 Processing over Encrypted Data

Confidentiality of data requires that when users outsource data, the cloud should not learn any information about the data it is storing and the operations performed over it. Although classical encryption algorithms ensure data confidentiality, they unfortunately prevent the cloud from operating over encrypted data. This makes any serious Database as a Service offering questionable and is the way many traditional DBMS like Sybase, Oracle, DB2 or solutions like Dropbox appear to work when they claim to encrypt data and provide cloud storage. Moreover, the queries issued by the user and the result of the queries should remain confidential to the cloud. Existing crypto primitives such as searchable encryption or private information retrieval cannot immediately be adopted by current cloud solutions.

Several solutions such as CryptDB and related works [11, 12, 13, 14, 15] address this problem with a combination of (i) SQL-aware encryption strategies that map specified SQL operations to "fitting" encryption schemes (e.g. deterministic, order-preserving, partially homomorphic or searchable) supporting the requested functionality and (ii) of onion cryptosystems that can adjust the encryption level of each item to the required level for the desired functionality. Although very promising, we postulate that these approaches are not yet directly applicable in real industrial strength use-cases. One reason is that businesses have to process not only structured (i.e., SQL) but equally semi-structured data in the same application context. Another reason is the inability to methodically decide (and accept) the inevitable trade-off regarding functionality, security and performance that has to be made when outsourcing data for secure processing in the cloud.

The inherently scalable nature of the cloud has resulted in the introduction of different processing paradigms; one such example is MapReduce, a widely used mechanism to process large data sets with a parallel, distributed algorithm. The same tauntness between confidentiality and processing applies here. PRISM [16] is a framework to enable the MapReduce paradigm in a privacy-preserving way; it assures both data and query privacy and takes advantage of the inherent parallelization akin to cloud computing. This work unfortunately does not allow more advanced features that, for example, the enabling of the search/lookup functionalities for multiple keywords in an efficient manner. Furthermore, current "privacy preserving word search" primitives only allow the owner of the data to query its data. Of course, delegating such operations should not enable third parties to jeopardize the confidentiality of the outsourced data. Moreover, the data owner should be able to revoke the access of such authorized parties at any time.

Among data processing primitives, word search (i.e., verifying, whether a certain word is part of a dataset) is one of the most fundamental operations. The majority of existing privacy solutions are either impractical as they are designed for centralized architectures or do not provide sufficient security guarantees.

Another important problem here is to verify the integrity of the computations over encrypted data which are carried out by the cloud: have all inputs been considered in the computation? Was the correct function applied? Have the results been tampered with? These are classic questions in systems security, and it is particularly relevant in the context of cloud computing. Various solutions

have been proposed that make assumptions about the class of computations, the failure modes of the performing computer, etc. However, deep results in theoretical computer science - interactive proofs, probabilistically checkable proofs (PCPs) coupled with cryptographic commitments, etc. -- tell us that a fully general solution exists that makes no assumptions about the third party: the local computer can check the correctness of a remotely executed computation by inspecting a proof returned by the third party. The problem is practicality: if implemented naively, the theory would be prohibitively expensive (e.g., trillions of CPU-years or more to verify simple computations).

Over the last several years, a number of projects have reduced this theory to near-practice in the context of implemented systems; we call this field proof-based verifiable computation. The pace of progress has been rapid, and there have been many exciting developments. There are still several challenges to be addressed to bring this theory to near-practicality: many of these questions cut across multiple sub-disciplines of computer science: complexity theory, cryptography, systems, parallel programming, and programming languages.

Fully-homomorphic encryption has been long sought as a means of achieving processing of encrypted data. Recently, Gentry introduced the first fully homomorphic cryptosystem [17]. Since his seminal work, several papers have addressed a number of shortcomings of the scheme, namely the required key length and the space and time complexity. Despite some very promising advancement, this technology is still not ready to be applied as a general-purpose solution to the problem of secure data processing.

2.1.5 Ensuring Data Integrity/Availability

Whereas PoW deals with the assurance that a client indeed possesses a given file, Provable Data Possession (PDP) and Proof of Retrievability (PoR) deal with the dual problem of ensuring – at the client-side – that a server still stores the files it ought to. PoR and PDP schemes address the requirement of data integrity (ensuring that data has not undergone malicious modifications) and availability (ensuring that data is still available in its entirety and can be accessed if needed).

PDP is formally introduced by Ateniese in [18]. A number of earlier works already address remote integrity verification (see [19], for more details). The protocol of Ateniese is based on asymmetric cryptography and uses the following methodology: prior to the upload, the data owner computes a tag for each data block. Any verifier can later prompt the storage service to answer a challenge on a subset of blocks: the computation of the response involves both data blocks and tags. Ateniese also present a dynamic PDP scheme [20] based on symmetric cryptography, and shows how relaxing the requirement of public verifiability allows a much more lightweight scheme. The scheme is dynamic in that it allows data blocks to be appended, modified and deleted. Erway [21] presents formal definitions of dynamic PDP together with two protocols allowing also block insertion. PoR schemes, introduced by Juels and Kaliski [22] combine message authentication code-based data verification with error-correcting codes (ECC) to allow a client to download pre-determined subsets of blocks and check whether their MAC matches the pre-computed one. The use of ECC ensures that small changes in the data are detected with high probability.

Trusted Execution Environments offer a way of securing PoR and PDP protocols. In particular, trusted computing-based systems can be used to generate proofs supporting properties on the lower layers of the software stack and the function set of the Trusted Platform Module (TPM). While feasible in theory, such approaches still suffer from the limitations highlighted in the previous sections.

Encryption keys stored in the hard disk are susceptible to tampering, TPM solutions offer a protected storage of keys through hardware and protection of authentication credentials by binding them to the platform, providing a stronger mechanism to prevent unauthorized access to the platform and thus, the integrity of the data stored. Authentication built on top of trusted computing services (based on the use of TPMs) provides higher degrees of assurance, but performance overheads introduced can be significant.

2.1.6 Data Access Policies

Securely enforcing data access policies is a challenge of paramount importance in existing clouds. In fact, current clouds: (i) do not implement any mechanism to ensure the secure deletion of their data and (ii) rely on the cloud to enforce data access decisions between different tenants. This latter

limitation becomes especially evident when the cloud is untrusted to perform such unilateral decisions.

On the other hand, an elegant solution for enforcing access policies when sharing data is to ensure that it can be recovered only upon agreement of a given threshold of users. This approach was proposed by Shamir [23]. The authors introduced a scheme where data is partitioned in n shares and each share is given to one user, so that only if a threshold of t ≤ n of users combine their shares, the original data can be reconstructed; the proposed scheme provides information-theoretic secrecy in case not enough shares (e.g., less than t) are available. That is, t shares are enough to recover the original data, while no information is leaked if any number of shares below t are available. The drawback of this scheme lies in the high computation/storage cost that it incurs, which makes it impractical to be used for sharing large files. Rabin (Rabin, 1989) proposed an information dispersal algorithm with smaller overhead than the proposal of Shamir (Shamir, 1979). However, this proposal does not provide any guarantees on the secrecy of the original data when a small number of shares (less than the reconstruction threshold) are available. Krawczyk (Krawczyk, 1993) proposed to combine both approaches ((Shamir, 1979) and (Rabin, 1989)) in order to accommodate large files and to provide guarantees on the availability of the data in case users have access to less shares than the reconstruction threshold; in (Krawczyk, 1993) a file is first encrypted using AES and then dispersed using the scheme in (Rabin, 1989), while the encryption key is shared using the scheme in (Shamir, 1979).

2.2 Innovation Potential

We have shown most existing solutions are not suitable for the market because they either provide security at the expense of the economy of scale and cost effectiveness of the cloud (e.g. data is encrypted before being outsourced, which prevents any computation to be performed in the cloud), or they meet the latter objectives at the expense of security (e.g., data deduplication and compression optimally use the resources of the cloud provider but require the customer to blindly trust its cloud provider).

The main aim of TREDISEC is to bridge this gap by developing tools and systems to address these shortcomings and to enhance the confidentiality and integrity of data outsourced to the cloud without affecting functionality, and storage efficiency.

From a practical standpoint, the ambition of this project is to develop systems and techniques that make the cloud a secure and efficient place to store data. We plan to step away from a myriad of disconnected security protocols or cryptographic algorithms, and to converge instead on a (possibly standardized) single framework where all objectives are met to the highest extent possible.

Based on our assessment, we identify the 7 key innovation points of TREDISEC shown in Figure 1.

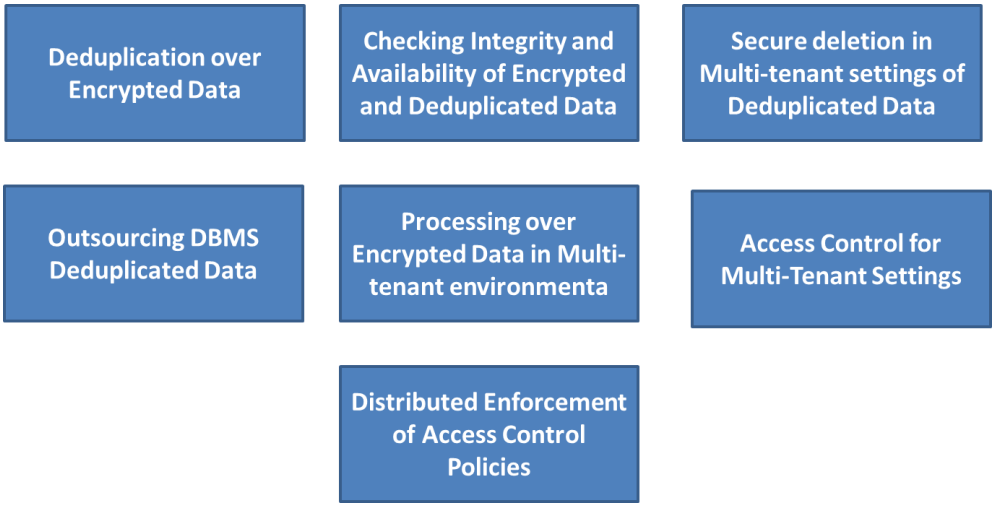

Figure 1 The 7 key innovation points of TREDISEC

We now detail those 7 innovation aspects:

1. **Deduplication on encrypted, multi-tenant data**: when we require cloud storage systems to handle data in encrypted form from multiple, mistrusting tenants, there are no known ways in which deduplication can be carried out. TREDISEC aims to leverage existing or novel cryptographic protocols and system security mechanisms which offer strong data confidentiality guarantees while permitting data deduplication across multiple tenants.
2. **Mechanisms to check the integrity and availability of multi-tenant data in presence of storage efficiency**: current PoR/PDP solutions have not been analysed in conjunction with data deduplication and require specific pre-processing of data by its legitimate owner prior to outsourcing. These techniques fail if data is shared in the cloud by multiple-tenants because either (i) the key material used cannot be shared amongst mistrusting entities or (ii) the pre-processing causes the data to be unsuitable for deduplication. TREDISEC will ensure data availability while data is deduplicated and enable the verification of data integrity and availability in multi-tenant settings. TREDISEC will also find a novel solution to store data encrypted within a multi-tenant environment while maintaining the possibility to search over the encrypted data.
3. **Secure deletion of multi-tenant data in presence of deduplication**: deduplication of data breaks the paradigm over which cryptographic secure deletion hinges, namely, that secure deletion can be obtained by breaking one of the links of the cryptographic chain of keys that decrypts the required datum; indeed, in presence of deduplication, multiple such chains exist from each of the individual, mistrusting uploaders. TREDISEC plans to develop new security protocols, possibly leveraging trusted execution environments and/or the novel cryptographic techniques developed by its partners, to enable secure deletion of one tenant's data on top of cross-tenant deduplication, i.e., ensure that a tenant's deduplicated data is securely deleted without disturbing the data access to other tenants.
4. **Storage efficiency in presence of securely outsourced DBMS data**: data outsourcing schemes used by clients to encrypt data prior to upload to the cloud do not lend themselves well to deduplication/compression, nor do they optimize on encryption performance. TREDISEC will devise novel secure data outsourcing schemes which, by design, can work atop compressed/deduplicated data. In parallel, TREDISEC will show new concepts to optimize the encryption process of large data sets which shall be migrated into the cloud.
5. **Secure outsourced analytics/processing in a multi-tenant environment**: how can multiple, mistrusting tenants outsource both data and computation into the cloud in such a way that (i) legitimate processing is possible (where the extent of what is legitimate is defined by the data owner) and yet (ii) the confidentiality of data and privacy of the users querying it are respected? TREDISEC will investigate new primitives such as delegated privacy preserving schemes, supporting revocation with no significant impact on performance.
6. **Trustworthy, consistent and conflict-free access control for multi-tenancy settings**: current attribute-based access control (ABAC) models have been proven limited in multi-tenancy cloud settings. In ABAC, authorization to perform an action is determined by evaluating attributes associated with the subject, object, and the requested action against a policy that describes the allowable set of actions for a given set of attributes. Authenticating users whose attributes are distributed across different trust domains becomes a problem due to the lack of confidence between principals and the means to verify a chain of trust. TREDISEC will provide the mapping between existing ABAC models to enforceable policies (e.g. XACML-based), enabling the definition of policies that can be verified against conflicts, and that effectively govern access to the growing number of cloud transactions when spanning different "circles of trust". TREDISEC will also develop a service able to evaluate these policies against distributed attributes, allowing immediate access to services/resources to tenants that belong to different circles of trust.
7. **Distributed enforcement of access control policies**: current cloud platforms are agnostic to the concept of shared ownership. These platforms force a collaborating group to elect one leader with unilateral access control powers, which he may very well abuse. The problem of collaboratively enforcing access control policies in the cloud is further hardened because existing cloud platforms do not allow deployments of third-party enforcement components. TREDISEC will leverage a novel set of cryptographic primitives which ensure that access to

data can be efficiently collaboratively achieved while preventing malicious tenants from combining their credentials and escalating their access rights.

3. Innovation Strategy Plan

In what follows, we detail our Innovation Strategy Plan (ISP). As mentioned earlier, the main goal of the ISP is to ensure that the innovation goals of the project do not lose relevance given changing market trends.

We achieve this by continuously monitoring the technological and scientific trends and breakthroughs in the market, and by comparing the current trends with the 7 key innovation points mentioned previously. This is achieved as follows:

- A number of deliverables of TREDISEC will contain a clearly labelled section which will outline current trends in the area (at the time of writing the deliverable) and clearly state the comparative innovation of the presented technology in relation to current state of the art. This section will be reviewed by the deliverable reviewers and the ID in order to better track the innovation aspects of TREDISEC as the project progresses. We identify the following deliverables which will contain an innovation summary section: D2.3, D2.4, D3.2, D3.3, D4.1, D4.2, D4.3, D4.4, D5.1 and D5.2.
- The ID and other Executive Board (EB) members will continuously monitor the technological trends in order to influence WPs action plans according to the evolving market trends.
- In each plenary meeting (i.e. General Assembly), a dedicated session for innovation management will be held. This session will be chaired by the ID in the presence of the WPLs. During the session, each WP leader will present the current innovation status of their WP, in order to acquire feedback from the ID and WPLs.
- During the innovation management session in plenary meetings, the ID and WPLs will re-evaluate the innovation-related risks outlined in Table 1 and assess the overall innovation level of the project or project's Innovation Health Level (IHL) (see section 4).
- Dissemination plans and Communication Strategy will be reported well in advance to the ID and EB. The ID, with the support of the members of the Press Office (which is composed by the Scientific and Communication Directors, the WP7 leader and the Project Coordinator), need to monitor every dissemination activity in order to ensure that no information that is published can harm the innovative project results.
- If, at any point in time, an innovation-related risk emerges (described in Table 1), the ID will call for an immediate Executive Board meeting in order to discuss possible measures to mitigate the alleged risk. When necessary, the ID, after consultation with the EB, will call for a re-definition of specific tasks in the corresponding WPs in order to ensure

Description of risk	**Impact**
State-of-the-art environment / project objectives lose relevance	High
Technological changes require significant redesign	High
Conflict between innovations produced by the project and existing / new patents	Low
Results produced by TREDISEC are not well exploitable	Medium

Table 1 Innovation-related risks

4. Framework for the continuous assessment of project innovation

This framework is a tool that helps TREDISEC project management in assessing the level of innovation of the project activities throughout the project duration.

A list of elements that represent the main lines of work within the project, such as research and technological advances, new developments and solutions, methodologies and conceptual models or business models, will be continuously monitored throughout the project with regards to a set of innovation indicators, grouped into three main dimensions: technological/scientific, market and organisational.

By regularly assessing the selected elements with regards to these indicators, the ID can control the project's Innovation Health Level (IHL), a RAG (Red-Amber-Green) score that reflects the potential impact that TREDISEC developments and research advances have with regards to current technologies and market trends. This information also helps the ID in putting in place the necessary actions to mitigate the deviation from the strategic innovation objective.

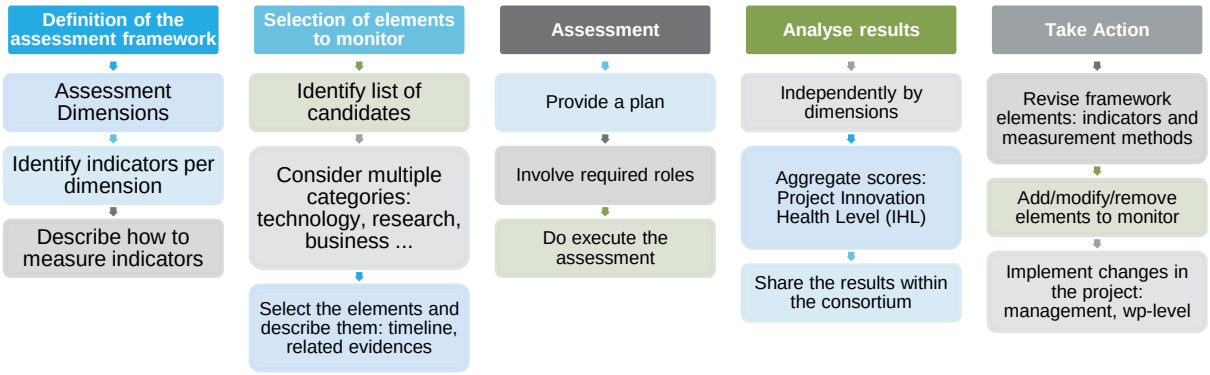

Figure 2 Innovation Assessment Process

The Innovation assessment process depicted in Figure 2 reflects the steps that will be implemented in TREDISEC, by the Innovation Management task within WP1, in order to set the grounds to continuously evaluate to what extent project activities are conducted in an innovation-driven manner.

In the following sections, we detail this process.

4.1 Framework definition

The diagram depicted in Figure 3 shows the three dimensions used to assess the innovation of TREDISEC: the scientific/technical dimension, the market dimension, and the organizational dimension. In what follows, we detail each of these dimensions.

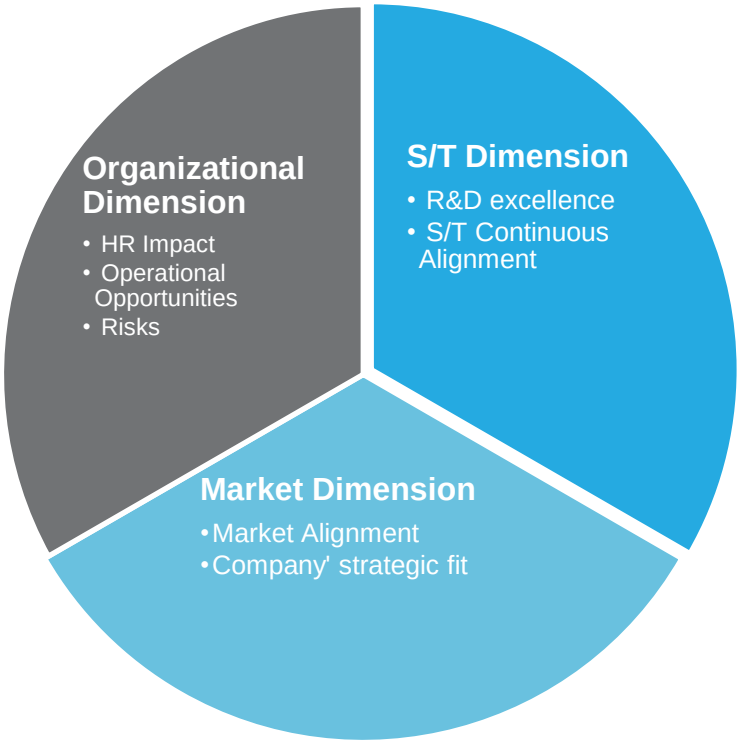

Figure 3 Framework Assessment Dimensions

4.1.1 S/T Dimension

The Scientific/Technical dimension focuses on assessing the alignment of the R&D activities to what constitute current and long term concerns from a scientific/technical point of view, and thus, in being relevant, timely and adaptable. S/T dimension also assesses the excellence of the R&D work, as a combination of high quality and relevance.

4.1.1.1 Project R&D Excellence

Project R&D excellence is an indicator that assesses the relevance of the project investigations and research developments by looking at their perception and impact within the R&D community. This assessment consists in measuring production, i.e. the quantity and frequency of dissemination activities of the consortium members; but puts a special emphasis on the quality of these publications, i.e. the specialized research capacity shown and the impact caused, for example as follow-up opportunities that may arise.

The following are tools to measure R&D excellence:

- Bibliometrics: provides reliable data (on an aggregate level) of both production (quantity) and scientific impact (quality)
  - Average of Relative Citations (ARC)
  - Number of accepted publications in conferences and journals
  - Number of accepted publications in top tier ranked conferences, journals or venues (A*, A,B) [2]

[2] http://core.edu.au/index.php/conference-rankings
http://icsd.i2r.a-star.edu.sg/staff/jianying/conference-ranking.html
http://faculty.cs.tamu.edu/guofei/sec_conf_stat.htm, http://webofscience.com

- Number of talks and presentations given to a specialized audience, co-located sessions
- Presence of the project in events of relevance and high impact
- Collaboration with other related projects and initiatives (nature and quality)
    - Peer reviewed publications
    - Publications co-authored with at least one author external to the project
    - Joint-workshops, joint-networking sessions
    - Cross-references in official dissemination mean (e.g. website, LinkedIn group, presentations, etc.)
- Contribution to designing educational programs, creation or participation in summer schools or related teaching courses, creation of masters/degrees, webinars, tutorials including fully/partially content developed in the project.

These indicators are consistent with the EU strategy for dissemination and communication of research activities in H2020 [24]. Therefore, these indicators will mainly assess the activities carried out in "Task 7.2 Dissemination" mainly, but also in "T7.5 Communication" because of its objective to reach a wide audience and create high impact and awareness besides R&D community.

Excellence implies consistently producing research that, according to one's peers makes an important contribution to the field of study in question. In addition to creating impact within the R&D community, R&D excellence can also make an impact in policy development, which benefits from high quality research achievements. Public administration's demand for research results is increasing since that may help to legitimise policy measures and influence interest groups outside of the state and other centres of power in civil society.

R&D excellence has also a societal dimension component: ensuring a high quality contribution that the dissemination of TREDISEC research makes to social development promoting a research culture.

4.1.1.2 S/T continuous Alignment

The TREDISEC proposal already incorporated an analysis (non-exhaustive) of the state of the art cloud security technologies and solutions. This analysis has been revisited in sections 0 and 2.2 in order to incorporate approaches that popped up a year after. However, a comprehensive scientific-technology watch over all aspects related to cloud security that may arise along the three years duration of the project is out of the scope of this Innovation Strategy Plan, because of the lack of resources necessary and second, because putting a watchdog in every advance of a broad field such as security in the cloud would not guarantee a success from an innovative point of view. To this end, it is necessary to identify the strategic objectives for the S/T Watch that (i) are oriented towards meeting existing research gaps in the field of security in cloud and, (ii) target priority and mandated issues that reflect both the current and the long-term scientific-technical concerns of the EU.

To measure S/T alignment of TREDISEC key innovation points, the ISP will closely monitor the following instruments that regularly provide an insight on current research and technological gaps and reflect long-term, ongoing objectives at the EU-level.

- EU Digital Agenda [25]
- NIS WG3 Strategic Research Agenda[3], secure ICT landscape
- Trust in Digital Life[4] ecosystem
- Results of the FP7 CAPITAL project evaluation

Assessing S/T alignment means also measuring the adaptability of the TREDISEC technologies, and the models and algorithms in which these technologies are built upon. Evaluating their capacity to respond to new emerging concerns, both scientific and political, and to adjust the R&D directions as new technological paradigms and models arise, contributes to increase the probabilities of technology adoption and a successful innovation. Some of the aspects to assess are the use of standards, and model-driven methodologies for the design of the primitives. The use of Agile software development techniques also promotes adaptive planning, evolutionary development, early delivery, continuous improvement, and encourages rapid and flexible response to change [26].

[3] https://resilience.enisa.europa.eu/nis-platform/shared-documents/wg3-documents
[4] http://www.trustindigitallife.eu/publications.html

4.1.2 Market Dimension

4.1.2.1 Market Alignment

This indicator assesses the alignment of TREDISEC key innovation points to current and forecasted market trends. This assessment entails a continuous monitoring of market evolution with regards to some strategic aspects:

- Evaluate how TREDISEC results help bypassing known barriers that prevent adopting cloud services
- Evaluate how TREDISEC results contribute to building incentives that foster adoption in both public and private organizations.
- Evaluate how TREDISEC-based cloud solutions could constitute a benefit that SMEs may exploit (since SMEs do not always understand all the information security risks and opportunities of cloud computing).

The ISP Market Watch is an instrument that monitors market trends focusing on the above-listed objectives, mainly with the support of the following relevant resources:

- Gartner Hype Cycle[5]
- Atos specialized publications: Ascent Journey[6], Trusted European Cloud
- NEC specialized publications: NEC Technical Journal
- EU-related sources: ENISA's Cloud Security Guide for SMEs and SME Cloud Security Tool[7], NIS WG3 report on business cases and innovation paths[8]

4.1.2.2 Company's Strategic Fit

This indicator evaluates how by ensuring alignment of TREDISEC outcomes to individual project partner's strategic market objectives/lines (e.g. business/exploitation plans), have a positive effect in maximizing the success of the Technology Transfer (TT) process and thus, the exploitation opportunities.

For instance, some risks that may endanger a successful TT process in an organization are:

- Fragmentation of ownership
- User requirements that do not adapt to company's lifecycle and needs
- Assumptions and simplifications done in the research project
- Lab versus operational environment
- Maintenance issues (e.g. know-how, costs etc.)

In order to assess company's strategic fit, the framework will measure the following elements for each of the industrial partners of the consortium:

- Intellectual property: number of patents
- Contribution to company's portfolio: contribution to service offering, product, etc.
- Number of exploitation opportunities created, prospects done
- Time-to-market (TRL-based estimation): measure rapid adoption…

4.1.3 Organizational Dimension

Technology can have an impact on how an organization is structured and how work flows. This dimension focuses on the impact that an end-to-end security model for cloud services and solutions, such as the one proposed by TREDISEC, may have in organizations, assessing both the benefits and the risks that may be derived from its adoption.

[5] http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp
[6] http://ascent.atos.net/
[7] https://www.enisa.europa.eu/activities/Resilience-and-CIIP/cloud-computing/security-for-smes/cloud-security-guide-for-smes https://www.enisa.europa.eu/activities/Resilience-and-CIIP/cloud-computing/security-for-smes/sme-guide-tool
[8] https://resilience.enisa.europa.eu/nis-platform/shared-documents/wg3-documents/business-cases-and-innovation-paths-final-version-1.1/at_download/file

4.1.3.1 Impact in Human Resources

Technology can create/eliminate positions within a company. This indicator evaluates influence in traditional organizational structures, for instance in IT Security or Operational departments, of the TREDISEC technologies and models.

In order to assess this indicator, the framework will measure the following aspects at the level of each individual consortium organization:

- Staff diversification: new roles created in the company structure, new Labs, Units, etc.
- Competences strengthening: enhancement of existing capabilities or new knowledge creation
- Corporate culture shift: incorporation of end-to-end security cloud solutions to corporate tools and procedures, commitment of top management roles in related events, influence in corporate strategy and governance model (from static perimeter security to decentralization of security policy and security governance)

4.1.3.2 Operational Opportunities

This indicator looks at risks and benefits with regards to optimization of operational aspects of businesses. The following are a set of aspects to assess:

- Emergency response: increased availability of staff derived from a better coordination through the use of cross-disciplinary teams on the virtual space.
- Collaboration and communication: Increased flexibility and mobility (working remotely), capacity to work with virtual teams.
- Data management: enhancement of data availability by deploying secure cloud-based storage and computation technologies.

4.1.3.3 Risks

There are multiple risks derived from the adoption of TREDISEC technologies and models that may negatively impact the innovation assessment.

ENISA groups risks that affect the Consumerization of IT into three different areas of influence[9], which we will adopt in the ISP:

- Risks related to the protection of corporate data: a weak implementation of enforcement of security policies, security protocols to protect data on the move, vulnerabilities related to the maturity of the solutions, etc. They pose a risk of loss of corporate data, for instance, as a result of unauthorized sharing of information on used cloud services.
- Risks related to legal and regulatory issues: it will be difficult for businesses to enforce their entire suite of policies and working regimes, such as HR policies, legal scope and context and claims of ownership on intellectual property. Additional difficulties might arise in maintaining compliance with data protection regulation through loss of individual privacy, enterprise data and data integrity.
- Risks related to costs: uncontrolled use of Cloud Computing services, social media, drop boxes, browser data and software and applications installed or used in mobile devices. Through improper use of such services, users may neglect existing security policies and transfer company information outside the security domain, thus enabling access to non-authorized individuals.

4.2 Selection of the elements to monitor

In what follows, we identify candidate research lines, technological advances, business models and any other results/outputs of the project to assess in terms of innovation, throughout the project duration.

The candidate project elements to monitor are listed in Table 2. Each element is mapped to the related project-level 7 key innovation points depicted in Figure 1.

[9] https://www.enisa.europa.eu/activities/risk-management/evolving-threat-environment/enisa-thematic-landscapes/consumerization-of-it-top-risks-and-opportunities

Candidate Element	Related Key Innovation point	Category
D2.3: TREDISEC architecture and initial framework design	1-7	Research – models Research – algorithms
D2.4: Final architecture and design of the TARDIS framework	1-7	Research – algorithms Research – models
D3.2: Specification and Preliminary Design of Verifiability mechanisms	2	Research – algorithms Research – models
D3.3 Complete Design and Evaluation of Verifiability mechanisms	2	Research – algorithms Research – models
D4.1: A Proposal for Access Control Models for Multi-tenancy	1,3,5,6	Research – models, concepts
D4.2: A Proposal for Resource Isolation in Multi-Tenant Storage Systems	3,5	Research – algorithms Research – models
D4.3: A Proposal for Data Confidentiality and Deduplication	1	Research – algorithms
D4.4: A Proposal on Secure Enforcement of Policies in the Cloud	3,6	Research – algorithms
D5.1: Design of Provisioning Framework	2,4	Research – models, concepts
D5.2: Optimization of outsourcing activities and initial design of privacy preserving data processing primitives	4,5	Research – algorithms

Table 2 Mapping between candidate deliverables and key innovation points

4.3 Continuous Assessment

The selected elements will be assessed using the framework indicators using a RAG score system. The aggregated score of each indicator will result in a RAG score per dimension, constituting the IHL of the project.

Figure 4 shows the plan for implementing the process described in section 4. As a follow-up to this document, the first phase of definition of the framework elements will entail a complete description of indicators, getting as input the updated S/T Watch and Market Watch. The first assessment of the indicators will take place at M12, right after the first year of the project where R&D advances and preliminary outputs (e.g. models, algorithms, etc.) will be available

Figure 4 Assessment Plan for the entire project duration

The conclusions of the first assessment in M9 will result in a revision of the framework indicators and updates on the S/T and Market Watch. The revised framework will be used to conduct a second assessment at M20, taking into account advances achieved in the project with regards to previous assessment and incorporation of new developments. The revised framework, as well as the results of the assessment will be fully detailed in the first innovation management report (D1.6). A similar process will be followed with the analysis of results of the 2nd assessment, a revised version of the framework indicators, will be used to conduct a 3rd assessment at M30. This 3rd assessment will focus specially on the implementation of the TREDISEC primitives and the unified framework in order to prove the resilience of TREDISEC approach and in particular of the ISP. A final assessment will be conducted at the end of the project as part of the Evaluation task in WP6. The outcomes of the 3rd and final assessments will be included in D1.7.

5. Outlook

In M20, we plan the delivery of an intermediate innovation management report. This report will review the initial ISP and incorporates modifications if needed. Additionally, the document reports on the alignment of the project architecture designs and developments to the current technological and market trends.

A final innovation management report is additionally planned in M36. This document will report on the alignment of the project achievements to the current technological and market trends. The document also includes an assessment of the maturity of the project results after the deployment and evaluation in the use case scenarios.

6. References

[1] European Comission, “How to convert research into commercial success stories? Analysis of EU-funded research projects in the field of industrial technologies,” 2013. [Online]. Available: http://ec.europa.eu/research/industrial_technologies/pdf/how-to-convert-research-into-commercial-story-part2_en.pdf. [Accessed June 2015].

[2] Microsoft, “Microsoft cluster shared volumes,” 2013.

[3] M. Johnson, P. Ishwar, V. Prabhakaran, D. Schonberg and K. Ramchandran, “On compressing encrypted data,” *IEEE Transactions on Signal Processing,* 2004.

[4] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon and M. Theimer, “Reclaiming Space from Duplicate Files in a Serverless Distributed File System,” ICDCS, 2002.

[5] M. Bellare, S. Keelveedhi and T. Ristenpart, “Message-locked Encryption and Secure Deduplication,” in *EUROCRYPT*, 2013.

[6] M. Bellare, S. Keelveedhi and T. Ristenpart, “DupLESS: Server-Aided Encryption for Deduplicated Storage,” IACR Cryptology ePrint Archive, 2013.

[7] D. Harnik, B. Pinkas and A. Shulman-Peleg, “Side channels in cloud services: Deduplication in cloud storage,” in *IEEE Security & Privacy*, 2010.

[8] S. Halevi, D. Harnik, B. Pinkas and A. Shulman-Peleg, “Proofs of ownership in remote storage systems,” in *Computer and Communications Security (CCS)*, 2011.

[9] NIST, Computer Security Division, “Attribute-based access control”.

[10] X. Jin, R. Krishnan and R. S. Sandhu, “A unified attribute-based access control model covering DAC, MAC and BAC,” in *DBSec*, 2012.

[11] R. A. Popa, C. S. Redfield, N. Zeldovich, H. Balakrishnan and M. Catherine, “CryptDB: Protecting Confidentiality with Encrypted Query Processing,” in *Symposium on Operating Systems Principles (SOSP)*, 2011.

[12] F. Kerschbaum, P. Grofig, I. Hang, M. Härterich, M. Kohler, A. Schaad, A. Schröpfer and W. Tighzert, “Adjustably encrypted in-memory column-store,” in *ACM Conference on Computer and Communications Security*, 2013.

[13] F. Kerschbaum, M. Härterich, P. Grofig, M. Kohler, A. Schaad, A. Schröpfer and W. Tighzert, “Optimal Re-encryption Strategy for Joins in Encrypted Databases,” in *DBSec*, 2013.

[14] F. Kerschbaum, M. Härterich, M. Kohler, I. Hang, A. Schaad, A. Schröpfer and W. Tighzert, “An Encrypted In-Memory Column-Store: The Onion Selection Problem,” in *ICISS*, 2013.

[15] A. Schaad and F. Kerschbaum, “Experiences and observations on the industrial implementation of a system to search over outsourced encrypted data,” in *GI Sicherheit*, 2014.

[16] E. Blass, R. Di Piedro, R. Molva and M. Onen, “PRISM: Privacy-Preserving Search in MapReduce,” in *Privacy Enhancing Technologies Symposium* , 2012.

[17] C. Gentry, “Fully homomorphic encryption using ideal lattices,” in *STOC*, 2009.

[18] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson and D. Song, “Provable data possession at untrusted stores,” in *ACM Conference on Computer and Communications Security*, 2007.

[19] G. Ateniese, R. Burns, R. Curtmola, J. Herring, O. Khan, L. Kissner, Z. Peterson and D. Song, “Remote data checking using provable data possession,” 2011.

[20] G. Ateniese, R. Di Pietro, L. Mancini and G. Tsudik, “Scalable and efficient provable data posession,” in *SecureComm*, 2008.

[21] A. K. C. P. R. T. C. Erway, “Dynamic Provacble Data Possession,” in *16th ACM Conference on Computer and Communications Security (CCS)*, New York, USA, 2009.

[22] B. K. A. Juels, “PORs: Proofs of Retrievability for Large Files,” in *14th ACM Conference on Computer and Communications Secruity (CCS)*, New York, USA, 2007.

[23] A. Shamir, “How to Share a Secret,” *Communications of the ACM,* pp. 612-613, 1979.

[24] European Conmission, “Communicating EU research and innovation,” 2014. [Online]. Available: http://ec.europa.eu/research/participants/data/ref/h2020/other/gm/h2020-guide-comm_en.pdf. [Accessed June 2015].

[25] European Commission, “Net-Cloud Future,” 2013. [Online]. Available: https://ec.europa.eu/digital-agenda/sites/digital-agenda/files/NET-CLOUD_DIGITAL-AGENDA_clickable_0.pdf. [Accessed June 2015].

[26] Agile Alliance, “"What is Agile Software Development?",” 2013. [Online]. Available: http://www.agilealliance.org/the-alliance/what-is-agile/. [Accessed April 2015].