JASEHN
Joint Action to Support the eHealth Network

Co-funded by
the Health Programme
of the European Union

REPORT

on

the use of cloud computing in health

Document Information:

Document status:	For information to the members of the eHealth Network at their 8th meeting on 23 November 2015
Document Version:	V2.0
Document Number:	D7.2.1
Document produced by:	Joint Action to support the eHealth Network • WP7: Exchange of Knowledge • Task 7.2: Secondary use of Health Data
Author(s):	Jeremy Thorp, HSCIC (UK) Greg Fletcher, HSCIC (UK) Jurgen Wehnert, GEMATIK (Germany) Christof Gessner,GEMATIK (Germany) Luc Nicolas, FPS (Belgium)
Member State Contributor(s):	Belgium, Finland, France, Germany, Greece, Hungary, Sweden, Italy, Luxembourg, Norway, Portugal, UK
Stakeholder Contributor(s):	EHR4CR

TABLE OF CONTENTS

1. Executive summary

Cloud computing offers many opportunities for providing high-quality, high-bandwidth, resilient access to computing resource that can be rapidly provided or released as required. There are different models of cloud computing, in terms of the levels of service and the implementation model (e.g. public, private or hybrid). The selection of an appropriate model is dependent on the business requirement, in particular the levels of performance, security and resilience.

This report, the first deliverable of key task 7.2 of the Joint Action to support the eHealth Network, considers the pros and cons of cloud computing for eHealth, with a focus on the use of cloud facilities to support uses of data other than for the direct care of an individual patient.

For healthcare organizations, there are compelling reasons for exploring cloud solutions. The speed of technological change, the need for continuous re-investment, the huge growth in storage requirements as resource-intensive applications such as 3D-imaging and genomics and the pressures on recruitment and retention of skilled staff, are all good reasons for seeking a Cloud Service Provider (CSP) who can ensure reliable, scalable services. However, medical data are sensitive, and HealthCare Organizations (HCOs) have legal and ethical obligations on security and data protection. Also, many HCOs operate within budgetary constraints, and solutions must be affordable. Any approach to the use of cloud needs to comply and be seen to comply with these requirements and hence gain the confidence of the public.

Whilst there has been caution in relation to the use of cloud in health, some recent developments (e.g. the US National Institute of Health policy statement[1], the EHR4CR project in Europe) are pointing to conditions within which the risks and disadvantages of cloud can be mitigated and the opportunities and benefits are realized. A series of proposals, for use in developing contractual agreements, sets out appropriate steps from a business, legal and technical perspective.

The follow-on deliverable will develop a code of conduct on how to handle secondary use of health data. It is intended to work more closely with other relevant teams and projects (e.g. those involved in reference networks and disease registers) and with other relevant bodies such as the OECD, who have recently released a report on "Health Data Governance: Privacy, Monitoring and Research".

[1] http://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-086.html

2. Introduction

2.1 Purpose

This report forms part of Task 7.2 of the Joint Action on eHealth. The Description of Work introduces the work by saying "*This task will address the following subjects:*

- *The pros and cons of the use of cloud computing in health,*
- *Publication of a code of conduct on how to handle secondary use of health data*
- *Recommendation on de-identification of data for secondary use*".

The deliverables from this task are as follows:

- D7.2.1 Report on the use of cloud computing in health (month of delivery: M7, November 2015)
- D7.2.2 Code of conduct on how to handle health data for purposes other than patient care (month of delivery: M13, May 2016)

This report is D7.2.1. Whilst it identifies the pros and cons of cloud computing, the report aims to exceed that brief by providing guidance on appropriate actions to take.

2.2. Scope

Cloud technology in healthcare can be applied in a number of ways: (a) using cloud technology for connecting mobile devices, (b) storing patient-related data during treatment for patient care (c) using healthcare data for clinical research and public health. The task 7.2 is focussed primarily on c) although parts of this document have wider applicability.

Chapter 4 forms the body of the report in that it

- introduces the concepts of cloud computing, its characteristics and operating models.
- provides the policy background, and then
- applies the aspects of cloud to the field of healthcare, introducing a number of relevant national and international projects in the domain of secondary use for clinical research, supported by further detail in the Annexes.

The Chapter offers practical guidance based on business, technical and contractual points of view.

Chapter 5 provides initial conclusions and introduces next steps, including the actions leading to the production of a code of practice in deliverable D7.2.2. The development of the code of practice and the accompanying recommendations on de-identification of data will involve engagement with a number of projects and representative stakeholder groups.

3. Report

3.1. What is the Cloud?

3.1.1. Cloud's Five Essential Characteristics

The National Institute for Standards and Technology (NIST) in the United States has defined cloud computing as follows:

"Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." [DN check ISO/IEC 17788:2014]

Any true implementation of cloud computing must have the five essential characteristics. If any technology purporting to be cloud does not have these, then it is NOT a cloud-based technology.

On-demand self-service

A cloud customer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.

Broad network access

Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops and workstations).

Resource pooling

The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state or datacenter). Examples of resources include storage, processing, memory and network bandwidth.

Rapid elasticity

Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time.

Measured service

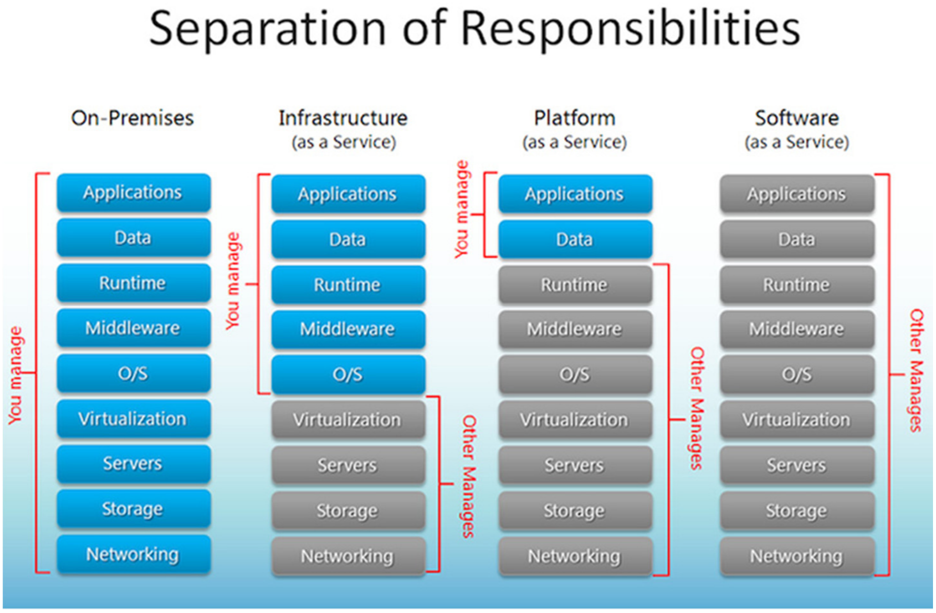
Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth and active user accounts). Resource usage can be monitored, controlled and reported, providing transparency for the provider and consumer of the service.

3.1.2. Cloud Infrastructure

A cloud infrastructure is the collection of hardware and software that enables the five essential characteristics of cloud computing. The cloud infrastructure can be viewed as containing a physical layer and an abstraction layer. The physical layer consists of the hardware resources that are necessary to support the cloud services being provided, and typically includes server, storage and network components. The abstraction layer consists of the software deployed across the physical layer, which manifests the essential cloud characteristics. Conceptually the abstraction layer sits above the physical layer.

3.1.3. Cloud Service Models

The diagram below illustrates cloud service models.



Software as a Service (SaaS)

The capability provided to the cloud client is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g. web-based email), or a program interface. The cloud customer (sometimes termed cloud consumer) does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.

Platform as a Service (PaaS)

The capability provided to the cloud client is to deploy onto the cloud infrastructure client-created or acquired applications created using programming languages, libraries, services and tools supported by the provider.

The cloud client does not manage or control the underlying cloud infrastructure including network, servers, operating systems or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment.

Infrastructure as a Service (IaaS)

The capability provided to the cloud client is to provision processing, storage, networks and other fundamental computing resources where the cloud client is able to deploy and run arbitrary software, which can include operating systems and applications. Depending on the cloud deployment model, the cloud client may not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components (e.g., host firewalls).

3.1.4. Cloud Implementation Models

There are four recognized cloud implementation models. Each model comes with its own characteristics that must be considered when choosing to build or purchase cloud capabilities.

Private Cloud

The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed and operated by the organization, a third party, or some combination of them (i.e. owned by consumer, run by third-party), and it may exist on or off premises. The emphasis here is that one organization governs and controls the use of the cloud for its own purposes. Accountability implementation, operation and use of the cloud are managed solely by that organization. An example of a private cloud might be the conversion of all computing infrastructure (services, network and CPU) for a hospital to IaaS, which is then closed to external users. [Note: the distinction between laaS and the simpler model of virtualization, which has the benefits of consolidating resources, but without elasticity, self-service or charge back (metering / pay as you use)].

Community Cloud

The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g. mission, security requirements, and policy and compliance considerations). It may be owned, managed and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises. This implementation model is scoped by the user/customer base of the community of organizations served. The services provided are typically focused on the shared "business" of that community.

Public Cloud

The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed and operated by a business, academic or government organization, or some combination of them. It exists on the premises of the cloud provider. This implementation model is well suited for generalized capabilities that are not specific to any one sector, community of users or organization. An example of a public cloud implementation might be provisioning of email or office applications.

Hybrid Cloud

The hybrid cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

The hybrid implementation model is most often found in environments where the services provided cover a broad spectrum of the characteristics of privacy, governance and deployability. In a hybrid model, an organization may have some aspects of its technology deployed in a private model, other aspects of its business needs met by participating in one or more community models, and the most generic of its needs met by use of public cloud services (e.g. where an organization has implemented a private cloud for its mission-critical applications, a community cloud for collaboration with business partners, and it has chosen to consume generic office services from a public cloud).

3.1.5. Cloud Security Principles

The Cloud Security Principles are summarized in the table below (ref. UK.GOV).

Security Principle	Description	Why this is important
1. Data in transit protection	Consumer data transiting networks should be adequately protected against tampering and eavesdropping via a combination of network protection and encryption	If this principle is not implemented, then the integrity or confidentiality of the data may be compromised whilst in transit.
2. Asset protection and resilience	Consumer data, and the assets storing or processing it, should be protected against physical tampering, loss, damage or seizure (or duplicated in different venues)	If this principle is not implemented, inappropriately protected consumer data could be compromised which may result in legal and regulatory sanction, or reputational damage
3. Separation between consumers	Separation should exist between different consumers of the service to prevent one malicious or compromised consumer from affecting the service or data of another	If this principle is not implemented, service providers cannot prevent a consumer of the service affecting the confidentiality or integrity of another consumer's data or service.
4. Governance	The service provider should have a security framework governance	If this principle is not implemented, any procedural, personnel, physical and technical

	framework that coordinates and directs their overall approach to the management of the service and information within it	controls in place will not remain effective when responding to changes in the service and to threat and technology developments.
5. Operational security	The service provider should have processes and procedures in place to ensure the operational security of the service.	If this principle is not implemented, the service can't be operated and managed securely in order to impede, detect or prevent attacks against it.
6. Personnel security	Service provider staff should be subject to personnel security screening and security education for their role.	If this principle is not implemented, the likelihood of accidental or malicious compromise of consumer data by service provider personnel is increased.
7. Secure development	Services should be designed and developed to identify and mitigate threats to their security.	If this principle is not implemented, services may be vulnerable to security issues which could compromise consumer data, cause loss of service or enable other malicious activity.
8. Supply chain security	The service provider should ensure that each element of its supply chain satisfactorily supports all of the security principles that the service claims to implement.	If this principle is not implemented, it is possible that supply chain compromise can undermine the security of the service and affect the implementation of other security principles.
9. Secure consumer management	Consumers should be provided with the tools required to help them securely manage their service.	If this principle is not implemented, unauthorised people may be able to access and alter consumers' resources, applications and data.
10. Identity and authentication	Access to all service interfaces (for consumers and providers) should be constrained to authenticated and authorised individuals.	If this principle is not implemented, unauthorised changes to a consumer's service, theft or modification of data, or denial of service may occur.
11. External interface protection	All external or less trusted interfaces of the service should be identified and have appropriate protections to defend against attacks through them.	If this principle is not implemented, interfaces could be subverted by attackers in order to gain access to the service or data within it.
12. Secure service administration	The methods used by the service provider's administrators to manage the operational service should be designed to mitigate any risk of exploitation that could undermine the security of the service.	If this principle is not implemented, an attacker may have the means to bypass security controls and steal or manipulate large volumes of data.
13. Audit information provision to consumers	Consumers should be provided with the audit records they need to monitor access to their service and the data held within it.	If this principle is not implemented, consumers will not be able to detect and respond to inappropriate or malicious use of their service or data within reasonable timescales.
14. Secure use of the service by the consumer	Consumers have certain responsibilities when using a cloud service in order for this use to remain secure, and for their data to be adequately protected.	If this principle is not implemented, the security of cloud services and the data held within them can be undermined by poor use of the service by consumers.

3.1.6. Standards

The following standards are required to underpin security of cloud facilities

Standard Guidance on certification: it is possible to be certified as compliant with ISO/IEC 27001:2005 or ISO/IEC 27001:2013, but note that ISO/IEC 27001 certification will not verify that the controls implemented by the service provider are effective.

Cloud Security Alliance CSA CCM v3.0 compliance is achieved through CSA's STAR scheme, the first level of which is 'self-assessment', followed by 'certification' and 'attestation'.

ISAE 3402, The International Standard on Assurance Engagements 'Assurance Reports on Controls at a Service Organization', and SSAE 16, Statement on Standards for Attestation Engagements No. 16, replace the US Statement on Auditing Standards No 70 (SAS 70). SSAE and ISE both require a description of the service organization's 'system' and a written assertion by management.

ISO/IEC 30111:2013 gives guidelines for how to process and resolve potential vulnerability information in a product or online service. It is applicable to vendors involved in handling vulnerabilities for IT systems.

BS7858:2012 is the British Standard that specifies a Code of Practice for security screening of individuals and third party individuals to be employed in security environments by an organization prior to their employment.

ISO/IEC 27034 provides guidance to assist organizations in integrating security into the process used for managing their applications. It introduces definitions, concepts, principles and processes involved in application security.

ISO/PAS 28000:2007 specifies the requirements for a security management system, including those aspects critical to the security assurance of the supply chain. These aspects include finance, manufacturing, information management and the facilities for packing, storing, and transferring goods between modes of transport and locations.

3.2. EU policy and strategies

3.2.1. Overview

The European Commission has long been interested in the cloud computing industry, conducting public consultations as early as 2011 and as participants in earlier debates regarding the costs and benefits of cloud technologies. In 2012 the European Commission announced its commitment to embracing cloud computing through a comprehensive strategy that set out a framework for conducting research and exploring policy options to facilitate faster adoption of cloud computing in Europe.

"The strategy seeks to establish a common set of rules to develop a cohesive market structure among the EU member states for cloud service providers. Although the European Commission's strategy does not immediately foresee the creation of a 'European Super Cloud' – a dedicated cloud system for use across Europe in the public sector – one aim of the strategy is to ready the cloud market for public sector use. More specifically, the strategy states the EU policies will focus on 'enabling and facilitating faster adoption of cloud computing throughout all sectors of the economy which can cut ICT costs, and when combined with new digital business practices, can boost productivity, growth and jobs'. As part of the effort, the European Commission plans to address several key areas related to harmonizing laws across borders,

consumer protection, contracts and transactional fairness, and standards development".

Decision 2010/87/EC was adopted and focused *"on the standard contractual clauses for the transfer of personal data to processors established in third countries". It* states that: *"This Decision should contain specific standard contractual clauses on the sub-processing by a data processor established in a third country (the data importer) of his processing services to other processors (sub-processors) established in third countries". "In addition, this Decision should set out the conditions that the sub-processing should fulfil to ensure that the personal data being transferred continue to be protected notwithstanding the subsequent transfer to a sub-processor".* The sub-processing should only consist of the operations agreed in the contract between the data exporter and the data importer incorporating the standard contractual clauses provided for in this Decision and should not refer to different processing operations or purposes so that the purpose limitation principle set out by Directive 95/46/EC is respected (compare with "Clause 11" of Decision 2010/87/EU).

However, as cloud technology grows faster than the activities of the legislature, this issue is not only for the European Union and each member state, but is also a global issue. It still lacks a revised legal framework (within the context of privacy, but also within civil and criminal codes) that takes into account the innovation introduced by cloud computing and is able to provide adequate protection in respect of legal categories related to the adoption of distributed computing and data storage services. For example, the European legislation on data protection dates back to 1995. Furthermore in 2013 the "Data Retention Directive" was considered incompatible by the Court of Justice with the article 7 of The Charter of Fundamental Rights.

To solve the problem of legal fragmentation, an important policy change for the entire sector of electronic communications should take place with the approval of a new general regulation on data protection. The new regulation will introduce the same rules in Europe and in relation with non-European states (hence rewriting the code of privacy) and this should help to make it less complex and risky to use cloud services. One of the important innovations of this reform will concern the extension of the notification of security breaches that relate to personal data to all holders of data processing such as, for example, banks, insurance companies, local health authorities, local public bodies. When required, the persons concerned will be informed without delay of loss or theft of their data.

3.2.2. Other Policies

The National Institute for Health in the US has recently issued a Notice for Use of Cloud Computing Services for Storage and Analysis of Controlled-Access Data Subject to the NIH Genomic Data Sharing (GDS) Policy (March 2015). This Notice (see Annex D) provides guidance on the conditions within which NIH would allow researchers to use cloud facilities for data storage and access.

It is important to ensure there are mechanisms for ensuring that cloud service supplier meet requirements. Annex E provides further detail and offers guidance in case of difficulty.

3.3. Implications for health

3.3.1. Introduction

The previous chapters introduced the concepts of cloud computing, and identified some of the opportunities and the challenges. This chapter considers the implications for health, building on experience from specific countries (e.g. Belgium) and specific projects; the chapter then offers guidance on how to proceed. A list of references is provided in Annex A and a summary of relevant projects in Annex B. In particular, the EHR4CR project has developed requirements and specifications, included in Annex C.

3.3.2. Characteristics of eHealth

For healthcare organizations, there are compelling reasons for exploring cloud solutions. The speed of technological change, the need for continuous re-investment, the huge growth in storage requirements as resource-intensive applications such as 3D-imaging and genomics and the pressures on recruitment and retention of skilled staff, are all good reasons for seeking a Cloud Service Provider (CSP) who can ensure reliable, scalable services. However, medical data are sensitive, and HealthCare Organizations (HCOs) have legal and ethical obligations on security and data protection. Also, many HCOs operate within budgetary constraints, and solutions must be affordable. Any approach to the use of cloud needs to comply and be seen to comply with these requirements and hence gain the confidence of the public. In this chapter, three sets of issues are considered: business, technical and contractual.

3.3.3. Business Perspective

From a business perspective, it is important to understand the nature of the requirement, such as the answers to the questions below:

- What are our modes of operation?
- What data do we handle?
- What are the constraints related to data types?
- What are the risks specific to certain types of data?
- What data can we, controlling risks, outsource?

It is important for the HCO to distinguish between expected benefits and risks to safety or operations that are unique to the adoption operation of a "cloud computing".

Before considering the use of cloud computing, the HCO must clearly identify the data, treatment or services that might be hosted in the cloud and determine the return on investment, taking into account:

- Where data are subject to specific regulations, identify the minimum conditions or restrictions on their processing The additional costs of the implementation of changes, should be assessed

- classify data according to their sensitivity (confidentiality), their expected availability (availability) and quality (integrity) and characteristics (e.g. authenticity, accountability, reliability and traceability)
- a risk analysis and business impact analysis to determine the "tolerances" for security incidents and hence map expected service levels
- Identify the relevant approach for each cloud that consideration be given to subcontract
- Evaluate return on investment, for different collaboration schemes chosen.

3.3.4. Technical Approach

Building on the analysis of business requirements, the HCO can develop its strategy for the deployment of services. Typical aims might be to:

a) Provide a cost effective, secure and reliable platform(s) to host systems and applications, keeping the level of hosting diversity across the organization manageable.

b) Remove the HCO from undertaking the low-level plumbing ('tin and wires') of infrastructure, instead focusing efforts on more strategic issues

c) Provide agility to rapidly respond to new programmes, initiatives and changing demands.

d) Minimize vendor lock-in and protect the HCO from failing vendors (e.g. by ensuring ability to migrate data).

Section 3.3 introduced a number of service models (IaaS, PaaS, SaaS). In addition to those, HCOs may wish to consider:

- The physical hosting arrangements for the as-a-Service offering
- *Colocation:* offers the ability to rent space in a 3rd party Data Centre for IT equipment. The vendor provides power, cooling, physical security and network connectivity.
- *Fully Managed*: a bespoke system is provided, fully managed and hosted by an outsource partner (but still required to meet security and data protection requirements.

These are not mutually exclusive, as IaaS and PaaS can be provided using either a Public Cloud or Dedicated Cloud infrastructure. A Public Cloud is shared across multiple organizations. A Dedicated Cloud is built specifically for an organization and is not shared with any other organizations. A possible scheme of responsibility for providing each component of the technology solution is summarized in the table below:

Component	Data Centre	Hardware	Hypervisor	Infrastructure Stack	Application Stack
Hosting model					
On-Premise	HCO	HCO	HCO	HCO	HCO
Co-location	Hosting Provider	HCO	HCO	HCO	HCO
IaaS	Hosting Provider	Hosting Provider	Hosting Provider	HCO	HCO
PaaS	Hosting Provider	Hosting Provider	Hosting Provider	Hosting Provider	HCO
SaaS	3rd party	3rd party	3rd party	3rd party	3rd party
Fully Managed	3rd party	3rd party	3rd party	3rd party	3rd party

The deployment environment will restrict the characteristics of the cloud model it supports. In particular the ability to elastically scale rapidly will be limited by the capacity of blades or racks available beyond which new hardware must be provisioned and installed.

The primary motivation of the hosting strategy must be to address business needs: fulfilling security requirements, service availability, then value for money and agility. A hosting approach is illustrated in the matrix below as a product of these security (data confidentiality) and service (system criticality) requirements:

		System criticality >>>			
	Business Impact Level (BIL)	Bronze	Silver	Gold	Platinum
Security level >>>	BIL 4+	3rd party dedicated cloud	3rd party dedicated cloud	3rd party dedicated cloud	3rd party dedicated cloud
	BIL 3	Internal with co-location	3rd party dedicated cloud	3rd party dedicated cloud	3rd party dedicated cloud
	BIL 2	Internal with co-location	3rd party public (or dedicated) cloud	3rd party public (or dedicated) cloud	3rd party dedicated cloud
	BIL 1	Internal with co-location	3rd party public (or dedicated) cloud	3rd party public (or dedicated) cloud	3rd party dedicated cloud
	BIL 0	Internal with co-location	3rd party public (or dedicated) cloud	3rd party public (or dedicated) cloud	3rd party dedicated cloud

The HCO should assess the importance of the application, for instance using the Business Impact Level (BIL) (https://www.gov.uk/service-manual/making-software/information-security.html). Business Impact Levels (BIL), often shortened to Impact Levels (IL) is a classification system used to guide discussions of risk in government projects.

Similarly, there are formal data classifications such as OFFICIAL, SENSITIVE, PERSONAL (https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/251480/Government-Security-Classifications-April-2014.pdf).

Within the hosting model, the main focus is on confidentiality – the potential impact if the information is seen by those who should not see it.

The HCO should then assess the Service Category of the system – e.g. Bronze, Silver, Gold or Platinum. Each Service Category lists typical, not absolute characteristics such as availability, scale, numbers of users, incident resolution time, impact of failure, etc. so a best fit approach should be used.

The Security / Service intersection then details which hosting solution to use:-

i. ***3rd Party Dedicated Cloud*** is a Cloud infrastructure ring-fenced for HCO use only. The hosting provider is responsible for providing managed (virtual) servers and associated networking in a secure and resilient data centre.

ii. ***3rd Party Public Cloud*** is a multi-tenanted. i.e. the Cloud infrastructure shared with other organizations. The hosting provider is responsible for providing managed (virtual) servers and associated networking in a secure and resilient data centre.

iii. ***Co-Location*** where space is rented at Cloud Service Provider (CSP) and IT assets are provided and managed by HCO. The CSP is responsible for providing accommodation, power, cooling and physical security for HCO owned and managed IT assets.

Example:

The hosting strategy comprises of a hosting policy, associated governance and implementation approach. In essence it stipulates:-

- Adopting 3rd Party Cloud hosting model for all core business - IaaS, PaaS or SaaS
- Addressing the business need of security, service and agility/VfM in that order
- Virtualising by default
- Partnering with at least two Cloud providers
- Providing commodity based pricing to programmes
- Establishing a Design Authority to govern policy adherence and grant exceptions
- Maintaining a hosting 'heat map' to indicate how systems comply with the strategy

All systems should be hosted with an HCO specified hosting partner. Each of these vendors will have already been vetted, their data centre and operational procedures checked, SLAs defined, pricing confirmed and a contract in place. The pricing will include volume discounts with a number of thresholds which activate greater discounts to allow us to leverage economies of scale. Hosting demand will be balanced out between multiple vendors.

All systems should use HCO certified Cloud Building Blocks to construct all hosting solutions. These building blocks will have already been hosted by vendors and certified as being suitable for use by the HCO and are designed to accelerate the design and deployment of the solution.

Whilst security, followed by service, is the fundamental driver in the creation of the strategy, cost / value for money has also been considered.

An organization that plans to use a CSP should ensure that this provider has implemented appropriate security measures to protect against risks related to the Cloud as well as those relating to traditional data processing in especially regarding the protection of privacy.

There are many risks arising from not having suitable arrangements in place. These include the following (non-exhaustive) list; for each there needs to be an assessment of probability and impact, together with consideration of mitigating actions:

- Dependence on CSP's services (including difficulty of changing the solution or change to another provider) without data loss or unavailability
- Data hosted on a virtualized or shared system be altered or accessed by third parties
- Court orders, including foreign authorities, requiring access to data
- Fault with subcontractors, if the CSP uses third parties to provide the service
- Loss of governance over provision of services
- Loss of the ability to audit the provider due to its subcontractors management failure
- ineffective or insecure destruction of data, or too long-term conservation
- Access rights management issues
- Unavailability of the services, including network issues
- Lack of control in the event the provider is acquired by a third party;
- Regulatory Non-compliance, including international transfers.

3.3.5. Contractual

It is necessary for the institution to define its own requirements for technical and legal security. While the aim of the Cloud is to reduce certain operational tasks, it must ensure that the provider is able to offer a level of requirements at least equivalent to that requested in the HCO. For data and "trade" treatments, the HCO must particularly ensure the reversibility and a sufficient level of availability

There are two objectives, that each HCO should be able:

- to identify the data processing and quality requirements information in order to decide what information and what services can be outsourced.

- To specify the expected quality of service that will be subcontracted to ensure the CSP offers sufficient guarantees of data protection (in terms of confidentiality, integrity and availability), respect the law of privacy, data continuity but also with regard to the legal requirements and technical realization of benefits.

Certain legal and regulatory requirements applicable to the hospital sector should be taken into account if your hospital plans to save health data, and more specifically, patient records or medical records in a "cloud". Specific attention may be needed regarding data to be used for research.

The HCO which plans to process sensitive data in a "cloud" managed by a CSP must ensure the following contractual guarantees:

- **Subcontracting**
    - The CSP remains solely responsible vis-à-vis the HCO of execution therefore its obligations as if it subcontracts certain activities.
    - If specific tasks can be assigned to subcontractors, the contract must stipulate that the CSP first obtain the explicit consent of the HCO and agrees to formally refer all obligations in commitments he will contract with subcontractors.
- **Integrity, continuity and quality of service**
    - The CSP must have appropriate backup and recovery operations in place, tested and ready if needed.
    - A commitment to a service level (SLA) should be formalized in an agreement, including during and after any period of warranty, service availability and maximum restart time when interrupted after accidents and other criteria the business recovery (Recovery Time Objective and Recovery Point Objective).
    - Similarly, details of the measures to the continuity of service to be provided in Service Level Agreement (SLA) attached to the contract and financial penalties corresponding to the minimum financial loss in terms of hospital activity.
    - The CSP must commit to transparency in governance, and should, among other matters, report security incident information. The CSP should, in case of outsourcing of medical data, ensure continuity of service even in case of bankruptcy of the provider or one of his subcontractors.
- **Data storage**
    - The CSP agrees not to retain the HCO data beyond the time period determined in consultation with the HCO in relation to the purposes for which data were collected.

  - In case of early termination or end of the service, the provider agrees to restitution all of the HCO's data in an agreed manner and period, based on a format defined by mutual agreement, structured and commonly used so that the HCO can ensure continuity of service. Once restitution is done and with the HCO's agreement, the CSP undertakes to securely destroy all copies of the data in its possession, including backups and archives within a reasonable period and provide proof of destruction.

- **Data portability**
  - At the end of service, the CSP undertakes to provide the necessary assistance to the migration to a new provider

- **Audit**
  - The CSP agrees to allow audits sponsored by the HCO to collaborate closely and treat deficiencies observed as soon as possible. These audits can be performed by the HCO itself or by a trusted third party selected by thereof
  - Audits must allow a detailed analysis of compliance with contractual clauses and security rules.
  - Audits should also help to ensure that security measures on the confidentiality, availability, traceability and data integrity implemented only can be circumvented without this being detected and notified.
  - Where a claimant wishes perform audits on its infrastructure, it must first, according to a preset time, seek the agreement of the HCO so as not to impede availability of the information system.

- **Confidentiality of data**
  - The CSP must commit that neither they, any subcontractors and potential buyers (i.e. someone takes over the CSP), will use or disclose the data for their own purposes or that of third parties without express permission from the HCO
  - The CSP shall, within outsourcing contracts negotiated and concluded, introduce the concept of transfer of responsibility to include subcontractors of subcontractors the chain of responsibility of the HCO's data processing
  - The CSP must commit to protect and make available to the client all traces of the data management tools and applications
  - The CSP must inform the HCO of any anomalies it detects in these footsteps connection such that access attempts of non-authorized persons
  - The CSP must immediately notify the HCO of any inquiry or request investigation from a formal authority

- The CSP must provide the customer with a document describing in detail the policy computer security establishment and approved by its management and information the evolution of this policy.
- The CSP is required to comply with security requirements in force required by and for the HCO.
- The CSP establishes clear roles and responsabilités4 related to internal management information security and organizes the necessary contact channels with the HCO

In terms of data protection, there are five areas where security requirements must be implemented:

1) Sensitive data: the CSP must implement, consistently, processes for security of personnel management, inventory, qualification and traceability of data

2) the CSP must implement security management access for the physical data centres and technical devices ensuring the availability of data and their treatment. The data centres will be protected against fire and flood

3) logical access security: the CSP must have logical access controls ensuring adequate protection of sensitive data or not

4) system security: the CSP must have configured systems and protected from security breaches, especially for environments hosting the data, and minimizing the impact of breaches should they occur

5) network security: the CSP must have a secure network with isolation appropriate to third parties

The CSP should guarantee the following:

- the location of sensitive data is known and meets the HCO's requirements (computer centre, storage and servers);
- Safeguards and related IT systems backup plan are implemented, tested and can be checked by the HCO;
- it has a code of ethics that is applied by its staff and any subcontractors
- it has and will not exercise activities that may cause a risk of conflict of interest;
- staff regularly attend security awareness training, and will be monitored for compliance;
- it has centralized all elements of traceability and to guard against violations of privileges or behaviours
- it has a management of security incidents including the detection, warning, treatment up to resolution, identifying the causes and communication to the HCO.

For physical security, the CSP should guarantee the following:

- It has secure physical access control systems, intrusion detection, fire, flood and video surveillance;
- Access to the data centre shall be allowed only to authorized persons following appropriate approvals; they are traced and reviewed regularly;
- any subcontractor used for maintenance or repair of equipment containing sensitive data is subject to contractual clauses of confidentiality;
- any data storage media containing sensitive data will be destroyed on expiry.

In respect of system access, the CSP should guarantee the following:

- it applies the rules of granting access to data based on elements provided by the HCO (copy, consulting, creation, modification and deletion,
- user access to systems administrators and containing sensitive rely on mechanisms to ensure the confidentiality and traceability (audit trails on data access);
- it applies an authentication policy in line with that of the HCO.

In respect of systems, the CSP should guarantee the following:

- The saved data, regardless of the medium is encrypted.
- manages vulnerabilities in systems and organizes at least annually tests intrusion, the identified critical vulnerabilities are corrected immediately.
- servers that host sensitive data is configured with a security level strengthened; security patches are managed centrally, and previously tested applied within a time less than one month with prior agreement of the HCO (problematic applications that do not support patches);
- anti-virus software installed on servers, workstations, maintained and supervised;
- the use of USB drives and other portable storage media is controlled, managed and native prohibited on all systems containing sensitive data, including all types of external storage not explicitly provided for in the contract;
- processes and risk management practices, incident management and management re implemented and properly documented.

In respect of network access, the CSP guarantees the following:

- the entry points to the network are limited, secure, filtered and logged;
- the systems administration tasks are carried out from a network of Directors secure, dedicated and isolated by connecting with strong authentication mechanisms;

- changes in network equipment are tracked, documented and previously approved.
- In the case of a "Cloud computing" shared service:
  - Network access is allowed only to trusted devices
  - the network to which are connected systems hosting sensitive data is isolated from the third network.

4. Conclusions

It is clear that much work has already been completed in the field of cloud and secondary use of data, and many activities currently underway are addressing elements of this topic. European projects such as EHR4CR have developed highly relevant material, and the Joint Action on disease registries and rare diseases is adding an important component with the customer view of what is needed.

There are two main recommendations: that there should be dedicated solutions (private or community cloud) for hosting sensitive data and that there should be negotiation and drafting of specifications that address risk management in data security.

At the same time, a recent OECD report "Health Data Governance: Privacy, Monitoring and Research" (October 2015) has highlighted wide variations in the approaches taken in different countries, in turn emphasizing the need for clear guidance and a benchmark set of requirements for all to follow.

Task 7.2.2 will address two areas: publication of a code of conduct on how to handle secondary use of health data and recommendations on the de-identification of data for secondary use.

The development of the code of practice will need to engage with these and other groups, building on such input. A code of practice provides rules and guidance for all parties; an example might be "Protecting Patient Confidentiality: NHS Scotland Code of Practice", 2012.

The recommendations on de-identification will also be dependent on wider engagement, as this topic is of particular relevance to much current work.

5. References

Cloud Security Guidance: Standards and Definitions, Gov.uk, August 2014

PUBLIC SECTOR STUDY: CLOUD FOR EUROPE , FP7-610650, Version 1.0, 22/12/2014

Emerging Technology Series : Cloud Computing in Health , Canada Health Infoway, September 2012

Big Data and Data Protection, Information Commissioner's Office, July 2014

Data controllers and data processors: what the difference is, Information Commissioner's Office, May 2014

Big Data Analytics in Health, Canada Health Infoway, April, 2013

Patients' Rights in Cross-Border Healthcare in the European Union, EC Eurobarometer, 425, May 2015

A Business Logic System for Mining German Patient Records, Philipp Senger, Alexander Klenner, Juliane Fluck, GMDS 2013. 58. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V. (GMDS). Lübeck, 01.-05.09.2013. Düsseldorf: German Medical Science GMS Publishing House; 2013. DocAbstr.248

U. K. Schneider, Sekundärnutzung klinischer Daten - Rechtliche Rahmenbedingungen, mit einem Beitrag von A. Roßnagel und G. Hornung, TMF-Schriftenreihe Band 12, MWV, Berlin 2015

K. Pommerening | J. Drepper | K. Helbing | T. Ganslandt, Leitfaden zum Datenschutz in medizinischen Forschungsprojekten, Generische Lösungen der TMF 2.0, 1. Auflage, 240 Seiten, September 2014, € 64,95 [D], ISBN: 978-3-95466-123-7

Health Data Governance: Privacy, Monitoring and Research, OECD, October 2015

Protecting Patient Confidentiality: NHS Scotland Code of Practice, NHS Scotland, 2012

6. Annexes

6.1 Annex 1 -Projects

EHR4CR

The EHR4CR (http://www.ehr4cr.eu/) project involves a total of 10 companies from the pharmaceutical industry, 11 university hospitals, plus numerous academic groups and patient organizations. Together, the partners are establishing a technical platform that makes it easier to link electronic health records to research platforms and networks in the healthcare sector. In doing so, a great deal of importance is being attached to data protection in particular: the platform was being created in such a way that analysis of the de-identified data takes place at an early stage, in the relevant hospital. All disclosure of person-related data takes place only with explicit consent from the patient, who, as previously, is asked by his attending doctor whether he wishes to give his consent. Principal hurdles identified by project participants are to date: interoperability, legal (data protection) and ethical issues, and the integrity and trustworthiness of data. Therefore, user groups were integrated into the development process at a very early stage.

Ethical and legal aspects of cross-border secondary use of treatment data were discussed in a workshop held in Berlin on 12 January 2012 (see: http://www.tmf-ev.de/EnglishSite/News/ArticleType/ArticleView/ArticleID/1065/PageID/1033.aspx)

Cloud4health

On a national level, the project Cloud4health (http://cloud4health.de) is addressing related questions: This research project, was funded by Federal Ministry for Economic Affairs and Energy as part of the "Trusted Cloud" technology programme launched in 2010 (http://www.trusted-cloud.de).

The Cloud4health project taps into large medical raw data inventories for the data protection-friendly evaluation of various issues from the areas of Research, Development and Health Economy. The approach combines text analysis and data warehouse technologies and can be made available either privately or in the public cloud, depending on the need. Three application scenarios were investigated: the extraction and evaluation of information from anonymized patient data through the operative treatment of hip joints, the development of processes for the automated plausibility and profitability checks of medical treatments, as well as the early identification of undesired side effects of newly introduced medications with the help of automated processes.

An overview of the project results is published at http://cloud4health.de/projekt/publikationen

The current situation was summarized at a workshop held at TMF, Berlin on 29 October 2014 regarding data protection in medical research (see http://www.tmf-ev.de/EnglishSite/News/ArticleType/ArticleView/ArticleID/1648/PageID/1549.aspx and also http://www.tmf-ev.de/News/articleType/ArticleView/articleId/1628.aspx, with links to the presentations).

SAHRA

Another recent project addresses particularly the data protection legislation aspects when implementing secondary use solutions in healthcare: SAHRA – Smart Analysis - Health Research Access

In the German healthcare system, a great deal of data has been collected, but so far it hasn't been used sufficiently to research and improve patient care. The project "SAHRA – Smart Analysis – Health Research Access" aims to make it possible to combine billing, treatment, study and registry data in accordance with privacy legislation, and to make it available to health care research and authorised care CSPs. For this, a representative sample of comprehensive care data from medical practices will be combined with other health care system data sources via a highly secure, web-based analysis platform. This will then be analysable with scientific methods by using the most modern memory technology and be made available in compliance with the law. Data in the health sector is particularly sensitive. Patients therefore enjoy a high level of social data privacy. Therefore, a major focus of the SAHRA project is the strict adherence to data protection legislation, the technical, legal and organisational implementation of data privacy and the protection of business secrets.

The project partners are: GeWINO - Gesundheitswissenschaftliches Institut Nordost (lead), data experts GmbH, Hasso Plattner Institute of the University of Potsdam and TMF – Technology, Methods and Infrastructure for Networked Medical Research (contacts: Dr.-Ing. Thomas P. Zahn, AOK Nordost - Die Gesundheitskasse / GeWINO – Gesundheitswissenschaftliches Institut Nordost, Wilhelmstr. 1, 10963 Berlin, Guidelines on data protection in medical research projects).

Already in 2003, has the technology and methodology platform for networked Research (TMF, www.tmf-ev.de) a basic data protection concept for networked medical research projects presented. The data protection concept contained sample solutions for different types of medical research networks. 2006 was also voted a generic data protection concept for the establishment and operation of biobanks between the TMF and the Data Protection Officer of the Federal and State Governments. The concepts ensure the preservation of research interests at the same time adequate protection of patient data through certain procedures specifications (patient information and education, anonymisation and pseudonymisation, trustees use, technical and organizational measures).

Due to the further development of the framework an update of concepts has been made. The previous procedures recommendations for the protection of patient data have been retained.

The "Guide to Privacy in medical research projects" contains four modules: clinical module, study module, research module and Biobank module. The new guide has been coordinated with the working groups Science and Technology of Data Protection Commissioners. The Conference of Data Protection Commissioners of the Federation and the Länder decided on 27 and 28 March 2014 in Hamburg, medical research HCOs and -verbünden the new generic data protection concepts of the TMF as a basis for the specific configuration of data protection concepts recommended.

6.2 Annex 2 - EHR4CR Requirements

[Ref. **Electronic Health Records for Clinical Research (EHR4CR)** Deliverable 5.2 Data Protection Legal Analysis and Security Architecture Requirements Specification]

EHR4CR has been considering how to address the challenges to the use of personal health information which has been captured without obtaining patient consent. Explicit consent for processing of patient data enables a wide use, while obtaining this consent retro-actively is almost not feasible from an organizational and economic point of view. Deliverable 5.2: Data Protection Legal Analysis and Security Architecture Requirements Specification from this project describes how it has addressed the challenge of applying good practice in information governance.

Requirements relating to identification of patients who might be invited to take part in a clinical trial, have been categorized by priority according to the MoSCoW priority classification:

- M means that a requirement Must be met as part of the delivered service;
- S means that a requirement Should be met but can be superseded by a Must-classified requirement;
- C means that a requirement could be met if desirable, but is not considered either critical or a reason to delay delivery;
- W means Want and is to be considered a nice feature if possible, but in no way a priority or to be implemented ahead of any other, more critically categorized requirement.

ID	Requirement	Priority
Identity Management (of platform users)		
1.1	The identity of users attempting to access the EHR4CR workbench must be verified to provide assurance that authorized users can be reliably authenticated.	M
1.2	The identity of software components that form part of the platform must be verified so that each component can be successfully authenticated with the others when queries are being submitted	M
Authentication and Authorization (platform authorization at the service level)		
2.1	Users must be identified and their access to the workbench authenticated using a service that manages access control. This is needed for each component of the platform service, with appropriate tokens to maintain access	M
2.2	Authenticated users must be assigned appropriate access to data resources according to site and company agreements	M
2.3	Sites should control levels of access to authenticated users themselves and would like control over authorization mechanisms	S
2.4	Some sites have expressed the need to visually inspect queries before allowing execution. Others would like an automated formal mechanism for comparing queries at execution against a set of predefined constraints.	
Trust (establishing trust between service providers)		
3.1	Every service must have the capability to verify the user identity and user role, handling security tokens (e.g. certificates and passwords) and interchanging them within the	M

	infrastructure securely and reliably.	
3.2	Every service needs to have an identity that can be verified by the other services	M
3.3	Sites have stated that all communication paths should be secured using https: workbench<->orchestrator orchestrator<->site	M
3.4	A single orchestrator certificate should enable this, confining the responsibility of the platform; this is not about client authentication which is expected to be handled elsewhere.	S
3.5	Certificates, access credentials and identity keys should have their integrity and validity protected: this should protect against loss, misuse (accidental or deliberate), theft and technical compromise (e.g. disk corruption), and will require management and revocation / timed validity.	M
Audit (user and data subject or data-standard centric)		
4.1	A record of user access and any queries they submit through the workbench must be kept, including date, time, success or failure to authenticate and details of the user and their organization.	M
4.2	The Orchestrator should keep details of what data sources were referred to for querying to provide the counts, which will assist with provenance. It should also log all attempted authentications and service bindings, including date, time, outcome, client ID and possibly user identity, or a means to link this back to the logs kept by the workbench.	M
4.3	Audit of activity (processes) at the end point and nature of content accessed should be kept; this audit should be at both the study and query level and include detailed logging as well as methods of analysis, which should assist with provenance	M
4.4	The audit trail must be computer processable for analysis	M
4.5	The audit trail must be presented to the user in a human readable form	M
4.6	The audit trail and component logs should be protected from external editing and amendment, and stored separately from the running processes	S
Provenance		
5.1	The system should provide a means to give users confidence that the queries submitted are returning counts that are reliable and can be scrutinized and justified: for example, they should know where the counts have been derived from	M
Maintaining anonymity (Anonymization)		
6.1	The identity of patients that form part of the queried subset should not be disclosed, it should also be impossible to determine the identity of individuals based on low count scores and architected queries.	M
6.2	Sites need to prevent unauthorized identification of patients, requested by sites, refer to excel spreadsheet “Data access issues in EHR4CR”	M

Architecture

This section describes the architectural overview of the technical security implementations that have been developed within EHR4CR. These include implementations that manage authentication, authorization, trust, identity and credentials management, provenance and fuzzing.

The EHR4CR Security Architecture is defined as the set of services that provide the basic EHR4CR security infrastructure and the technical requirements to which EHR4CR compliant services must adhere. These requirements are formulated as restrictions on the security standards on which the EHR4CR Security Architecture is built (also called security profiles). At the same time, the EHR4CR Security Architecture features implementation and interoperability guidelines to facilitate secure and trustworthy interactions between EHR4CR compliant service consumers and providers. This approach is equivalent to the interoperability framework of profiles provided by the Web Services Interoperability (WS-I) initiative.

The following EHR4CR authentication services are available:

Service	Description	Supported standards
Identity Provider (IdP)	Issues security tokens to end-users based on username-password authentication	SAML v2 Browser profile1
Security Token Service (STS)	Issues security tokens to web service clients based on X.509 certificate authentication	WS-Trust v1.42 with SAML v23 tokens SAML Delegation profile4

The EHR4CR platform features a single-sign on (SSO) application. EHR4CR compliant web applications must accept security tokens issued by the authoritative EHR4CR IdP in the EHR4CR platform instance to which they are connected.

Web service clients are authenticated by the EHR4CR Security Token Service (STS). The STS is a WS- Trust v.1.4 compliant authentication service that issues SAML v2 security tokens. The supported authentication methods are username/password (WS-Security Username Token Profile5) and public key authentication (WS-Security X.509 Token Profile6). EHR4CR compliant web services must accept security tokens issued by the authoritative EHR4CR STS in the EHR4CR platform instance to which they are connected.

The EHR4CR STS supports the issuance of delegation tokens using the WS-Trust v1.4 ActAs feature. The issued tokens are SAML v2 tokens having the delegating user as subject and the delegate specified as part of conditions specified in the token's authentication statement. This is done in accordance with the SAML v2 delegation profile7. An EHR4CR compliant STS must support issuance of delegation tokens based on either a SAML token issued by the authoritative IdP or another delegation token issued by the authoritative STS.

The EHR4CR security infrastructure features a central authorization service for requesting platform- level authorization. This authorization service supports policy-based (also called attribute-based) access control according to the XACML v28 model. Relying parties (EHR4CR service providers) can request an authorization decision in accordance to the SAML 2.0 profile of XACML v2.09. Authorization decision queries are subject to a number of security requirements (expressed as WS- SecurityPolicy10 policies) including the requirement to include a security token issued by the EHR4CR STS in the request.

Before the EHR4CR authentication services can issue authentication tokens for an EHR4CR compliant web application or web service, the corresponding Service Provider (SP) must be enrolled in the EHR4CR *Circle of Trust* (CoT). Within this *Circle of Trust*, member SPs implicitly trust each other. The fabric from which this Circle of Trust is built, is provided by the Service Provider management component of the EHR4CR Identity and Access Management platform.

In order to use any EHR4CR compliant web application that requires end-user authentication, EHR4CR platform users must be enrolled in the EHR4CR Identity Management (IdM) system first. The EHR4CR IdM keeps track of platform user and their organizations and assigns them to trust domains. These trust domains form the basis of the security interoperability layer facilitating secure and trustworthy interaction between platform users and services.

In accordance with the EHR4CR requirements, end-users are uniquely identified by their e-mail address and are authenticated using password authentication. The EHR4CR IdM keeps track of the end-user credentials and enforces platform-wide policies such as minimal password strength and periodic password renewal. Platform services are authenticated using X.509 credentials.

EHR4CR-compliant web services and web applications shall maintain audit logs containing at least the following information. The EHR4CR Security Architecture currently does not dictate a particular audit log format or interface for querying audit logs matching certain criteria. This is planned for a later iteration. The audit support for the endpoint is therefore discussed under subsection 2.5.4 under Auditing Capabilities.

Provence capacity is increasingly employed in health informatics applications for facilitating auditability and traceability of data access and query transactions. In computing, provenance concerns the history of a piece of data. It aims to answer particular provenance-related questions. Typical provenance questions could be ”What resources were used for answering the query?” or ”How was a data access control decision made which allowed a user to query a particular data set?” One of the goals of the EHR4CR project is to address the patient recruitment issues commonly facing clinical trials. Finding adequate numbers of eligible patients is often very difficult and always time- consuming in a clinical trial project. The EHR4CR system aims to resolve such issues by integrating distributed hospital data resources and support automated queries. In principle, the enhanced availability of patient data will make patient discovery much easier and quicker. This implies many challenges given the complexity of the technical and non-technical arrangements for supporting the proposed functionality in a distributed clinical setting. The provenance capacity of the EHR4CR system focuses on addressing the traceability, cross-system auditability and compliance issues related to query processes and data access within the system. In particular, the provenance capacity aims to provide a means to answer a collection of questions regarding the provenance of data within the EHR4CR system from the stakeholders of the project, including clinical researchers, hospital data providers and system auditors.

Provenance capacity comprises a set of provenance-related functions to enable provenance data capture, store, query and integration. While the interface can be generic, the implementation of provenance capacity in each EHR4CR subsystem should follow a project standard specification to ensure constant cross-system interoperability. The provenance specification will be a part of the EHR4CR software specifications, which will systematically describe and standardize the detailed development aspects of provenance capacity in the EHR4CR project.

The architecture of the EHR4CR platform and, in particular, the setup at each pilot site is designed to protect the patient data. Queries to each pilot site must be in the form of eligibility criteria and the Eligibility Criteria Model (informally known as the Blue Model) constrains what may be asked and limits what is returned to sets of counts. The query endpoint at each pilot site submits low-level queries to an anonymised data warehouse so that data that directly identifies patients is absent. The process of Fuzzing is employed to obfuscate cases where low counts based on increasingly specific query criteria could be used to identify patients based on the details of those criteria and the numbers of counts that are returned.

References

1 See http://docs.oasis-open.org/security/saml/v2.0/saml-core-2.0-os.pdf

2 See http://docs.oasis-open.org/ws-sx/ws-trust/v1.4/ws-trust.html

3 See http://docs.oasis-open.org/security/saml/v2.0/saml-core-2.0-os.pdf

4 See https://www.oasis-open.org/committees/download.php/16076/sstc-saml-constrained-delega...

5 See http://docs.oasis-open.org/wss/v1.1/wss-v1.1-spec-os-UsernameTokenProfile.pdf

6 See http://docs.oasis-open.org/wss/v1.1/wss-v1.1-spec-os-x509TokenProfile.pdf

7 See https://www.oasis-open.org/committees/download.php/16076/sstc-saml-constrained-delega...

8 See http://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-core-spec-os.pdf

9 See http://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-saml-profile-spec-os.pdf

10 See http://docs.oasis-open.org/ws-sx/ws-securitypolicy/200702/ws-securitypolicy-1.2-spec-os.html

6.3 Annex 3 - NIH Genomic Data Sharing Policy

Notice for Use of Cloud Computing Services for Storage and Analysis of Controlled-Access Data Subject to the NIH Genomic Data Sharing (GDS) Policy

Notice Number: NOT-OD-15-086

Key Dates **Release Date:** March 27, 2015

Related Announcements

NOT-OD-15-027
NOT-OD-14-124

Issued by National Institutes of Health (NIH)

Purpose

The purpose of this Notice is to inform the research community that the National Institutes of Health (NIH) is now allowing investigators to request permission to transfer controlled-access genomic and associated phenotypic data obtained from NIH-designated data repositories1 that are under the auspices of the NIH Genomic Data Sharing (GDS) Policy to public or private cloud systems for data storage and analysis. This change is being made in light of the advances made in security protocols for cloud computing2 in the past several years and given the expansion in the volume and complexity of genomic data generated by the research community.

Provisions for the Use of Cloud Computing Services for Storage and Analysis of Controlled-Access Data Subject to the NIH GDS Policy

NIH expects cloud computing systems to meet the data use and security standards outlined in NIH Security Best Practices for Controlled-Access Data Subject to the NIH Genomic Data Sharing (GDS) Policy as well as the institution's own IT security requirements and policies.

Investigators who wish to use cloud computing for storage and analysis will need to indicate in their Data Access Request (DAR) that they are requesting permission to use cloud computing and identify the cloud service provider or providers that will be employed. They also will need to describe how the cloud computing service will be used to carry out their proposed research.

As with data stored in institutional systems, the institution's signing official, Program Director/Principal Investigator, IT Director, and any other personnel approved by NIH to access the data are responsible for ensuring the protection of the data. The institution, not the cloud service provider, assumes responsibility for any failure in the oversight of using cloud computing services for controlled-access data.

The NIH Security Best Practices for Controlled Access Data Subject to the NIH Genomic Data Sharing (GDS) Policy, which have been updated to include best practices for cloud computing, are available at http://www.ncbi.nlm.nih.gov/projects/gap/pdf/dbgap_2b_security_procedures.pdf. The Model Data Use Certification has also been updated and is available at http://gds.nih.gov/pdf/Model_DUC.pdf . More information on NIH Genomic Data Sharing may be found at http://gds.nih.gov

References

1. Current NIH-designated genomic data repositories are the National Center for Biotechnology Information's (NCBI's) database of Genotypes and Phenotypes (dbGaP) and the Sequence Read Archive, and repositories that have been established by NIH as trusted partnerships for the storage of NIH controlled-access data.
2. The National Institute of Standards and Technology (NIST) defines cloud computing as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." It defines a cloud service provider as an organization that offers some component of cloud computing to other businesses or individuals, typically Infrastructure as a Service (IaaS), Software as a Service (SaaS) or Platform as a Service (PaaS). See: http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf.

Inquiries

Please direct all inquiries to:

Genomic Data Sharing Policy Team

Office of Science Policy

6.4 Annex 4 - Compliance

Companies that do business on the Internet or in the cloud may fall under one or more compliance domains. In order to comply with regulations, companies must do everything in their power to adhere to data security guidelines within the various compliance standards. Examples include:

- US healthcare, which must follow HIPAA regulations as they relate to the protection of patient information.
- The European Union, which has Directive 95/46/EC regulation to ensure protection of personal data.

While compliance is difficult in its own right, cloud computing complicates the picture even further. If your organization is processing and/or storing sensitive data that is protected by compliance regulations, maintaining a compliant organization becomes a shared partnership between the CSP and the customer. Seven areas to watch:

1) Cloud Service Provider Is Compliant, You Are Not: it's easy to assume that as long as your CSP is compliant, so are you. This is far from the truth and leads to a scenario that includes fines and a potential halt of business operations if compliance failures persist. A key point: at the end of the day, the data controller is ultimately responsible for protecting the data This is not only true for the cloud in which you reside, but also the systems and applications you maintain within the cloud

2) Bad Tenants: Without proper supervision by the CSP, tenants could unwittingly introduce non-compliant processes that may bleed over into other tenants' space. While you can't control and audit other tenants within a cloud, you must verify that the service CSP has the proper security measures and processes in place so other tenants cannot impact your compliance duties

3) Where's My Data?: Cloud visibility remains a top problem in ensuring compliance. When data is moved between data centers and disaster recovery facilities, it becomes difficult to track where the data lives and how many copies of that data exist at any given time. As service CSPs grow, merge and upgrade, data may end up in places that are not considered compliant for the types of regulations your data requires

4) Poor Governance Strategy: The PCI Security Standards Council warns service CSPs and clients alike about clear lines of communication and tangible reporting and monitoring systems: "without a clear governance strategy, the client may be unaware of issues arising from use of the cloud service, and the [cloud service CSP] may be unaware of issues within the client environment that could impact their service provision."

5) That's Regulated Here?: for a global corporation, ensuring compliance is compounded due to the fact that countries, and even states or provinces within a country, can have widely differing compliance requirements. While a cloud CSP can offer assistance from the compliance knowledge it possesses, it's up to you to figure out what's needed, and whether your cloud CSP is certified under those compliance rules.

6) Failing To Regularly Audit The Service CSP: being compliant for something is simply a snapshot in time. Cloud services are still relatively new and very fluid. Because of this, it's up to the customer to regularly audit its service CSP to ensure they maintain the necessary certifications and meet the compliance standards that they originally met.

7) New or Modified Compliance Rules: Compliance rules grow and change on a regular basis. Whether due to new technologies, different interpretations, or a tightening of current standards, new and modified compliance rules must be reviewed regularly. Companies must stay on top of their service CSP to make sure that new or changed compliance rules are handled in a timely manner.