ACGT

Advancing Clinico Genomic Trials on Cancer

Risk analysis concerning the data security and data protection framework

Project Number: FP6-2005-IST-026996

Deliverable id: D 10.8

Deliverable name: Risk analysis concerning the data security and data protection framework

Date: 09.06.2010

Information Society Technologies

COVER AND CONTROL PAGE OF DOCUMENT	
Project Acronym:	ACGT
Project Full Name:	Advancing Clinico-Genomic Clinical Trials on Cancer: Open Grid Services for improving Medical Knowledge Discovery
Document id:	D 10.8
Document name:	Risk analysis concerning the data security and data protection framework
Document type (PU, INT, RE)	INT
Version:	FINAL
Date:	09.06.2010
Editor: Organisation: Address:	Brecht Claerhout Custodix NV Kortrijksesteenweg 214 b3 9830 Sin-Martens-Latem Belgium

Document type PU = public, INT = internal, RE = restricted

The ACGT Data Protection Framework (DPF) is a generic framework designed for dealing with data protection-related aspects in large, transnational research infrastructures such as ACGT. The DPF governs the processing of personal data within the ACGT project and specifies a combination of technical, contractual, legal and organisational data protection measures. The protection of patient privacy and the level of regulatory compliance that is attained within ACGT thus depend on the proper functioning of this framework. In order to assess the DPF's strengths and weaknesses, a risk analysis concerning data security and data protection was performed. This document aims to give a high-level overview of the data privacy risks within ACGT.

KEYWORD LIST: risk analysis, data protection, privacy, data protection framework, security, informed consent

MODIFICATION CONTROL			
Version	Date	Status	Author
1.0	10/06/2009	Initial Draft	S. Castille
1.1	02/10/2009	Draft	N. Forgó
2.0	09/04/2010	Draft	B. Claerhout
2.2	15/04/2010	Draft	B. Claerhout, D. Voets
2.3	5/5/2010	Draft	E. Egermann
3.0	9/6/2010	Final	B. Claerhout

List of Contributors (authors)

- E. Bonsma (Philips)
- S. Castille (Custodix)
- B. Claerhout (Custodix)
- E. Egermann (LUH)
- N. Forgó (LUH)
- A. Hoppe (USAAR)
- J. Karlsson (UMA)
- T. Krügel (LUH)
- A. Sanchez (UPM)
- J.-M. Van Gyseghem (FUNDP)
- D. Voets (Custodix)

CONTENTS

EXECUTIVE SUMMARY 6

1 INTRODUCTION 7

2 SCOPE 8

2.1 CONFIDENTIALITY 8
2.1.1 Privacy *8*
2.1.1.1 Context of Anonymity 8
2.1.1.2 Central Data Protection Authority 10
2.1.1.3 Trusted Third Party 10
2.1.2 Confidentiality and Open Access *11*
2.2 INTEGRITY 11
2.3 AVAILABILITY 12
2.4 PROJECT BOOTSTRAPPING RISKS 12
2.4.1 Partners not Signing Contracts *12*
2.4.2 Extensive Time Effort with respect to the Negotiation and Signing of Contracts by Data Exporters 13

3 NON-COMPLIANCE 14

3.1 INTRODUCTION 14
3.2 TECHNICAL COMPLIANCE 14
3.3 INFRASTRUCTURE MANAGEMENT PROCEDURES 15
3.3.1 User Registration and Removal *15*
3.3.2 Access Management *16*
3.4 THE HUMAN FACTOR 17
3.5 END-USER CREDENTIAL MANAGEMENT 18
3.6 LEGAL COMPLIANCE 18
3.6.1 De-facto Anonymisation *19*
3.6.1.1 Pseudonymisation 19
3.6.1.1.1 Lack of Pseudonymisation 19
3.6.1.1.2 Pseudonymisation key becomes known 20
3.6.1.1.3 Matching Procedure 20
3.6.1.1.4 Third Parties 20
3.6.1.2 Summary 21
3.6.2 Informed consent *21*
3.6.3 Additional Measures *23*

4 RISKS RELATED TO THE INFRASTRUCTURE AND DEPLOYED TECHNOLOGIES 24

4.1 SECURITY COMPONENTS 24
4.1.1 GAS *24*
4.1.1.1 Denial Of Service 25
4.1.1.2 GAS security 25
4.1.2 Data Access Services *26*
4.1.3 PKI *26*
4.1.3.1 Certificate Authority 26
4.1.3.2 OCSP service and CRL distribution point 26
4.1.3.3 Registration Authority 27
4.1.4 MyProxy *27*
4.2 OTHER RELEVANT COMPONENTS 28
4.2.1 ACGT Portal *28*
4.2.2 Resource, Data, Workflow Management *29*
4.2.3 Grid Nodes *29*

5 CONTEXTUAL ANONYMITY 30

5.1 INTRODUCTION 30
5.2 INFORMATION GAIN 30
5.2.1 Enriched Data 30
5.2.2 Unique Characteristics 30
5.2.3 Legal Consequences 31

6 LONG-RUN SUSTAINABILITY 31

6.1 PROJECT END 31
6.1.1 Who is responsible in case the CDP is gone? 31
6.1.2 EU – Legislation 32

7 CONCLUSION 33

8 ACRONYMS 34

Executive Summary

The ACGT Data Protection Framework (DPF) is a generic framework designed for dealing with data protection-related aspects in large, transnational research infrastructures such as ACGT. The DPF, described in deliverable D10.2 "The Ethical and Legal Requirements", governs the processing of personal data within the ACGT project and specifies a combination of technical, contractual, legal and organisational data protection measures.

The protection of patient privacy and the level of regulatory compliance that is attained within ACGT thus depend on the proper functioning of this framework. In order to assess the DPF's strengths and weaknesses, a risk analysis concerning data security and data protection was performed.

This document aims to give a high-level overview of the data privacy risks within ACGT. In this sense, no attempt was made to fully quantify risk, as estimating attack probabilities (estimating attacker gain and necessary investment) adds no value given the high number of assumptions that need to be made.

The main threats to the proper functioning of this framework that are discussed relate to:

- Non-compliance,
- Infrastructure and technology-related security risks,
- Risks associated to the concept of contextual anonymity,
- Risks related to sustainability,
- Risks introduced by long-term exploitation of the research infrastructure.

Most identified threats are associated with the leaking of de-identified data. In reality, these are likely to have a limited impact. In the end, it is highly unlikely that data will end up with people capable of re-identification (especially not with accidental breaches).

1 Introduction

The ACGT project has contributed considerable effort to solving the data protection-related issues which accompany the creation of a transnational biomedical research infrastructure. The goal of this work is to provide researchers with a convenient and easily implementable solution to deal with legal and regulatory compliance regarding data privacy. This solution was defined in the form of a framework consisting of a combination of technical, organisational and legal measures (D 10.2 “The ACGT legal and ethical requirements”).

This ACGT Data Protection Framework (DPF) builds upon the concept of “context of anonymity”, the establishment of a Data Protection Authority (DPA) and on the integration of a Trusted Third Party. In short, the main goal of the DPF is to ensure that within the ACGT environment, data can be considered “de-facto anonymous”. As such it frees researchers from the tedious administrative tasks connected to data protection compliance for each individual project, as long as they stick to the global ACGT rule set (procedures and contracts).

The protection of patient privacy and the level of regulatory compliance that is attained within ACGT thus depend on the proper functioning of the Data Protection Framework. In order to assess the DPF’s strengths and weaknesses, a risk analysis concerning data security and data protection was performed.

This document aims to give a high-level overview of the data privacy risks within ACGT. In this sense, no attempt was made to fully quantify risk, as estimating attack probabilities (estimating attacker gain and necessary investment) adds no value given the high number of assumptions that need to be made.

Four major issues are discussed in the following paragraphs:

- Non-compliance
    - Correct implementation and application of the DPF is an obvious requirement for its proper functioning. It needs therefore little explanation that non-compliance at all levels, be it technical, procedural or contractual, is one of the biggest risks.
- Infrastructure and technology-related security risks
- Contextual anonymity
    - Specific risk associated with the establishment of a controlled context in which data can be considered de-facto anonymous.
- Long-run sustainability
    - Risks associated with maintaining the research infrastructure on the long term.

2 Scope

2.1 Confidentiality

2.1.1 Privacy

Confidentiality is defined as "ensuring that information is accessible only to those authorized to have access" in ISO/IEC 27002.

Safeguarding the privacy of the involved patients is an important task within ACGT. Harming the privacy of the patient would have severe consequences for the trust of the patients and would very likely lead to patients withdrawing from the project and its clinical trials. Furthermore, privacy breaches are irreparable as an information leak cannot be undone. Therefore, it is of vital importance for the project to implement the Data Protection Framework (DPF) with all its fallback-scenarios. However, failure of the DPF can lead to loss of privacy. Numerous possible causes for the failing of the DPF can be discerned. As the DPF consists of a safety net with different supporting pillars, there are potential weaknesses in every pillar of this net. However, as fallback-mechanisms have been built into the DPF, for a total loss of privacy, a combination of multiple failures would have to occur at the same time.

Potential threats can relate to every pillar of the safety net, including the anonymisation of the data, guarding of the anonymous context, the consent forms and the national laws. Furthermore, every pillar can be associated with several factors which could lead to a loss of privacy in that particular context. These threats might cause a pillar to collapse; thus pose a risk to the data protection framework. They will be analysed later in this deliverable (see 3.6).

The DPF aims to divide equal attention to technical, organisational and legal measures. It builds on a context of anonymity, the establishment of a Data Protection Authority and on the integration of a Trusted Third Party. The main goals of these three parts of the DPF will be described in the following sections. Additionally, the possible threats that exist will be depicted in order to provide an overview before commencing the detailed analysis of the respective threats in Chapter 3.

2.1.1.1 Context of Anonymity

In order to establish a context of anonymity, all patients' data used for research and processed in the project must be pseudonymised. Pseudonymisation means replacing a person's name and other identifying characteristics with a label in order to preclude identification of the data subject or to render such identification substantially difficult. In addition to pseudonymisation, contracts have to be concluded between a governing body of the project which will be the Data Protection Authority (DPA) of the project on the one hand and each project partner on the other hand. This is needed in order to guarantee that the de-identified data are used only within the scope of the project and that each partner complies with the applicable data security standards. Thus, the context of anonymity consists of pseudonymisation and the establishment of a closed user group through contracts and security measures.

In order to accommodate for the use of pseudonymised data only, the data flow must be organised as follows: data concerning a patient are collected by the treating physician in a hospital, analysed and stored at the hospital. With the patient's consent, the physician may transmit the patient's data to the research project database. During transmission, the data are pseudonymised by removing all directly identifying characteristics and replacing them with a pseudonym. The key that connects a pseudonym to a particular patient (and thus allows re-identification) is stored at a Trusted Third Party (see 2.1.1.3). All partners within the project may not have access to these keys.

Within the closed user group, the project participants may not be able to identify the patient concerned with a proportionate amount of time, expense and labour. Additionally, all data must stay inside this closed user group. Therefore, the context of anonymity ensures that the data contained in the project's databases that are accessible for the project partners cannot be directly or indirectly linked to a person. Re-identification is only possible with additional information (e.g. the key for pseudonymisation). Thus, the data maintained and processed within the project can be classified as de-facto anonymous.

> De-identification is difficult and the state-of-the-art related to de-identification, re-identification and anonymity calculation shows that there are no fully satisfying practical solutions available (i.e. demonstrably secure in a practical environment). The balance between reduction of identifying information in a data-set and maintaining a sufficiently rich information content is difficult to find, and case-dependent.
>
> This means that in reality, if micro-data is required for research, re-identification is usually theoretically possible. ACGT advocates a pragmatic approach by establishing a "context of anonymity", and encouraging the use of practical tools such as the Custodix Anonymisation Tool (CAT).

The legal framework in support of the context of anonymity consists of contracts concluded between the Data Protection Authority (see 2.1.1.2) and each research organisation that wants to be part of the network. These contracts impose a legal obligation on the project participants not to undertake any steps in order to re-identify the patients. Furthermore, they contain obligations regarding the use of the security infrastructure.

Regarding the process of pseudonymisation, possible privacy breaches may be due to the fact that pseudonymisation was completely lacking or incomplete. The latter can be a consequence of the fact that data was either not completely pseudonymised (an identifying characteristic was not removed) or that the key needed to perform the pseudonymisation became known to a partner (or all partners) within the project, hereby rendering pseudonymisation ineffective. Each of these threats will be analysed in detail (see 3.6.1).

Threats to maintaining anonymity include failure of the security infrastructure or unlawful behaviour of one of the partners.

Threats related to the concept of the closed user group consist of the possibility of a project partner to re-identify the concerned data subject with a proportionate amount of time, expenses and labour and of the possibility that data becomes accessible outside the project.

2.1.1.2 Central Data Protection Authority

A central data protection authority must be installed in the project to control the partner's compliance with the aforementioned contracts and to serve as a central contact point for participants. The data protection authority needs to be legally capable to enter into binding contracts with the project participants and empowered to impose a penalty for infringement.

The Center for Data Protection (CDP)[1] acts as the central Data Protection Authority (DPA) within ACGT. It is a non-profit organisation under Belgian law.

The most important risk here is that the Data Protection Authority ceases to exist. Almost similarly important would be the loss of the ability to conclude and enforce the contracts. The data protection framework depends strongly on the CDP. Therefore, its possible loss of power would pose a great risk to the data protection framework and the privacy of the patients. The question of who would be responsible should the CDP cease to exist will be answered in the context of the long-run sustainability (see 6.1.1).

2.1.1.3 Trusted Third Party

A Trusted Third Party is an independent entity which facilitates interactions and establishes trust between several parties who all trust the third party.

Within ACGT, the Trusted Third Party implements appropriate technical and procedural measures to protect personal data against accidental or unlawful destruction or accidental loss, alteration, unauthorised disclosure or access, in particular if the processing involves the transmission of data via network, and against all other unlawful forms of processing. Having regard to the state of the art and the cost of their implementation, such measures shall ensure a level of security appropriate to the risks represented by the processing and the nature of the sensitive data to be protected.

The Trusted Third Party is a separate legal entity and in any case independent from the Data Protection Authority. Ideally, it is also not directly involved in the project. Most importantly, it shall be contractually bound to certain duties and responsibilities by an agreement with the Data Protection Authority. Among other things, the agreement establishes that the Trusted Third Party is only allowed to re-identify a data subject if a new medical treatment has been developed from which the patient concerned could benefit and if the particular patient wants to be informed. Apart from that, the Trusted Third Party has to safeguard the key which can be used to re-identify the patients against unauthorised access. Thus, the main task of the Trusted Third Party is to protect the key. It should be noted, that it is not very likely that it will have to re-identify a patient.

Possible risks for the data protection framework regarding the Trusted Third Party are on the one hand its possible cessation and on the other hand the inability (whether voluntarily or not) to fulfil its duties and responsibilities as laid down in the agreement with the Data Protection Authority. This could mean that the technical and organisational measures to protect personal data cannot/are not taken or that the key is used to re-identify a patient/data subject without either or both of the conditions established for this fulfilled.

[1] See http://www.privacypeople.org/

If either the Center for Data Protection (acting as DPA) or the Trusted Third Party cease to exist or cease to function according to the principles of the data protection framework, this poses such a great risk to privacy and confidentiality that the processing of personal data within the project's settings should immediately stop.

2.1.2 Confidentiality and Open Access

A number of works created within ACGT that contain information that needs to be kept secret are protected by copyright. There might also be information subject to secrecy protection that is not protected by copyright.

Regarding the copyrighted works that contain confidential information, it is important to use protection in a way that the information may not become known outside the project. Thus, the right holders need to ensure that those (parts of the) works that contain the secret information cannot be accessed. This restriction of access extends to all parties that do not have the obligation to keep information secret.

The Consortium Agreement (CA) establishes that "any access rights to knowledge and pre-existing know-how shall be granted only upon written request specifying the scope and duration of their application, and are subject to a separate agreement."[2] Furthermore, it states that "Access Rights to Software do not comprise access to *Source Code* but only to *Object Code*."[3] Additionally, the CA obliges the parties to keep information that is communicated on a confidential basis secret.[4] Thus, all project partners are obliged to keep secret information confidential.

If a third party gains access to secret information by acting maliciously, the project partners may claim damages. However, if the information is accessible due to insufficient care taken in the scope of the project, there is no legal remedy against this third party. In that case, remedy is only available against those partners that acted negligently.

2.2 Integrity

Integrity is defined as "safeguarding the accuracy and completeness of information and processing methods" in "ISO/IEC 27002"

Integrity (in the context of information security) deals with making sure that no data can be altered by unauthorised persons (either intentional or accidental). Integrity of the data used in ACGT is relevant for the correctness of research, and indirectly for patient safety (e.g. when basing treatment on ACGT results). It is not further discussed in this document.

[2] See Section IV.4.1.1, p. 58 of the CA.

[3] See Section IV.4.7.1 Access to Source Code for Software, p. 61 of the CA.

[4] See Section IV.6.1 Confidentiality obligations, p. 65 of the CA.

2.3 Availability

Availability means that end-users should be able to use the infrastructure and have access to the secured data as intended. Denial Of Service (DOS) attacks are threats to availability. Such an attack usually fits a larger hacking attempt (e.g. eliminating one of the vital services in order to render another service vulnerable). Unlike integrity and confidentiality, loss of availability is seldom permanent. Once services or data have been restored, normal operation can continue.

Except for mission critical systems (which a research platform like ACGT is not), the impact of attacks on availability is often limited to annoyance and waste of money due to inactivity of people and underuse of resources.

2.4 Project Bootstrapping Risks

2.4.1 Partners not Signing Contracts

A possible threat to the data protection framework is that the partners refuse to sign the contracts.

There are two kinds of contracts: the data transfer agreement and the contract on data protection and data security within ACGT.

The data transfer agreement is concerned with the data transfer from a data exporter to the ACGT network. It states that the healthcare organisation or hospital is responsible for and has control over data analysis and storage performed within the boundaries of the organisation including the data transfer to the Center for Data Protection (CDP). As soon as the data is transferred to the ACGT network its processing is under the sole responsibility and control of the CDP. It further poses on the data exporter the obligation – among others – to collect, process and transfer the data in accordance with the laws applicable to the data exporter, to transfer the data only if the patients concerned have been sufficiently informed and therefore have signed all corresponding ACGT consent forms, to implement a pseudonymisation tool and transfer data only after initiating pseudonymisation, to ensure technical and organisational separation of ACGT data from other data, to process data contained in the ACGT databases only if he/she has signed the end-user contract as well as not to run any matching procedures and not to use any other means to identify the data subject/patient concerned.

The contract on data protection and data security within ACGT states the conditions under which an ACGT end-user is allowed to retrieve data from the CDP and to conduct research as well as the obligations of the CDP regarding the ACGT end-users. It establishes that the CDP is responsible for the transfer of the data within the network, whereas the ACGT end-user processes data on its own responsibility within its own organisation as a data controller. Furthermore, it poses on the end-user the obligation – among others – to process the data in compliance with national law, Directive 95/46/EC and the provisions of the contract. The provisions of the contract oblige the end-user to ensure that the data from ACGT is technically and organisationally separated from other data, not to carry out any matching procedure or any other means in order to identify the data subject or patient concerned as well as to conduct only scientific research that is in line with the ACGT policies. Additionally, the end-user must not disclose data or any other information acquired from the CDP to any other person unless this is needed in pursuit of their duties as detailed in the contract.

However, the risk of not signing the contracts can be placed outside the scope of the project, because partners that have not signed the relevant contract either cannot access the data in the project or may not contribute any data to the project. This is due to the fact that without signing the contract, the partner is under no obligation to pseudonymise the data they contribute to the project or to use the specific consent form which was drafted for the use of the data within the framework. Both pseudonymisation and consent forms are important parts of the data protection framework. The de-facto anonymisation and consent are the respective first and second pillars of the safety net. Without an obligation to implement these two parts of the data protection framework, there is no legal certainty with regard to the processing of patient data, given the fact that the third fall-back scenario, i.e. the legal situation in the Member State of the data controlling partner, may change. To base the contribution to the project solely on Member State law, is therefore considered a risk which could easily lead to a loss of privacy. In order to avoid this risk, data of partners who have not signed the data transfer agreement cannot be contributed to the project.

Similarly, partners who have not signed the contract on data protection and data security within ACGT may not access the data contained in the project's databases. This is due to the fact that only the contract ensures that the context of anonymity is upheld by the data processors. Without the contract, partners might be allowed to carry out matching procedures which could lead to re-identification of the data subject or patient concerned.

Thus, if the partners do not sign the contracts, a situation arises in which no data is available in the project. Therefore, the signing of the contracts is crucial to the project.

2.4.2 Extensive Time Effort with respect to the Negotiation and Signing of Contracts by Data Exporters

Another outside risk can develop with respect to an extensive time effort regarding the negotiation and signing of contracts. The contracts are quite complex and specifically designed for ACGT. On the one hand, some of the data exporters or more precisely their legal departments might not want to include certain clauses or might not be able under their national law to do so. On the other hand, the contracts can only be changed individually for each partner to a certain extent; extensive changes would make management of the contracts very difficult. Moreover, changing those clauses in the contracts that relate to the data protection framework is not possible, because this would undermine the framework. The contracts are optimised for the data protection framework and any change would weaken its constitution. Therefore, there is not much room for negotiation which leads to the problem of data exporters or possible end-users not signing the contracts and the consequences stated above.

The severe risk arising from this situation is a deadlock with respect to the delivery of data into the project as well as to the export of the data.

3 Non-compliance

3.1 Introduction

Procedures, contracts, End User License Agreements (EULA) and other (legal) agreements have been introduced to fill the gaps that technical security measures leave open.

These gaps remain due to architecture and design limitations or simply because it is technically impossible to prevent certain actions. For example: a system administration can forbear to disable accounts of people leaving an organisation (not following employee removal policies): those people can still access data, although they are no longer authorised. A second aspect of compliance is the technical compliance to ACGT best practices (cf. D11.4) of all components (services) branched into the infrastructure. A security infrastructure is designed assuming that all services deployed fulfil a minimum set of security requirements. If this minimum set is not fulfilled, the security infrastructure will not function properly.

The contracts of the project include the “Consortium Agreement (CA)”, “data transfer agreement” and the “contract on data protection and data security within ACGT”. The contracts refer to the “General Terms”. The contract on data protection and data security within ACGT is supplemented by an annex dealing with “technical and organisational measures” and an annex called “the ACGT end-user agreement”. The data transfer agreement refers to the informed consent given by the patients or their legal representatives.

Non-compliance (all forms) can arise from ignorance, sloppiness or malicious intent. In all cases, the consequences can be devastating as technical security architecture can only function properly if the legitimate users adhere to the policies and the different components behave as expected.

Only organisations and people who signed a contract which specifically includes financial penalties for non-compliance to the ACGT use policy are accepted into the ACGT community. Financial penalty is a typical and effective stimulant for enabling compliance. In addition to that, people not complying with applicable law may face criminal prosecution.

3.2 Technical Compliance

ACGT services can receive the “ACGT enabled” label if they are deployed in compliance with D11.4. Technical compliance goes further than mandatory integration of several ACGT security components. Software must be built with security in mind and securely deployed. Failure to secure the environment, in which ACGT services are deployed, can open up additional attack vectors that could enable outsiders to compromise the service.

Compliance with D11.4 is a minimal requirement for ACGT enabled services. Since the amount and type of services running on the system vary from host to host, it is infeasible to make a detailed risk analysis. The potential impact resulting from non-compliance with the minimal security requirements depends from service to service and can range from Denial of Service to leaking of secure data.

Services thus need to be validated (cf. quality assurance) before being allowed to operate on the ACGT infrastructure (at least for dealing with sensitive data). Periodic re-evaluation is a must. If the ACGT platform is to be exploited on the long run, this rather informal validation should be replaced with a formal certification process.

3.3 Infrastructure Management Procedures

Considerable risks related to confidentiality are associated with (in)correct management of the infrastructure (any information system for that matter). Management is performed by humans, who are prone to honest mistakes and can be instigated to abuse.

Management procedures are put in place to reduce the risk of mistakes and facilitate the detection (and handling) of malicious actions. Compliance to the management procedures that are put in place is a prerequisite to a secure system.

3.3.1 User Registration and Removal

The DPF depends on the ACGT Registration Authority (RA) to verify the identity of the ACGT end-users when enrolling them. Not only is their identity used to assign roles, but all users need to be accountable for their actions and therefore be legally bound to the ACGT usage policy. These stringent demands on user registration require physical verification of identity, which cannot be practically done by a central registration authority due to the international nature of the project.

Delegate Registration Authorities are therefore created in several places to perform local identity checks. In practice, a delegate RA is assigned to each organisation that participates in ACGT. In that way, each organisation manages the enrolment for (and removal of) its own “employees”. This approach has the possible benefit that user lists are better kept up-to-date (removal from ACGT should be included in the employee removal procedure). Before a delegate RA is assigned to an organisation, it must agree to the procedures and regulations set by the Data Protection Framework (DPF) and sign a contract with the Center for Data Protection.

The main threats coming from delegate RAs are insufficient validation of new end-users and late or non-removal of accounts of users that no longer take part in ACGT as well as registration of fake users (malicious intent of the RA or one of its employees, see also 3.4). These threats pose considerable risks. Moreover, executing oversight over the RAs operational procedures is difficult due to their geographical spreading. Although little access is granted by simply registering users (some resources are free for all though), some policies might grant access to members of an organisation or Virtual Organisation (VO) and resource managers might be tricked into giving that fake user access to a restricted resource. This impacts confidentiality (IP, breach of the context of anonymity) and integrity.

Note that when fake accounts are used, audit trails are of little help to handle incidents and try to find those accountable for the abuse.

3.3.2 Access Management

It needs no explanation that correct management of access control (AC) is a requirement for assuring integrity and availability of the ACGT resources. In the ACGT collaborative environment, access needs to be managed at different logical and infrastructural levels: Grid level, VO level, Resource level. The different responsibilities also reside with different people and organisations.

This paragraph roughly (non-exhaustively) lists the three main levels at which access control of the ACGT infrastructure is managed. The risk associated with incorrect access management (non-compliance) within ACGT is the possible erroneous elevation of access rights (with the associated issue of a possible privacy breach or IP issue) or removal of access rights (availability issue) of registered ACGT users (cf. user registration). It needs no explanation that the possible impact is extremely high.[5]

- **Grid Management**
  Grid management is performed by a few dedicated people whose only ACGT responsibility is to manage security at the lower level of the Authorisation Service. Those Grid Managers are the only ones able to add resources to the ACGT infrastructure (they are thus responsible for checking resource compliance). As these people have a "super user" status over the central Authorisation service, they are in principle capable of accessing all resources on the ACGT infrastructure. It is clear that only trustworthy security professionals can fulfil this role. Further, it is encouraged that additional operational constraints are built into authorisation service management in order to minimise the risk on unauthorised access (e.g. administration login could be location-based).

- **VO Management**
  Within ACGT, the different dynamic groups of collaborating partners that originate from the different trials and research initiatives are reflected in the formation of Virtual Organisations (VOs). The initiative to form such groups originates from the end-users themselves, it is therefore logical that they are also responsible for VO management.

  Grid Managers can delegate VO management capabilities to specific users by appointing them the role of VO Manager. VOs created by a VO Manager can be managed by a Grid Manager or by other end-users assigned as VO administrators for that specific VO. The VO Manager is responsible for VO user management (VO membership) and can manage access to all VO resources. He can thus also add unauthorised persons (ACGT members) to a VO and grant them access to all VO resources (i.e. only resources which have been allocated to the VO, thus not all ACGT resources or resources belonging only to other VOs).

[5] All actions performed by the Resource and VO managers, including those assigned for sensitive data are logged for auditing. This enables the ACGT DPF to regularly check for abuse and inconsistencies.

The fact that VO Managers are typically neither security professionals, nor used to dealing with management elevates the risk of non-compliance. These people might be more inclined to cut the corners procedure wise because they are not familiar with the common threats and their possible impact. Still, VO Managers need to be explicitly appointed and thus be educated for awareness and proper knowledge of the guidelines.

- **Resource Management**
  End-users can share resources with ACGT. When sharing a resource, an end-user is considered to be the owner of that resource and is thus able to manage access control rules for that resource. For example (s)he can decide to share it with the whole of ACGT, or assign it to a VO (implying that a VO Manager will manage part of the access). The resource owners can also assign other end-users as access managers for their resources.

  A special rule is applied to data resources that are protected by the Data Protection Framework. The rule stipulates that these resources are managed by a DPA assigned administrator instead of the owner of the resource (can be the owner if (s)he is properly trained).

  In general, end-users should be educated in order to raise awareness about the potential implications that may follow from the publishing of sensitive data to the whole ACGT community or to a specific VO (see also 3.4).

3.4 The Human Factor

It is a known fact that the majority of successful attacks is not performed by clever technical hacks, but rather by exploiting the human aspects associated with security. “Insider attacks” and “social engineering” are therefore listed among the top threats to any security system.

An insider attack is basically a user abusing his lawfully attributed rights to perform an unlawful action. It is deliberate non-compliance by users (usually administrators). The motif for this action can range from curiosity and revenge to bribes. It goes without saying that the impact associated with an “inside attack” depends on the rights that were attributed to the user in question. The remarks made in the previous paragraphs about the different actors in the ACGT management are thus valid here.

> *“You could spend a fortune purchasing technology and services... and your network infrastructure could still remain vulnerable to old-fashioned manipulation.”*
> - Kevin Mitnick
>
> (convicted computer criminal, infamous for social engineering attacks)

Social engineering is the act of manipulating people into performing actions or divulging confidential information. Through social engineering one tries to incite people to behave in a non-compliant way (i.e. not to follow their normal procedure) by threatening them or relying on their helpfulness or laziness.

Again, this has most impact at the system management level: user registration and access control management. Through social engineering, a person might convince an ACGT RA to skip the identity validation so he can register under a false identity. A VO Manager or Resource manager might be convinced to grant access to resources the person should not be able to access. This opens the path to unauthorised data access, either directly or through a breach of contextual anonymity.

The fact that ACGT is an international collaboration in which many strongly hierarchical organisations participate increases the probability that a social engineering attack might work (people don't know each other personally, it is easier to impersonate a high-ranked person from a foreign organisation, ...).

3.5 End-user Credential Management

At registration each end-user receives personal credentials in the form of a secret key and certificate. These credentials are strictly personal and must never be shared with other people. Failure to keep the credentials secure and personal can lead to unauthorised access by third parties to all data, information and services to which the end-user has access (cf. same remark on impact as listed above).

Users must thus be made aware of the need to protect their credentials and the consequences of sharing credentials, such as data leaking but also the fact that the end-user is responsible for actions taken with his/her credentials. The latter in particular is an effective way for mitigating the risk of credential sharing.

One aspect of great importance is the compliance to standard best-practices of password management (i.e. sufficiently complex and random passwords, not writing them down, not using the same password for every application, etc.). When using ACGT services, users are obliged to choose passwords at several points in time. The most critical point is at registration, where the main ACGT credentials are generated and stored (password-protected) on the client computer (or USB stick for that matter). Users are free to choose their password. Although technical measures help to avoid choosing a weak password, proper handling (i.e. not writing it down) is the responsibility of the end-user. The same is true for the selection of MyProxy delegation passphrases (cf. ACGT passport and visa).

Furthermore, it is also necessary for users to report any suspected compromise of abuse of their credentials as soon as possible, so that their credentials can be revoked and access rights adjusted.

3.6 Legal Compliance

Non-compliance with the contracts by a partner after signing them is one of the threats that might lead to a loss of privacy. From a legal perspective, there are two major threats if a partner does not comply with the contracts: the conduct of the partner leads to a failure of the de-facto anonymisation or the informed consent of a patient is lacking. Due to the complexity of the DPF, partners may unintentionally violate either contract and thereby threaten to harm the framework. However, a partner could also simply refuse to comply with the DPF. Thus, one needs to consider both intentional and unintentional behaviour in the risk estimation.

However, as the difference does not lie in the risks themselves, but in the consequences this has for the partners, these will be described after the assessment of the risks (see 3.6.3).

In the following, these forms of behaviour and the threats to the data protection framework resulting thereof will be examined. The two pillars of the safety net involved in this are the de-facto anonymisation and consent. Both are (at least partly) laid down in the relevant contracts. Thus, the CDP may enforce compliance in this field. However, compliance with national law cannot be enforced within the project and is to be ascertained by the project partners.

3.6.1 De-facto Anonymisation

De-facto anonymisation is the first pillar of the safety net. It ensures that the data processed within the framework of ACGT can be considered as non-personal data and that the researchers therefore do not risk unlawful processing of data. It thus forms part of the context of anonymity.

However, in case de-facto anonymisation fails, the researchers can rely on the consent forms signed by the patients. Therefore, it is important to know for which cases the consent forms play an important role in upholding the data protection framework.

The threats that exist regarding de-facto anonymisation are either related to pseudonymisation or to the closed user group concept (security infrastructure).

Both types of threats are of equal importance to the data protection framework.

3.6.1.1 Pseudonymisation

Pseudonymisation may fail for a number of reasons: pseudonymisation is lacking in total, pseudonymisation measures were incorrectly implemented (i.e. not according to specifications), pseudonymisation was not successful (i.e. an identifying characteristic was not removed) or the key required for re-identification became known to a data processing partner, rendering the data personal again. Other failures include successful matching procedures carried out on the pseudonymised data or data exposures to third parties.

3.6.1.1.1 Lack of Pseudonymisation

Possibly, a project partner refuses to use the pseudonymisation tool when contributing data to the project’s databases. Irrespective of the fact that this constitutes a violation of the data transfer agreement, this also is a considerable threat to the de-facto anonymisation. Another possibility is that a project partner accidentally does not use the pseudonymisation tool when contributing data to the project’s databases. In both cases, personal data would be available. This is due to the fact that the pseudonymisation tool removes all directly identifying characteristics and replaces them by a pseudonym. As examined in D10.2 this data is de-facto anonymous, if it is only processed in a closed user group like ACGT.

Similarly, in case not all directly identifying characteristics are removed, the data remains personal so that data protection legislation applies. This could be due to a wrong use of the pseudonymisation tool or to a failure of the software itself (a bug).

In this case, legitimate processing of data is only possible with an informed consent of the patient or a legal basis (for example an exemption for scientific research as foreseen in Art.

13 (2) Directive 95/46/EC). Both the General Terms and the Patient information sheet (D 10.1) state that the data collected by the hospital/investigator will be pseudonymised. However, the Patient information sheet which is included into the consent states under '8. Risks and benefits of this project': "*Before your data can be used for research, personal information (e.g. name, address, etc.) is disconnected from data identifiers (…). There is a residual, albeit extremely small risk that these data might be linked back to your person.*" Therefore, the consent also covers the processing of data that was - due to a failure of the context of anonymity for whatever reason - not pseudonymised by the hospital or investigator. Whether the hospital or investigator acted on purpose is from a purely legal point-of-view irrelevant.

The risk of pseudonymisation lacking is in itself low. However, due to the fact that the informed consent acts as a safety net in this case, the consequences for the data protection framework are minimal.

3.6.1.1.2 Pseudonymisation key becomes known

Another threat to the de-facto anonymisation is the possibility that the pseudonymisation key becomes known to a project partner, especially if this concerns a partner that is processing de-facto anonymous data. As soon as the key becomes known, the data is re-identifiable without much effort and thus has to be classified as personal. This might happen by mistake during configuration of the de-identification tools or due to an issue with the Trusted Third Party security infrastructure.

This risk is, however, considered to be very low.

3.6.1.1.3 Matching Procedure

Carrying out a matching procedure in which data from the project is matched with other (nominative) data e.g. stemming from earlier (or later) trials constitutes a violation of both the data transfer agreement and the contract on data protection and data security within ACGT. If the de-facto anonymised data is successfully matched with non-anonymous data, the data becomes personal again. All data-sets that escape out of the context of anonymity can no longer be qualified as de-facto anonymous data, but have to be qualified as personal data.

Without the anonymisation, the processing of data is only lawful if the patient consented to the processing or if there is a legal basis. The matching of data is covered by the clause stating that there is an extremely small risk that data may become personal again. Therefore, matching of data is covered by the patient's consent, and thus the second pillar of the safety net applies here.

The risk that a matching procedure is carried out is not high. As the patients' consent includes this scenario, the threat to the data protection framework can be considered to be low.

3.6.1.1.4 Third Parties

Furthermore, a partner could violate either contract by releasing data to a third party. This could happen both intentionally and unintentionally, e.g. by releasing research findings in order to publish in a scientific journal or by transferring the data. In case the data is released, third parties could use the data to match it with other data they have. By matching the data-sets, they could be able to re-identify data subjects from the pseudonymised data-set and (at least part of) the data would become personal. Again, the context of anonymity is destroyed.

As stated above, without anonymisation, the processing of data is only lawful if the patient consented to the processing or if there is a legal basis. As the consent forms do not include transferring the data to third parties, either separate consent would have to be given or a legal basis is needed. Again, relying on the different national laws poses a threat to the data protection framework as the laws might be subject to change.

The risk that data is released to a third party is not very high. However, as the informed consent cannot act as a safety net in this scenario, but the data controller has to rely on national law, the threat to the data protection framework is the highest of those described in this section. Still, the overall threat is not serious.

3.6.1.2 Summary

It can be seen from the aforementioned, that the context of anonymity is the core of the data protection framework. If it is destroyed, the lawfulness of data processing relies on the consent of the patient and the legal situation in the respective Member State. If the pseudonymisation failed, the key became known or a matching procedure was carried out, thus rendering the data personal again, the informed consent can be used as a safety net. However, this is not the case with respect to third party exposure. In case national law does not offer an exemption for the latter or an existing exemption is later on repealed, this could constitute unlawful processing of data.

Most of the risks described above can be lessened by enforcing the contracts within ACGT.

Unlawful processing of data is penalised under the law of all European Member States and can be the subject of a fine or a claim to damages. The fine or claim has to be directed at the person or institution responsible for the processing of the data. Depending on the circumstances, either the institution carrying out the unlawful processing or the Data Protection Authority is responsible. For any data processing within the project, the responsible institution is the Data Protection Authority, which is the CDP in case of ACGT. For other data processing, this is the institution carrying out the processing. The data transfer agreement states that the healthcare organisation or hospital is responsible for data analysis and storage within its own organisation. Similarly, the contract on data protection and data security within ACGT states that end-users process data on their own responsibility within their own organisation as a data controller.

Still, any data processing that remains in the project's sphere lies in the responsibility of the CDP. However, a partner violating either contract is obliged to relieve the Data Protection Authority of any claims to damages or fines.

3.6.2 Informed consent

Informed consent of the patients constitutes the second pillar of the data protection framework. This is the safety net in case the de-facto anonymisation failed. Therefore, obtaining the informed consent is very important for the data protection framework. Additionally, this has several advantages: it involves the patient in the whole procedure, thereby leading to transparency and generating trust; it gives legal certainty regarding the processing of data as well as covers ethical requirements.

However, in order for the consent to be valid, the patients (or their legal representatives) have to give it freely i.e. voluntarily knowing all circumstances of the processing of their data. Therefore, all data flows and all data processing have to be transparent. This means that the patients have to be informed about the identity of the data controller, the purpose of the processing of their data, their rights, the security of the data processing, the categories of data concerned and the recipients or categories of recipients. Furthermore, the data processing institution has to be able to prove that consent was given. All this is regarded if a patient freely signed the consent forms drafted for ACGT.

However, the informed consent may be faulty. This may be due to the wrong consent form being used, consent not being given voluntarily or by the wrong person, the patient not being properly informed beforehand or the consent form not being signed. These issues vary slightly in their degree of severity with respect to the project's data protection framework.

There are three basic distinctions to be made. First, consent exists but there is no proof. Second, no consent exists at all. And third, consent exists but is invalid.

In case the consent was not given voluntarily or by the wrong person (e.g. there should have been a legal representative), it is fairly clear that consent itself is lacking. The same applies if the patient hasn´t been informed properly beforehand as such consent is considered invalid.

In case the consent form was not signed, but the patient was duly informed and orally consented, there is simply no proof that the consent exists.

In case the wrong consent form was used, it may be either way.

If there is no informed consent and consequently no second pillar of the safety net, all processing of personal data would be unlawful, if a national permission cannot be found (assumed that the context of anonymity does not exist either).

In case the consent existed, but there is no proof, this constitutes a problem in case of auditing or legal action. However, as soon as the consent form is signed, this problem is solved. It is important to note in this context that later consent does not lead to an ex ante healing of unlawful data processing. Consent is only valid ex post. Therefore, informed consent needs to be given before any data is being processed within the project.

Thus, the worst case scenario would be that a partner forces a patient to sign the (wrong) consent form and afterwards exports the data to the project's databases without using the pseudonymisation tool. In this case, the context of anonymity is destroyed and informed consent was not given. Thus, the lawfulness of the data processing relies on the applicable national law. As health care data is classified as sensitive data, there are high requirements to allow lawful processing. Directive 95/46/EC allows Member States in Art. 13 (2) to introduce exemptions for scientific research and some states have introduced such exemptions. However, such an exemption does not exist in every national law and it is possible that an existing exemption is repealed by later legislation. Thus, without a context of anonymity and without the consent, the data protection framework is based on a very frail third pillar.

This example shows that any violation of the data transfer agreement of a partner regarding the informed consent leads to a great risk within the project as the second pillar of the data protection framework would be destroyed. If additionally the context of anonymity is destroyed, we have to turn to the third pillar, the national law, for help. Taking all different national legislation into account will lead, for different reasons, to a complex and legally uncertain infrastructure. On the one hand, such a framework would have to be modified whenever a partner from another Member State joins the project; and on the other hand, it is organisationally impossible to keep the framework in line with several national laws, as they may change permanently. This again shows that relying solely on Member State law poses a threat to the stability of the data protection framework.

However, the (substantial) risks described above can be mitigated by enforcing the contracts within ACGT to ensure the context of anonymity and auditing the procedural elements of the Data Protection Framework.

3.6.3 Additional Measures

If in any case of the aforementioned possible threats, patients or other entitled persons claim damages or impose a fine and the claim or fine was directed at the project, i.e. the CDP, the CDP can claim compensation for the costs incurred.

Furthermore, both the data transfer agreement and the contract on data protection and data security foresee a penalty for negligent breach of contract. Thus, even in the case there was no claim or fine from a third party, there is a liability of the partner that negligently violated either contract. This measure ensures compliance with both the data transfer agreement and the contract on data protection and data security within ACGT.

Another threatening measure in case of non-compliance is exclusion: The General Terms of ACGT state in Article 3 that ACGT has the right to exclude a physician, hospital/investigator or user (researcher) from the project in case of violation of the general terms, contracts, agreements or informed consents of ACGT or national and international legislation.

4 Risks related to the infrastructure and deployed technologies

The ACGT project defines a set of interdependent base components that form the foundation of the entire infrastructure. Risks associated with each of these individual components logically have an impact on the complete infrastructure. This chapter gives an overview of these risks.

4.1 Security components

The ACGT infrastructure is a Grid infrastructure based upon the GT4 and GRIDGE middleware. As such, the foundation of the security infrastructure lies with GSI (Grid Security Infrastructure), which is widely discussed. The Grid Security Infrastructure (GSI) is a widely-used specification for secret, tamper-proof, delegatable communication between software components in a Grid computing environment. Only the most important aspects are mentioned here.

Authentication and delegation is covered by GSI:

- **Based on PKI technology:** Authentication of services and end-users is based on x509 certificates. These certificates are issued by the ACGT PKI. The PKI also provides live services for certificate verification. The PKI is thus necessary not only for end-user registrations but also for providing up-to-date information about the status of the certificates.

- **MyProxy credentials repository:** Grid operations rely on credential delegation to allow for complex workflows and long running tasks. The MyProxy credentials repository allows end-users to store temporary proxy certificates and make them available to execution agents that perform tasks for the end-users.

ACGT relies on a centralised authorisation service, part of the GRIDGE middleware:

- **GAS:** The Gridge Authorization Services is a centralised Access Control List (ACL) repository and Policy Decision Point (PDP) for ACGT services.

- **ACGT Data Access Services:** The data wrappers are the gateway between the ACGT infrastructure and the ACGT databases. They act as a Policy Enforcement Point (PEP) for access to ACGT data.

4.1.1 GAS

As a central PDP for ACGT, GAS is one of the most critical services in the ACGT infrastructure. As all services depend on this central GAS service for authorisation decisions, it is both an attractive point of attack and single point of failure.

4.1.1.1 Denial Of Service

A Denial of Service (DOS) situation on GAS (intentional or not) causes serious disruption to the entire ACGT infrastructure. A DOS can be the result of a deliberate action (see 2.3) or other causes (power failure, network failure, …). Irrespective of that, the impact on ACGT will be huge. For obvious reasons, all ACGT services relying on GAS are configured to deny all requests if they fail to obtain an access control decision from GAS. A DOS situation on GAS should thus be considered to cause little risk to the confidentiality and integrity of the infrastructure, but has a huge impact on availability (deliberate trade-off).

Several measures have been taken to reduce the probability and impact of such an event. The operational GAS service is deployed according to state-of-the-art measures for avoiding downtime. It is hosted in a commercial-grade environment (with frequent backups, redundant power supply and network setup, liveliness monitoring, etc.). It is possible to deploy a redundant GAS setup; however, in this research project (where a DOS attack is improbable), this choice has not been made due to the involved operational cost.

4.1.1.2 GAS security

The GAS service is protected against unauthorised access (both for administration as for services requesting AC decisions) by requiring authentication with a valid ACGT certificate over an encrypted communication channel. This shields the service from anyone or every service not in possession of a valid ACGT credential.

At the moment there are no known vulnerabilities in GAS that can cause the service to crash, make unauthorised changes or issue incorrect authorisation responses. Note that the GAS authentication implementation is based on widely used and reviewed standard implementations (GSI, OpenSSL).

As mentioned earlier, a risk with serious impact is the loss of credentials by GAS managers (see 3.3). The impact of this attack is limited by the actions authorised by GAS based on these credentials. Each ACGT access control rule has an ACL associated with it. On a per rule basis, a decision can be made on who can read or alter that specific ACGT access rule. This prevents users and administrators from changing rules that are not managed by them. If the attacker obtains credentials of a normal end-user, then only ACLs of resources managed by that end-user can be compromised. With credentials of administrators of secure data or even the ACGT Grid Manager, the impact of the attack is much higher: sensitive data can be compromised.

Several measures have been taken to reduce the probability of both malicious attacks and mistakes made by authorised administrators:

- Feature-limited configuration tools: The configuration tools used by VO administrators and Resource Administrators are feature-limited. The tools can only change a small subset of ACLs stored in GAS (protection against accidental exposure or data loss).

- Restricted administration for secure data: Administration of access rules for secure data under control of the CDP is limited to a small set of trained administrators.

4.1.2 Data Access Services

The data access services act as a Policy Enforcement Point (PEP) for access to ACGT data (no sensitive data should be accessible through other means). A new instance is created for each distinct data-set that is made available to ACGT.

4.1.3 PKI

The ACGT PKI is composed of several interdependent modules. The Certificate Authority (CA) module is the central component that issues and signs certificates, the OCSP service provides real-time information on the validity of issued certificates while Certificate Revocation Lists (CRLs) provide this information at predefined intervals. The registration site allows user registration and management by Registration Authorities.

X509 PKI infrastructures are widely used and well known. The implementation that is used within ACGT is also commercially deployed.

4.1.3.1 Certificate Authority

The CA component is used only for creation of end-user and service certificates. The CA, based on the EJBCA software, is shielded from end-users by firewalls and has access control enabled. The ACGT Registration can contact the CA through a secured connection and on a custom built module. Additional protective measures for the CA include the use of Hardware Security Modules (HSM) to store the secret keys.

The probability of unauthorised access to the CA module is extremely small. Access to the CA requires authentication using physical tokens and operations that change the CA configuration require two out of five physical tokens. In the event that an unauthorised party does gain control over the CA he/she will be able to issue new ACGT certificates, giving him/her the possibility to impersonate other ACGT end-users and access the data they have access to or to revoke valid certificates leading to a Denial of Service. However, due to the protective measures taken (physical access control, firewalling, ..) this scenario is very unlikely.

Certificates are created automatically and do not require any human interaction. All communication between the CA component and the RA component is encrypted and authenticated.

4.1.3.2 OCSP service and CRL distribution point

The OCSP service and CRL distribution point provide up-to-date information about the validity of ACGT certificates. This information is essential for services that handle secure data in order to make sure that compromised certificates are denied access immediately. The OCSP service allows for real-time revocation status checking for each certificate that is encountered during credential validation while a CRL is valid for a longer time. The choice between using OSCP and CRL is a trade-off between security and overhead: while a CRL causes much less overhead and delay, there is a limited lap time between the time a certificate is revoked and the time the CRL is updated.

Both the OCSP and CRL services provide public information. Therefore, there is no risk of sensitive data being leaked. The integrity of the OCSP queries and CRLs is protected by several mechanisms; encrypted communication not only provides confidentiality but also integrity controls, the CRLs and OCSP responses are signed by the CA and OCSP service so that answers can be validated. The services are read-only and typically retrieve their data directly from a CA module. CRL or OCSP fraud is thus very unlikely.

The most likely attack would be a Denial of Service. Through a DOS attack on the OCSP service, one could delay (DOS attack is assumed to be only temporarily possible) the propagation of the revocation status of certificates. It might thus allow revoked users to still access resources after revocation of their certificates (as far as they have not been removed from e.g. GAS). The impact of this scenario is limited.

4.1.3.3 Registration Authority

The ACGT registration site represents the Registration Authority of the ACGT PKI. Through the registration site, end-users obtain their main ACGT credentials. These credentials consist of a private key (which is password protected) and the corresponding certificate. They are called an *ACGT passport*.

Users submit a registration request which can be approved (or denied) by the appropriate RA (RAs also perform this action through the registration site). Upon approval an end-user receives a one-time activation key which authorises the credential generation process. The Registration site connects to the ACGT CA through a VPN.

Compromising the registration site or an RA account could allow an attacker to enrol fake users. The impact of this depends on the default rights given to a user of a certain organisation. However, no resources containing sensitive data should allow access to “all” people from a certain organisation, regardless of their role, thus restricting the possible impact of the issue.

4.1.4 MyProxy

Complex and long running tasks require that the end-users are able to delegate credentials to software agents such as the workflow enactor, so that these agents can act on behalf of the user while the user is offline. Delegation can also be used in case an end-user wishes to use his/her credentials on another computer and wants to minimise the risk of credential disclosure. Delegation means that a user “transfers” his access rights (typically for a restricted period of time) to another actor (service). Thus, using a delegated credential has potentially[6] the same effect as using a normal end-user credential, the same data can be accessed and the same services used. Therefore, it is very important that the delegatee can be trusted, the delegated credential’s validity is limited in time and that the delegated credential is securely maintained.

Delegation in ACGT is provided through GSI and a MyProxy service. Delegated credentials can be stored in the central ACGT MyProxy credential repository in a secure way, protected by a username and password chosen by the end-users. An ACGT proxy certificate is called a Visa.

[6] Although the delegatee acts on behalf of the delegator, resource managers can differentiate between user credentials and delegated credentials and thus impose additional restrictions on the operations that can be performed on behalf of the user.

As the repository is a central place where all delegated credentials are managed, it is an interesting target for attackers.

MyProxy is widely used and thus also assessed for security and vulnerability[7,8]. Proxy credentials are stored in an encrypted way, hence compromising the repository does not unlock all credentials. It would, however, allow a hacker to get access to all user credentials generated after the MyProxy service is taken over by the attacker. A successful attack on (and continued unnoticed control over) the hardened MyProxy service is unlikely.

There are inherent risks associated with delegation (one should trust to whom one delegates). There are measures which reduce this risk: delegate credentials can never outlive the original credential (possible abuse is thus limited in time), restrictions can be set on who is allowed to store and request delegations and access control restrictions can be imposed on delegation-based authentication tokens.

In the initial test phase, ACGT adopted a liberal policy in which all ACGT users and services were allowed to request and store delegated credentials. This implies that a high degree of trust is placed on services accepted in the ACGT infrastructure. As the maturity of the ACGT infrastructure and services has significantly increased in the mean time, delegation capabilities will be subjected to stricter requirements and conditions.

In conclusion, the security of this service depends to a high degree on the adoption of a strong password policy for protecting access to delegated credentials.

4.2 Other Relevant Components

4.2.1 ACGT Portal

The portal site is the entry point for end-users into the ACGT infrastructure[9]. It offers a graphical interface to the ACGT services. Users log in on the portal using the username and password that was used when storing the delegated credential on the ACGT MyProxy server. The portal is thus privileged to contact the MyProxy server and request delegations, and therefore a potential target for malicious persons trying to obtain valid (proxy) credentials of other users.

Compromising the ACGT portal would allow an attacker to obtain delegated credentials from end-users. However, portal logout triggers the deletion of a user's delegated credentials. If a user forgets to logout, the delegated credential will expire according to the lifetime as specified when requested from the MyProxy service. An intruder will thus only be able to use delegated credentials for currently logged-in users and be allowed to retrieve delegated

[7] J. A. Kupsch, B. P. Miller, MyProxy Vulnerability Assessment, May 2008. http://www.cs.wisc.edu/mist/papers/myproxy_vuln_assess.html

[8] Vineet Kumar Saini, Qiang Duan. Attack Tree-Based Security Analysis for MyProxy Online Credential Repository. http://research.acxiom.com/downloads/ALAR2007_Proceedings/0-0-0--Saini-Duan--Security-Analysis-for-MyProxy-Credential-Repository--paper.pdf

[9] There can be multiple instances of the portal, so it should not be considered as a DOS single point of failure.

credentials from users subsequently logging in. As mentioned earlier, long-time exploitation of a well- managed host is very unlikely (note also that the technologies used in the portal site are widely used and regularly updated).

4.2.2 Resource, Data, Workflow Management

Any actor that (workflow enactor, GRMS and GDMS) acts as an agent for end-users that wish to perform complex tasks on the ACGT infrastructure needs proxied credentials to inherit the access rights of the user.

In that sense, similar to the portal, such services are interesting to attackers. Access is restricted to them as much as possible.

If such a service was compromised, an intruder would be able to obtain proxy credentials for the users actively using such a service.

4.2.3 Grid Nodes

Grid nodes represent the GRID computational resources. These grid nodes are used for the computational intensive tasks of ACGT: e.g. simulation of tumour growth or statistical analysis. Grid nodes are assigned automatically to a task based on availability at the time of execution. The nodes are controlled by the Globus Toolkit 4 middleware.

If a grid node is compromised, this can lead to the following situations:

- Jobs submitted on the grid node can be interfered with,
- Credentials of other users can be stolen if the job requires delegation,
- Data used for jobs can be leaked.

Grid nodes execute code provided by end-users and are thus particular vulnerable to local exploits in case that an intruder has obtained the necessary credentials (or “insider threat”).

The dynamic nature of remotely executed jobs makes job executors more susceptible to potential exploits. Within ACGT, care has been taken to ensure rigorous configuration of the job execution environments according to security best practices and principles to avoid susceptibility to unanticipated issues. E.g. when creating temporary files in the process of job execution, one must make sure that these files are not accessible by other parallel executing jobs (because they run under the same user account or because file privileges are incorrectly set). The employed Gridge middleware takes care of this.

5 Contextual Anonymity

5.1 Introduction

The DPF is based upon several pillars that support the safety net for protecting the involved patient's privacy (amongst other things). These include technical and organisational measures to avoid accidental or intentional misuse of personal data, the most important being the achieved de-facto anonymisation. However, the resulting anonymity is only conditional as it depends on assumptions based on external factors (i.e. not within the scope of the DPF). These assumptions have been taken into the equation in order to find the right balance between data subject privacy, usability and cost. The most important factors are the availability of external data sources and background knowledge of ACGT users.

5.2 Information Gain

There is a possibility that treating physicians or researchers gain information on their patients by matching patient data with anonymous ACGT data.

5.2.1 Enriched Data

One can assume that treating physicians who upload data into ACGT can always match their local records with the uploaded and de-identified data. This raises in itself no issues for privacy, as these physicians have lawful access to the nominative record, i.e. there is nothing to find out, except what they already rightfully know.

However, if additional data was linked within the pseudonymous ACGT context to those records, the physician would gain this information about his patient.

Note that there is a therapeutic relationship between the physician (gaining the information) and the patient. Therefore, the impact of the information gain is likely to be limited. Furthermore, the chance that this scenario occurs in practice is extremely small.

In case of serious indications of misuse, one could rely on auditing of data access requests in order to discover traces of such undesired information gain.

5.2.2 Unique Characteristics

Even in the case, where the physicians or researchers did not submit any data to ACGT themselves, they could deliberately search for or accidentally find data in ACGT that they can immediately associate with one of their patients. This is the consequence of the fact that the physicians or researchers dispose of a level of information that exceeds the anticipated level of information appointed to an attacker in the threat model that lies at the basis of the selected technical anonymisation measures.

5.2.3 Legal Consequences

From a legal point of view, there are two possibilities. The physician or researcher either intentionally matches data from outside the databases in order to gain information about patients or to see whether any of their patients is taking part in another trial. Or the physician or researcher unintentionally is able to identify his patient in the database.

The first scenario is prohibited by the contract on data protection and data security within ACGT. Thus, the physicians or researchers would be in breach of contract if they intentionally matched any data contained within ACGT with other data. Therefore, proper enforcement of the ACGT contracts will mitigate this risk.

The other scenario, that a physician or researcher is able to identify his patient through the data contained in the ACGT database, is covered by the informed consent. It is very unlikely, but there is this extremely small risk that the patient may be identified. Therefore, the consent forms provide a clause which highlights this risk.

In conclusion, an information gain of physicians and/or researchers does not pose a risk to the DPF.

6 Long-run Sustainability

6.1 Project End

When addressing long run sustainability, one of the major questions is what happens at the end of the project.

The first question that arises is who is responsible for taking the network apart? Generally speaking, this is the task of the management board in coordination with all the partners. The management board needs to know all rights and shares that have to be allocated with the different partners and needs to be able to prevent and solve any disputes among the partners related to this. Furthermore, the partners may need to conclude contracts that govern their relation after the project end, if they need to work on or work with jointly developed ideas or works. Thus, taking the network apart is a task of all project participants, coordinated by the management board.

6.1.1 Who is responsible in case the CDP is gone?

If the CDP ceased to exist, an important part of the data protection framework would be gone.

First of all, the question arises who is responsible for the data processing within the project. Both hospitals/investigator that transfer data and end-users like scientific researchers could not be considered to be data processors on behalf of the CDP anymore. Thus, they would have to be considered as controlling the data and therefore become responsible. This means that not the single person, but the institution they work for is responsible for the processing that is carried out during their work.

Similarly, Clause 10 of the contract on data protection and data security within ACGT states regarding the liability that if a data subject or patient is not able to bring an action arising out of a breach by the ACGT end-user of any of his obligations against the CDP because the CDP has disappeared factually or has ceased to exist in law or became bankrupted, the ACGT end-user agrees that the data subject or patient may issue a claim against the ACGT end-user as if he were the CDP.

Thus, the institutions that are data processors on behalf of the CDP as long as the CDP exists become data controllers as soon as the CDP ceases to exist.

The second question that arises is who will ensure compliance with the data protection framework? As the contracts were concluded between the CDP and the respective partners, there is no party that might be able to enforce the contracts. The only possibility left is that of the Consortium to exclude partners who violated the general terms, contracts, agreements or informed consents of ACGT or national and international legislation. Thus, compliance with the data protection framework could partly be enforced by the Consortium itself.

6.1.2 EU – Legislation

The project's legal settings were based on the existing EU-legislation. This especially relates to the data protection framework, but also for example to the analysis of the intellectual property rights.

Although it is less likely that EU-legislation changes substantially than it is likely that Member State law changes, the possibility remains. In that case, it has to be assessed whether the changes also apply to previously established settings. If this is the case, it has to be assessed whether the changes have any effect on the legal framework of ACGT. If they do, there might be the need to update the legal framework of ACGT to the new legal situation.

Currently, there is some movement on European level to review and amend the Data Protection Directive (Directive 95/46/EC). However, as there is no current official document that indicates what the aim of the review respectively any amendments would be.

Therefore, it is idle to discuss and describe possible changes and effects on the legal framework of ACGT before the legislative procedure started.

7 Conclusion

The ACGT Data Protection Framework (DPF) is a generic framework designed for dealing with data protection-related aspects in large, transnational research infrastructures such as ACGT. The framework combines technical, organisational and legal measures in a uniform multidisciplinary approach.

This document has discussed the main threats to the proper functioning of this framework:

- Non-compliance,
- Infrastructure and technology-related security risks,
- Risks associated to the concept of contextual anonymity,
- Risks related to sustainability,
- Risks introduced by long-term exploitation of the research infrastructure.

Most identified threats are associated with the leaking of de-identified data. In reality, these are likely to have a limited impact. In the end, it is highly unlikely that data will end up with people capable of re-identification (especially not with accidental breaches).

Due to the complexity and scale of the ACGT infrastructure, it is clear from the above that enforcing compliance to the DPF is not an easy task. In a long-time operational infrastructure, sufficient resources will need to be dedicated to technical and procedural audits. This again confirms[10] the need for a separate governing body such as the Center for Data Protection (CDP).

[10] Deliverable D10.2 "The ACGT ethical and legal requirements"

8 Acronyms

ACGT	Advancing Clinico-Genomic Trials on cancer
AC	Access Control
ACL	Access Control List
CA	Certificate Authority
CAT	Custodix Anonymisation Tool
CDP	Center for Data Protection
CRL	Certificate Revocation List
DOS	Denial of Service
DPA	Data Protection Authority
DPF	Data Protection Framework
EULA	End User License Agreement
GAS	Gridge Authorization Service
GDMS	Gridge Data Management System
GRMS	Grid Resource Management System
GSI	Grid Security Infrastructure
GT4	Globus Toolkit version 4
HSM	Hardware Security Module
IP	Intellectual Property
OCSP	Online Certificate Status Protocol
PDP	Policy Decision Point
PEP	Policy Enforcement Point
PKI	Public Key Infrastructure
RA	Registration Authority

VO	Virtual Organisation