Integrating Data Custodians in eHealth Grids – Security and Privacy Aspects

Jochen Fingberg[1], Marit Hansen[2], Markus Hansen[2], Henry Krasemann[2], Luigi Lo Iacono[1], Thomas Probst[2], Jessica Wright[3]

[1]C&C Research Laboratories, NEC Europe Ltd.
Rathausallee 10, 53757 Sankt Augustin, Germany
{fingberg, lo_iacono}@ccrl-nece.de

[2]Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein
Holstenstraße 98, 24103 Kiel, Germany
{LD10, LD63, LD9, LD91}@datenschutzzentrum.de

[3]School of Law, University of Sheffield
Conduit Road, Sheffield, United Kingdom
jessica.wright@sheffield.ac.uk

Abstract: This work introduces Grid computing, shows its use in eHealth environments and elicits trends towards the integration of custodians in eHealth Grids. It elaborates security and privacy requirements for the use of Grid computing in eHealth scenarios and discusses the possible integration of different types of data custodians. Finally the paper concludes and gives an outlook on the development and deployment of eHealth Grids in the near future.[1]

1 Introduction

This section describes Grid computing and presents an overview of Grid projects in the eHealth sector. In addition some scenarios for eHealth Grids are introduced. Needs of integrating data custodians are briefly elaborated, motivating the further discussion.

1.1 Grid Computing

Grids can combine aspects of clustering (multiple physical entities operating as one logical entity) and virtualisation (multiple logical entities operating on one physical entity). In a Grid, multiple entities (Grid Nodes, GNs) that are not centrally administered

 A short version of this text will be published in the Proceedings of “INFORMATIK 2006” (Workshop Electronic Data Custodianship: Applications, Methods, Foundations, October 2006, Dresden), Lecture Notes in Informatics (LNI), Springer, Germany, 2006.
URL of this document: http://www.ccrl-nece.de/publications/paper/public/LR-06-262.pdf.

interconnect, e.g., via Internet, and combine their resources to perform – among others – computational tasks [Fo02]. As Grids can serve multiple purposes and the GNs are independently administered, there is a need for highly flexible sharing relationships, ranging from client-server to peer-to-peer; for sophisticated and precise levels of control over how shared resources are used, including fine-grained and multi-stakeholder access control, delegation, and application of local and global policies; for sharing of varied resources, ranging from programs, files, and data to computers, sensors, and networks; and for diverse usage modes, ranging from single user to multi-user and from performance-sensitive to cost-sensitive and hence embracing issues of quality of service, scheduling, co-allocation, and accounting [FK+01].

1.2 eHealth Grids

Research using population-based data repositories gains increasing importance in academia, industry and governmental bodies. In life sciences, e.g., there is a compelling demand for the integration and exploitation of heterogeneous biomedical information for improved clinical practice, medical research, and personalised health-care. In this context Grid technologies are becoming a common infrastructure in order to federate different data sources to enable researchers as well as medical professionals to query and access distributed information in a unified and integrated way and to seamlessly provide computing resources [HGA04].

Various Grid-related research projects focus on the development, enhancement and implementation of Grid infrastructures in health-care. In 2002 the EC-funded FP5 projects BIOGRID (www.gridstart.org/BIOGRID.shtml), GEMSS (www.ccrl-nece.de/gemss/) [HC+04] and MammoGrid (www.mammogrid.com) started to introduce Grid technologies in eHealth. In autumn 2002 a first workshop was organised by the EC "ICT for Health Unit" on the new topic of the use of Grid technologies in the health domain. The term *HealthGrid* was coined. This term is very much linked to HealthGrid activities in eSciences, which is why in the following discussion the term *eHealth Grid* will be used to classify this kind of Grids. In January 2003, the first *HealthGrid Conference* was held in Lyon (France). On April 4th, 2003 the European HealthGrid Association (www.healthgrid.org) was established, and in 2004 (with support from Cisco Systems) it published a white paper on the concepts, benefits and opportunities of applying the emerging Grid technologies in a number of applications in health-care [HGA04]. In early 2006, six HealthGrid-related projects, resulting from FP6 call 4, were launched:

- @neurIST – Integrated biomedical informatics for the management of cerebral aneurisms, www.aneurist.org [AB+06]
- ACGT – Advanced Clinico Genomic Trials on Cancer, www.eu-acgt.org
- Health-e-Child – An integrated platform for European paediatrics based on a Grid-enabled network of leading clinical centres, www.health-e-child.org [FC+06]
- Immunogrid – The European Virtual Human Immune System Project, manage.zope.cineca.it/immunogrid/theimmunogridproject/

- Sealife – A Semantic Grid Browser for the Life Sciences Applied to the Study of Infectious Diseases Project Reference, www.biotec.tu-dresden.de/sealife/ [SB+06]
- SHARE – Supporting and structuring HealthGrid Activities & Research in Europe, www.eu-share.org

Beside the EC-funded projects, numerous national projects focus on this topic including CLEF (www.clef-user.com), which is a MRC[2]-sponsored project in the UK's e-Science programme, and MediGRID (www.medigrid.de) as part of the German D-Grid initiative (www.d-grid.de), which is funded by the Federal Ministry of Education and Research.

As powerful computing and data management capabilities provided by Grids continues to evolve, more and more computing and data sources will be combined in order to achieve a deeper knowledge about the human body and the treatment of diseases. @neurIST is, for example, a research project which will provide integrated biomedical informatics to aid the management of cerebral aneurysms. It aims to integrate heterogeneous data sources such as patient medical records, questionnaire results, images, genetic information with public genetic, public literature and private databases which spans all length scales, from molecular, through cellular to tissue, organ and patient representations. These data are increasingly heterogeneous in form, including textual, image and other symbolic structures, and are also diverse in context, from global guidelines based on the broadest epidemiological studies, through to knowledge gained from disease-specific scientific studies, to patient-specific data from electronic health records. @neurIST will develop various tools both to integrate and exploit these, and will provide a clinical decision support system which both analyses the risk of aneurysm rupture and provides treatment options. Grid computing will support the interface between the data sources and allow complex processing operations to be carried out, such as simulations.

1.3 Motivation for Data Custodians in eHealth Grids

The attempt of including as much knowledge (or data) into medical research environments and clinical decision-support systems as possible is targeted towards an enhanced and more efficient health-care. This raises, however, serious concerns regarding the protection of the data especially in the case of sensitive patient data.

Current practice is that the patient's personally identifiable information (PII) is provided for medical research after signing an informed consent form. This is bound to well-defined purposes and mostly to a rather short-term time-frame in accordance with the length of the research project. As long-term research collaborations are becoming increasingly important in some areas of medical research, the handling of informed consent forms might not be as feasible. This would also apply to short-term research projects where the knowledge base (including patient data) may form part of a yet larger database – such as the Virtual Physiological Human – in the future, and may therefore never be destroyed. The patient may not be able to take in all the information necessary

[2] Medical Research Council

to give proper consent to these future uses of his/her data, despite them being known at the time of collection. A larger discussion is needed on the impact of new Grid technologies and ICT-driven research on patient understanding. It is also questionable how it can be ensured that the patient data are used solely for the agreed purposes and that the patient does not lose the control over his/her data in the context of the Grid.

The extremely distributed nature of Grids seems to make the control of the PII of a patient particularly difficult. Furthermore, Grids incorporate an "amplifying" character meaning that the federated and integrated infrastructure also might facilitate unauthorised data collection and correlation, which might enable mining pseudonymised patient data and turning them into PII by accumulating identifiable information. These specific conditions have to be considered when designing or deploying pseudonymisation mechanisms.

To overcome these problems, trusted third parties in the form of electronic data custodians in charge of taking care of confided data might provide an efficient solution. To further analyse the role of and the security and privacy requirements for electronic data custodianship, different scenarios will be considered of which one is chosen for a more detailed elaboration in section 2.

1.4 Scenarios

Scenarios for eHealth Grids are manifold. They can be categorised according to their goals which include the improvement of clinical practice, medical research and personalised health-care. To enhance, for example, clinical practice and medical research, new technologies are used to incorporate imaging and simulations into diagnosis [HGA04]. Assume, for example, that including blood flow simulations based on scanned images in a patient's examination would aid the clinician to deduce the most appropriate and maybe cost-efficient treatment plan. Since these kinds of simulations are very complex and require a lot of computational resources, they are usually conducted by specialised service providers located outside of the attending hospital or physician. Thus, eHealth Grids serve – instead of an ordinary laboratory – the needs of the hospital or medical practitioners, and are commissioned by them.

- A *clinical treatment scenario* may start with a standard health check, the investigation of a particular disturbance, or an accident where some finding is diagnosed incidentally. The patient is referred to a specialist. The specialist retrieves the patient's data for further analysis by invoking – possibly located outside the specialist's domain – corresponding compute, analysis and simulation services. Furthermore his/her decision-support system will include other information sources so that the specialist is finally able to give the diagnosis and then suggests treatment options.
- A path through a *medical research scenario* may include the federation of various biomedical data sources such as gene sequences and epidemiological information in order to find correlations between the patient characteristics and the targeted research goal. As part of this, a link may be provided from hospital electronic health

records to researchers. This information could be integrated with other public and private databases, with a portal allowing access to researchers around the world. Modelling, simulation and data mining could take place to analyse the information.

- Industrial *scenarios including drug discovery or health equipment design* may access biomedical databases and medical research resources in order to improve their existing products or even develop totally new ones.

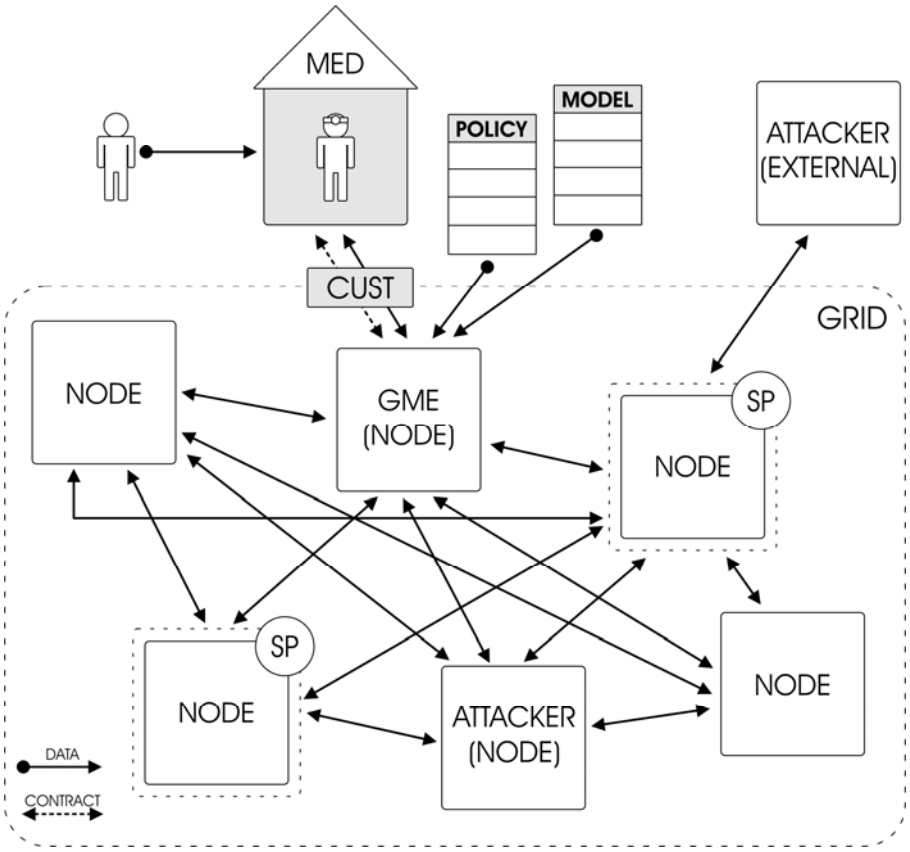

Figure 1: Main Roles in the Clinical Treatment Grid Scenario

Focusing on the clinical treatment scenario, the main roles, machines, and objects involved in health-care environments supported by eHealth Grids can be deduced (see Figure 1): The *patient* presents himself and thereby his/her PII to a *medical practitioner (MED)*, e.g., in a hospital. To analyse the data, the medical practitioner establishes a contract with a *Grid provider* who is operating a *Grid NODE (GN)* as a *Grid Management Entity (GME)*. The PII as well as a processing *model* (offered by the *model provider* who also may be the practitioner himself) and a *policy* get transmitted to the GME. The GME allocates resources within the Grid and transmits data and processing instructions to other GNs. The GNs might have a *security policy (SP)* of their own to

prevent them from providing resources for unsolicited tasks. The GME should be aware of such restrictions and should not allocate resources.

The GNs are interconnected and – according to the processing instructions they received – communicate with each other to solve the computational problem and transmit the results back to the GME which combines them and forwards them to the medical practitioner. Furthermore, *attackers* (external or within the Grid) may want to intercept communication, gain access to the data, or manipulate data, processing and results to spy out or even sabotage certain research. A possible task of a *custodian (CUST)* is to manage identifiers or pseudonyms, see e.g., [PR+05]: As an intermediary, it pseudonymises medical data before they are transmitted from the medical practitioner to the GME and the GNs and de-pseudonymises their reported results before delivering them to the MED.

2 Relevant Grid Properties with Respect to Privacy and Security

Although there seems to be no major differences between Grid technology and common commissioned data processing, the fact that various actors are involved in Grid computing (e.g, GMEs and GNs) leads to a different situation from a privacy and security point of view. This section describes these differences as well as relevant Grid properties and analyses privacy and security issues.

2.1 General Grid Properties

According to our setting, the principal (i.e., the medical practitioner) is not in charge of processing patient data through a Grid model himself. So the principal commissions a Grid to perform the task by making a contract with the GME. This contract typically includes privacy and security requirements to be fulfilled by the GME itself and the GNs. Involving a custodian will require appropriate contracts between principal, custodian and GME. Depending on the tasks of the custodian, some of privacy and security requirements might be burdened onto the custodian. An important factor is the degree of technical and/or organisational control of the GME over the GNs [cf. e.g., EGA05], or in other words the degree of autonomy of the GNs and their providers.

Even in the case of full digital control of the GME over the GNs there are differences with the scenario where the tasks are fulfilled directly in a laboratory associated to the hospital, as the GNs may be located in various places:

- The GME usually has no full physical control over the remote GN machines.
- The GNs may be located in multiple nations and therefore various legislative areas.
- The GNs require network access to exchange software and data which also opens ways for potential attacks.

The custodian can address some of these differences, e.g., by managing PII and keeping them away from the GME and the GNs. This would bypass problems of different

legislative areas. Also security issues such as software distribution and security configuration of GNs distrusting their GME might be handled by a custodian. However, other security aspects, e.g., the correctness and availability of computational results from GNs, cannot be guaranteed by the custodian.

2.2 Security Analysis

As Grid computation is characterised by the cooperation of multiple actors, computational resources and responsibilities, it is obvious that each entity has own, possibly conflicting, security interests. A trusted third party, such as a custodian, might help to overcome some of these conflicts. In order to understand the possible role of a custodian, it is necessary to analyse these conflicts. The following analysis characterises the security interest of each actor within the scenario shown in Figure 1. They include the classical protection goals such as confidentiality, integrity and availability, as well as accounting issues.

Some interests are commonly shared:

- As always all contractual parties are interested in the possibility to dispute and to impose liability on their contractors if interests are violated.
- Since most scenarios will be economically relevant, fair accounting for all actors is desired.
- Taking part in the Grid scenario means for most actors establishing an on-line connection to the Internet or another network not fully controlled by themselves. These actors are interested in protection from third party interference, e.g., hacking attacks.

Other interests are quite specific for the respective actor:

Actor	Interest	Contract with
Patient	The primary interest of the patient is good and affordable medical treatment. Of course confidentiality, integrity and availability are taken for granted.	MED
Medical practitioner (MED)	Primarily the MED is interested in surviving economically which is related to medical excellence and trust from the patient. When involving Grid technology, the MED is interested in correct computational results, delivered in time, including the possibility to check the quality of the results. Its interest in safeguarding the patient's data (in particular confidentiality and integrity) has to be further propagated to its other contractors.	Patient Model Provider GME Custodian
Model Provider	The model provider is interested in providing an accurate model (both algorithm and its implementation). The model running in the Grid must not be corrupt – in particular integrity and availability are primary interests. Intellectual property may also be an issue, meaning protection of confidentiality of the model.	MED
GME	As the GME is located in a central position in the data flow, communicating with MED, model provider, GNs and custodian, it is primarily interested in the perfect orchestration of the tasks and workflows. As mediator between GNs and MED, it has to rely on proper and in-time fulfilment of tasks by the GNs.	MED GNs
GN	Each GN is interested in protection from malware or dysfunctional software endangering integrity and availability of own computational resources and confidentiality of own data. In addition the GNs may be interested in undisturbed assignment of tasks by the GME, e.g., if they can account for their services on a task-by-task basis.	GME
Custodian	The custodian is interested in keeping its trustworthiness by safeguarding the PII according to the contract with the MED, at best being able to prove correct behaviour. Related is the interest in a well-working pseudonymisation method which does not reveal PII to unauthorised parties.	MED

Table 1: Overview of the interests of specific actors

For a comprehensive security analysis, four phases of the Grid scenario have to be considered:

1. Contract negotiation:
   All contracts have to be clearly stated and concluded by the parties. For technological support, the statements in the contract should be expressed and communicated via machine-interpretable policies. If a commonly agreed semantics is defined, it could be checked whether the GN's own security policies comply with what the GME demands, and whether this is in line with what the MED stipulates.

2. Data processing:
   According to what is regulated in the contracts (or policies) the Grid computation takes place: The MED determines the computational model and releases patient data to the custodian which forwards them without identifiable information to the GME (in some cases the custodian has to know details of the applied computational model in order to choose an appropriate way of eliminate PII, see section 4). The GME gets the model from the model provider and sends both model and pseudonymised data to the participating GNs. After the calculation the GME puts together the result and sends it to the custodian. The custodian knows the patient this result is belonging to, and communicates the information to the MED.

3. Accounting:
   All parties must monitor the services they provide during the previous phase in order to send invoices to the appropriate body.

4. Quality assurance:
   For the whole process it is important that errors are avoided. In some cases the MED may have to do a re-calculation of the Grid model – possibly with the same GNs as before or deliberately with other GNs not having been involved beforehand.
   Getting feedback to improve the applied methods is valuable both for the model provider, which may adapt the model according to the results, and for the custodian when enhancing the pseudonymisation method is possible or necessary (cf. section 4).

Several of the depicted security interests also arise in distributed non-Grid computing and can be tackled by well-known security mechanisms (e.g., authentication methods or cryptographic mechanisms of a VPN can be used to secure network traffic between the principal, the GME, and the GNs). Therefore, we will focus on those security issues that are specific to Grid computing and might lead to conflicts.

One of these issues is the security control of the GN components. From the GME point of view, the security status of a GN component should be the same as if the GN belongs to the GME. This includes correct operating system software and application software including patch level, control of network connections, absence of malware, access control etc. One possibility to implement such a status is a total (re-)configuration of the GN components by the GME (see [EGA05], Chapter 3.1.1.1). In this scenario, the GN temporarily loses control of its components. It requires a clear and clean change of control and configuration from the GN to the GME and back, e.g., "sanitation" of the

component after use, including the deletion of data and software and reconfiguration of the component. In [EGA05], Chapter 3.1.1.3, the proposed sanitation processes also include the deletion of forensic data on the component (after transferring such data to the GME to support its revision) and the re-configuration of BIOS and Flash-ROMs. From a technical point of view, it is not evident if this is possible via remote access of the GME, or fully possible at all (e.g., hidden areas of hard disk not accessible to operating systems).

From the GN point of view, the GME might be only a guest (among others) on the GN components. Therefore, the GN provider seeks maximum security from the applications and data coming from the GME. The GN provider's interest is not to transfer control of the component to the GME, but to keep control at all times. This especially holds when the owner of the GN and the GME use the component simultaneously, but also in a scenario when the GN temporarily surrenders the whole capacity of the component and parts of the control to the GME. In the latter case, the GN provider is interested in checking configuration changes by the GME and detecting any security-relevant remains of the GME (e.g., malware). Therefore, the GN provider is interested in log files and forensic data. Furthermore, the GN provider might not trust the re-configuration and sanitation of the GME and deploys its own configuration mechanism.

3 Ethical and Legal Aspects with Respect to eHealth Grids

Many ethical and legal aspects have to be considered concerning eHealth Grids. Confidentiality and privacy requirements are a major topic and form the focus of this section, but there are other issues, for example, regarding the Intellectual Property Rights of the models, algorithms and software which are being used for processing the data, or contractual Service Level Agreements.

The medical practitioner sending information to the Grid has a duty of confidentiality towards his/her patients. The World Medical Association International Code of Medical Ethics states that a physician shall "preserve absolute confidentiality on all he knows about his/her patient even after the patient has died" [WMA49]. This means it is the duty of the medical practitioner to ensure that processing patient data is done in a secure and trustworthy way. The patient him/herself also has a right to informational privacy. This can be enumerated in many different ways, but is often conceived as a right for the individual to control "to what extent information about them is communicated to others" [We70]. Both the duty of confidentiality and the right to privacy are often seen to stem from the concept of personal autonomy, which is "at a minimum, self-rule that is free from both controlling interference by others and from limitations, such as inadequate understanding, that prevent meaningful choice" [BC+01]. The concept of autonomy also gives rise to the requirement for informed consent as a moral rule. Proper informed consent would include providing the patient with information on the identity of the person processing the information and the purposes for which it was processed, and obtaining agreement to this.

The duty of confidentiality and the right to privacy can, however, be outweighed by other competing interests. These could include the patient's best interests and the public interest. It is often argued, for example, that medical research is in the public interest. It could also be argued that the use of an eHealth Grid to expand upon the clinical information available to the practitioner would certainly be in the patient's best interests, for it could aid diagnosis and better inform treatment decisions.

However, using these competing interests to justify possible disclosures and security concerns on the Grid should be seen as a last resort. The duty of confidentiality a doctor has to his/her patient is very strong, and can often only be overruled in situations of extreme public interest (for example, a mental patient discloses that he/she may kill someone). The right to privacy can be interpreted as more than just a concern that the information can identify a patient – for example, some commentators refute the idea that privacy interests stop at anonymisation or concealment of the patient's identity (although it can help protect it). The principle of personal autonomy is frequently highlighted as important in recent international statements on bioethics, for example the UNESCO Declaration on Bioethics and Human Rights states in Article 5 that "[t]he autonomy of persons to make decisions … is to be respected" [UN05]. Informed consent is therefore still seen as a *prima facie* rule when it comes to either medical treatment or research.

This means that, as far as possible, patients should be informed about and allowed control over the processing of their own information. Appropriate technical and security measures must be put in place to ensure safeguards to help protect privacy and respect confidentiality. These demands/needs might require some form of independent and trustworthy party taken care of the patient's data to ensure the proper protection and providing a single point of control. To better understand how the role of a data custodian in eHealth Grids may function, further details on related legal aspects are introduced, before the possible approaches are presented and discussed in section 4.

Personal data is governed in Europe by European Data Protection Directive 95/46/EC[3], which covers all individuals and organisations processing personal data (i.e. PII) and other (national) sector-specific regulations[4], unless the processing takes place for domestic or personal purposes. In Article 2, the following roles are defined:

- "data subject": natural person whose personal data are processed;
- "controller" shall mean the entity[5] which alone or jointly with others determines the purposes and means of the processing of personal data;
- "processor" shall mean an entity which processes personal data on behalf of the controller.

---

[3] All European countries have now transformed the Directive into national law.
[4] E.g., for health data being processed in online applications in Germany there are on the one hand canons of professional ethics with specific obligations for documentation and the doctor-patient confidentiality (§ 203 StGB (Penal Code), § 9 Ärztliche Berufsordnung), on the other hand requirements from telecommunications law (Telecommunication Act, TKG) and multimedia law (in particular Tele Services Data Protection Act, TDDSG).
[5] Entity: natural or legal person, public authority, agency or any other body.

In general personal data processing is lawful if the data subject gives an informed consent or if there is a specific legal basis (according to the laws of the Member State). In addition the principle of data minimisation applies: Only personal data which are necessary for the task shall be processed. If possible, data should be anonymised, especially because disclosing data in a digital world means that they leave the control sphere of the individual. Additionally the purpose-specification principle has to be considered. Personal data should only be collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes.

Processing health data, which are considered as "special category data" (or sensitive personal data) with a higher level of protection (Art. 8 95/46/EC), is lawful when it is required for the purposes of preventive medicine, medical diagnosis, the provision of care or treatment or the management of health-care services, and where those data are processed by a health professional subject under national law or rules established by national competent bodies to the obligation of professional secrecy or by another person also subject to an equivalent obligation of secrecy (Art. 8 par. 3 95/46/EC). However, this rule should not preclude consent and it is not certain whether medical research falls under this provision. An additional restriction could be the criminal law of the different countries. The professional discretion of people working in the health sector may restrict the possibility to transmit data of patients to third parties / processors like the GME. Not every country has special regulations about how this will be handled or gives special permission for this kind of transmission. In these cases it could be necessary to get an explicit consent by the patient.

In our setting the medical practitioner is the controller who is responsible for processing the personal data of the patient as data subject. The GME acts as processor, being governed by a contract from the practitioner which stipulates in particular that the processor shall act only on instructions from the controller (Art. 17 par. 3 95/46/EC).[6] Both the controller and the processor must implement appropriate technical and organisational measures to protect personal data against accidental or unlawful destruction or accidental loss, alteration, unauthorised disclosure or access, in particular where the processing involves the transmission of data over a network, and against all other unlawful forms of processing (Art. 17 par. 1 95/46/EC). The controller must, where processing is carried out on his behalf, choose a processor providing sufficient guarantees in respect of the technical security measures and organisational measures governing the processing to be carried out, and must ensure compliance with those measures (Art. 17 par. 2 95/46/EC). This includes the choice of the GME and processing nodes. For the purposes of keeping proof, the parts of the contract or the legal act relating to data protection and the requirements relating to the measures referred to in Art. 17 par. 1 shall be in writing or in another equivalent form (Art. 17 par. 4 95/46/EC).

Art. 23 95/46/EC settles the basic liability from the view of data protection: "[...] any person who has suffered damage as a result of an unlawful processing operation or of any act incompatible with the national provisions adopted pursuant to this Directive is

[6] The legal situation is different when the GME has own interests processing the data (e.g. own research). In this case an explicit consent of the data subject is the minimum requirement.

entitled to receive compensation from the controller for the damage suffered.", continuing in par. 2: "The controller may be exempted from this liability, in whole or in part, if he proves that he is not responsible for the event giving rise to the damage." Other liability obligations for failings, failures and misuse depend on the contracts between client (patient), controller and processor. Generally everyone is responsible for his/her area of control.

The data subject has several rights, e.g., to get information when data is collected (Art. 10-11 95/46/EC), the right of access (Art. 12 95/46/EC) including the right to rectification, erasure or blocking, and the right to object (Art. 15 95/46/EC). To exercise these rights the controller needs to be addressed, who will need to forward the request to the processor, e.g., when access to personal data is required. This may also affect the GNs which have to give access as far as the data processing is organised in a way that personal data are concerned.

Directive 95/46/EC guarantees a consistent level of protection inside the EU. The transfer of personal data which are undergoing processing or are intended for processing after transfer to a third country is problematic (Art. 25 95/46/EC). Generally this is only allowed if the third country in question ensures an adequate level of protection. But this is officially acclaimed only for some countries like Switzerland or Canada. For transferring personal data to other countries, the controller has to adduce adequate safeguards with respect to the protection of the privacy and fundamental rights and freedoms of individuals and as regards the exercise of the corresponding rights (Art. 26 par. 2 95/46/EC). Basically these safeguards can result from appropriate contractual clauses between sender and receiver of the data. It is necessary that this contract deploys irrevocable rights for the data subject. A contract like this is only possible if the receiver in the third country really has the potentiality to assure the adequate level of protection. If the (legal) circumstances in the third country are opposed to these safeguards (like special rights of access of the officials) regarding the concerned kind of data, the transfer to this country is not possible. This also applies if the controller himself is only operating and administrating the server in the third country but authorities have the right for, e.g., detention of the computer.

In the case of transferring personal data to GNs outside the EU, all these legal points must be considered. In most cases the deployed safeguards have to be assured – including a contract between the GME and the GN providers.

In relation to medical purposes and medical research, there are some important exemptions in Directive 95/46/EC not already mentioned above. Firstly, when data are collected for a specific purpose, it cannot be further processed in a way incompatible with these (Art. 6 par. 1 b). However, processing for "scientific purposes" is not considered as incompatible as long as appropriate safeguards are provided. Data should not be kept in an identifiable form for longer than necessary for the purposes obtained for (Art. 6 par. 1 e). Data collected for scientific purposes can be, subject to appropriate safeguards. When data is not collected directly from a data subject (Art. 11) and will be used for scientific research, information does not need to be provided where it would be a disproportionate effort to do so, if safeguards are provided. Implementing countries

may have also restricted the right of access when data are processed solely for scientific research, subject to certain conditions.

However, even in the field of scientific research a custodian supporting the patient's privacy rights may not only be stipulated because of ethical reasons, but also may implement the demanded "appropriate safeguards" and increase acceptance of Grid technology.

4 Approaches for Solutions

Especially in cases of PII to be distributed and processed, the GME must be able to safeguard that the data cannot be accessed by unauthorised third parties, e.g., operators of other GNs. As technical considerations will face worse problems than those the media industry tries to conquer, using Digital Rights Management or similar methods to protect PII will be required [HM05]. This implies that the GME has to define a policy the GNs have to comply with and that the Grid environment, i.e. the set of protocols and tools that allow for interoperation of the nodes, has to have a policy enforcement mechanism. Also, the GNs should have a policy of "acceptable" tasks, e.g., "military research tasks will always be rejected while medical research tasks can be accepted if they do not have to do with birth control and if idle resources are available". Furthermore, the Grid environment should be aware of policies the nodes have to comply with, such as different legal implications in different countries. For example, PII may only be exported from the European Economic Area under certain circumstances (cf. section 3).

Taking into account privacy and security principles, data processing in eHealth Grids has to be supported by a variety of measures. The currently often used solution of getting the individuals' consent for processing their PII is questionable because this would require that everybody really understands the risks. Instead, data minimisation techniques should be applied which may rely on a third party as a custodian.

We understand "custodian" as an independent and trustworthy third party taking care of provided data (possibly including software and configuration data), processing them in an agreed-upon manner, ensuring that provided data are used only for the agreed-upon purpose in the agreed-upon time period, are not forwarded to unauthorised parties, and are protected from external and internal attacks.

A primary task for a custodian could be to (reversibly) detach PII from the data for the duration of the processing. There are several possibilities, depending on the structure of the medical data and the computational task.

1. *Pseudonymisation*, i.e. exchanging names and other identifiers through the use of pseudonyms, and back, when transmitting data between practitioner and GME, to enable linkage between data relating to the same pseudonyms and to make re-identification possible e.g., for communicating the data processing results to the user. Pseudonymisation includes the administration of the relationship between identifying data and pseudonyms. If necessary, multiple pseudonyms can be used.

This task might include not only the modification of meta data such as file names containing PII, but also modification of medical data that are considered as "originals", e.g., change of patient names in an X-ray picture. Depending on the data structure, this can be a difficult task, e.g., if the patient's name is included in the picture as a watermark.

2. *Segmenting* the computational tasks and processes and dispatching them to the GME or GNs in a way that the tasks reveal no PII, (e.g., dispatching an image in small parts to different GNs/GME ([HGA04], Chapter 8.11)). It depends on the computational model if segmentation into pieces can prevent identification.

3. *Pre- and post-processing* of computational tasks in a way that the remaining data cannot reveal PII. This is similar to the previous option, could also include "encryption" / "decryption" processes by manipulating the computational tasks and data in a way that they can be performed by the GNs or GME (e.g., a random change of scale), but neither input data nor results reveal PII.

Note that while reliably removing the link from PII to the related patient is relatively easy with most alphanumeric data, it is impossible with, e.g., biometric data such as a tomographic scan of a head, where – while the single "slice" does not necessarily allow one to recognise the person – the whole set of "slices" allows for computation of a 3D model making the person identifiable. Just removing the name from certain data does not make them anonymous, i.e., non-identifiable.

Another task for a custodian could be to offer a trustworthy archive for the huge data amounts which may occur in Grid computing, e.g., a central repository storing medical data from different hospitals as a third party (outsourcing). This requires multi-client capability of the registrar in order to keep files separate between different clients.

All these tasks (pseudonymisation, segmentation, pre-processing, storage etc.) may be executed by the hospital itself or on contractual bases by a data processor on behalf of the hospital. But there may be several advantages of using a custodian as defined in the beginning of this section:

- As a security and privacy expert, a custodian might have a deeper knowledge of the specific legal and security requirements and do a better service.
- A custodian can help to overcome internal conflicts in a hospital, including conflicts of interests of different departments (e.g., a demand of the finance department to get the actual address of a patient from research files).
- A custodian serving multiple hospitals can simplify and thereby cheapen the transmission of pseudonymised (or anonymised) data for research purposes in-between research facilities, as the data formats of the pseudonymisation / anonymisation are compatible. This includes also the uses of multiple GME infrastructures operating on the same data, e.g., for benchmarking computational complexity and accuracy of different algorithms.

From the legal perspective the custodian must not have own interests in the patients' data, but has to demonstrate its independency and reliability. Of course appropriate contracts which regulate the obligations of all parties involved must be set up before data processing starts. The custodian will also have to adhere to data protection law, including the support of the patient's right to access.

5 Conclusions and Outlook

Grid computing is becoming steadily more important for intensive computing tasks as well as distributed and federated data access. Its usage is not limited only to the eScience domain anymore, and Grid computing is well recognised as powerful outsourcing technology in enterprises. Its capabilities are additionally relevant for the eHealth sector, since modern medical treatment and research is demanding for high-capacity computing, e.g., for image processing. Hence eHealth Grids will become increasingly visible in the following years.

Security and privacy requirements have to be considered when designing the workflow dealing with patient data. Preventing identifiability of patient data is not trivial. This is especially true for Grids, since due to their highly distributed nature, the control of the use of patient data becomes very complex if not impossible. Custodians can help to implement concepts for increased control and trustworthiness by keeping track of the patient's data and pseudonymising them in different ways, partially tailored according to the model to be used, and thereby separating the Grid context (calculation or data access) from the clinical context (treatment or research).

Many parts of security functionality are meanwhile addressed by Grid designers (e.g. [FK+01]), but still aspects of multilateral security and legal requirements for privacy are not fully solved. Privacy policies which formulate requirements depending on location and national legislation of the Grid Nodes should be supported. Data Protection Commissioners should be integrated early in Grid projects to give feedback in design phases.

References

[AB+06] Arbona, A.; Benkner, S.; Fingberg, J.; Frangi, A. F.; Hofmann, M.; Hose, D. R.; Lonsdale, G.; Ruefenacht, D.; Viceconti, M.: Outlook for Grid Service technologies within the @neurIST eHealth environment. In Proceedings of HealthGrid 2006, Valencia, Spain, 2006.

[BC+01] Beauchamp, T. L; Childress, J. F.: Principles of Biomedical Ethics (5th ed.), Oxford: Oxford University Press, 2001, p. 58.

[EGA05] Enterprise Grid Alliance Security Working Group: Enterprise Grid Security Requirements, Version 1.0, 2005. Available online at: http://www.gridalliance.org/en/workgroups/GridSecurity.asp.

[FC+06] Freund, J.; Comaniciu, D.; Ioannis, Y.; Liu, P.; McClatchey, R.; Morley-Fletcher, E.; Pennec, X.; Pongiglione, G.: Health-e-Child: An Integrated Biomedical Platform for

Grid-Based Paediatric Applications. In Proceedings of HealthGrid 2006, Valencia, Spain, 2006.

[FK+01] Foster, I.; Kesselman, C.; Tuecke, S.: The Anatomy of the Grid: Enabling Scalable Virtual Organizations. In International J. Supercomputer Applications, 15(3), 2001. Available online at: http://www.globus.org/alliance/publications/papers/anatomy.pdf.

[Fo02] Foster, I.: What is the Grid? A Three Point Checklist. In GRIDToday, July 20, 2002. Available online at: http://www-fp.mcs.anl.gov/~foster/Articles/WhatIsTheGrid.pdf.

[HC+04] Herveg, J. A. M.; Crazzolara, F.; Middleton, S. E.; Marvin, D.; Poullet, Y.: GEMSS: Privacy and Security for a Medical Grid. In Proceedings of HealthGrid 2004, Clermont-Ferrand, France, 2004. Available online at: http://www.ccrl-nece.de/gemss/Reports/Herveg-healthgrid2004.pdf.

[HGA04] HealthGrid Association; Cisco Systems: Healthgrid White Paper; 2004. Available online at: http://whitepaper.healthgrid.org/.

[HM05] Hansen, M.; Möller, J.: Digital Rights Management zwischen Sicherheit und informationeller Selbstbestimmung. In: Bundesamt für Sicherheit in der Informationstechnik (BSI) (Hrsg.): IT-Sicherheit geht alle an!, Tagungsband zum 9. Deutschen IT-Sicherheitskongress des BSI, Gau-Algesheim 2005, pp. 159-171. Available online at: http://www.datenschutzzentrum.de/vortraege/050510_hansen-moeller_bsi.htm.

[PR+05] Pommerening, M.; Reng, M.; Debold, P.; Semler, S.: Pseudonymization in medical research – the generic data protection concept of the TMF. In GMS Medizinische Informatik, Biometrie und Epidemiologie 2005; 1(3): Doc17. Available online at: http://www.egms.de/en/journals/mibe/2005-1/mibe000017.shtml.

[SB+06] Schroeder, M.; Burger, A.; Kostkova, P.; Stevens, R.; Habermann, B.; Dieng-Kuntz, R.: Sealife: A Semantic Grid Browser for the Life Sciences Applied to the Study of Infectious Diseases. In Proceedings of HealthGrid 2006, Valencia, Spain, 2006.

[UN05] UNESCO: Declaration on Bioethics and Human Rights, 2005. Available online at: http://portal.unesco.org/shs/en/file_download.php/46133e1f4691e4c6e57566763d474a4dBioethicsDeclaration_EN.pdf.

[We70] Westin, A.: Privacy and Freedom, London: Bodley Head, 1970, p. 7.

[WMA49] World Medical Association: International Code of Medical Ethics, 1949. Available online at: http://www.wma.net/e/policy/c8.htm.