D1.1. State of the Art Analysis

Medolution

Medical Care Evolution

ITEA3 – Project 14003

Document Properties

Edited by:	David Kuik, Lora Kushner, Norima Consulting Inc.
Authors	Medolution Partners
Date	11/24/2016
Visiblity	Public
Status	PMT review

History of Changes

Release	Date	Author, Organization	Changes
V.0.1	04/01/2016	David Kuik, Jacek Hunek, Lora Kushner (Norima)	An initial outline
V.0.2.	02/02/2016	Contributors: Gokce Banu Laleci Erturkmen, Anil Sinaci, (SRDC); Frerk Mueller (OFFIS); Celine Badr, (Prologue); Anna Litvina (Materna); Henning Brümmer (TUDO); Stéphane ZENG (Bull)	Input on the outline
V.0.3	04/02/2016	Contributors: Jacek Hunek, David Kuik, Lora Kushner, (Norima)	An updated outline
V.0.4	17/02/2016	Nils Reiss (SSK); Anil Sinaci (SRDC); Celine Badr (Prologue); Anna Litvina (Materna); François Exertier (Bull)	Further input on the outline
V.0.5	25/02/2016	Jacek Hunek, Lora Kushner (Norima)	An updated outline
V.0.6	25/04/2016, 11/05,2016; 14/06/2016; 19/08/2016	Gokce Banu Laleci Erturkmen, Anil Sinaci (SRDC)	Contribution to Chapter 8.2, 7.1, 9.3, Appendix A 1.1.1, Appendix B 2,3.
V.0.7	02/05/2016; 19/07/2016; 12/08/2016	François Exertier, Stéphane Zeng, Claudine Chouet (Bull)	Contribution to Chapter 3, 9.1, Appendix B 4.
V.0.8	03/05/2016; 17/07/2016	Mihai Mitrea (IMT)	Contribution to Chapter 8.4, Appendix B 1.
V. 0.9	28/04/2016; 21/06/2016; 08/08/2016	Anna Livina, Ingo Lieck, Lukas Int-Veen (Materna)	Contribution to Chapter 6, 9.4, Appendix B 8,9, 13-16
V.1.0	13/06/2016	Jacek Hunek, Lora Kushner (Norima)	Consolidated draft
V.1.1	08/06/2016 19/06/2016; 25/07/2016; 22/08/2016	Wolgang Thronicke (ATOS)	Contribution to Sections Chapter 3, Appendix B 4.
V.1.2	13/05/2016; 28/06/2016; 05/08/2016	Frerk Müller-von Aschwege, Jenny Roebesaat (OFFIS)	Contribution to Chapter 5, 9.3, Appendix A 1.4.1, B 5.
V.1.3	17/06/2016;	David Kuik, Jacek Hunek, Meenakshi Gupta, Lora Kushner (Norima)	Contribution to Chapter 8.3, 9.6, 10

Release	Date	Author, Organization	Changes
	14/07/2016; 12/08/2016		Appendix A 1.1,1.2, 1.3, B 2.
V.1.4	09/06/2016; 13/06/2016; 26/08/2016	Mauvigner Pierre, Bechir Taleb Ali (Prologue)	Contribution to Chapter 5, 9, Appendix A 1.4.2, B 6, 7.
V.1.5	27/07/2016	Henning Brümmer, Egor Kudrjaschow (TUDO)	Contribution to Chapter 4, 9.2, Appendix B 10.
V.1.6	11/08/2016; 19/08/2016	Robert Ponsen, Hubrecht de Bliek (Philips)	Contribution to Chapter 2, 7.2, 9.5, 10, Appendix A 1.3
V.1.7	19/08/2016	Jacek Hunek, Lora Kushner (Norima)	Updated version, References added, Overall Review
V1.8	26/08/2016	Henk van den Brink (Technolution)	Contribution to Chapter 3
V.1.9	30/08/2016	Güven Fidan, Selim Nar, Emine Ferraro	Contribution to Chapter 7.4, 9.5, Appendix B 11,12
V.2.0	05/09/2016	Jacek Hunek, Lora Kushner (Norima)	Updated version, Executive Summary, Overall Review
V.2.1	08/09/2016	Gokce Banu Laleci Erturkmen (SRDC)	Review
V.2.2	14/09/2016	Henning Brümmer (TUDO); Anna Livina (Materna); Selim Nar (Argedor); Celine Badr (Prologue); Frerk Müller-von Aschwege (OFFIS)	Minor updates
V.2.3	21/09/2016	Jacek Hunek, Lora Kushner (Norima)	Updated version
V.2.4	27/09/2016	Silvia Delgado Olabarriaga (AMC)	Review
V. 2.5	30/09/2016	Jacek Hunek, Lora Kushner (Norima)	Updated version
V. 2.6	16/10/2016	Elsemiek ten Pas (Sopheon), Henk Marquering (AMC), Abdelghani Chibani (UPEC/Maidis)	Contribution to Chapter 7, 9.5
V. 2.7	02/11/2016	Henning Brümmer (TUDO); Anna Litvina (Materna); Selim Nar (Argedor); Celine Badr (Prologue); Frerk Müller-von	Minor updates

Release	Date	Author, Organization	Changes
		Aschwege (OFFIS); François Exertier, Stéphane Zeng (Bull); Robert Ponsen, Hubrecht de Bliek (Philips)	
V.2.8	10/11/2016	Jacek Hunek, Lora Kushner (Norima)	Updated version
V.2.9	14/11/2016	Mihai Mitrea (IMT)	Review
V.3.0	17/11/2016	David Kuik, Lora Kushner (Norima)	Updated version
V.3.1	21/11/2016	Gokce Banu Laleci Erturkmen (SRDC)	Review
V.3.2	24/11/2016	David Kuik, Lora Kushner (Norima)	Final draft to sign off by PMT members
Final	28/11/2016	Frank van der Linden accepted	Move to final

List of Figures

List of Tables

Table of Contents

Executive summary

This document describes the State of Art analysis of the technologies, trends and approaches relevant for the Medolution project, as they may deliver conceptions and solution elements input to the development of the Medolution architecture, model, platform and demonstrator applications. The technology status presented in the document will be considered as the project's baseline that the project aims to advance. The document is intended to be read by a general informed research, development and integration engineer, be it from industry or academia.

Since the project schedule plans two iterations of the state of the art analysis, it appears in two versions. The Deliverable 1.1 is the first version.

After an introduction to this document, an overview over the concept of Big Dependable System, its key characteristics and key architectural elements is provided as an entry point to the relevant Medolution topics and their interconnection. A Big Dependable Healthcare System (BDHS) can be described as a system that supports the provisioning of health services towards the general population as well as patients. Such a system must be capable of supporting reliable decision making on a wide range of automatic input parameters for a multitude of simultaneous users, while providing safety and security for the users and their medical information. Consequently, the state of art on a number of related topics has been examined in the following Chapters.

In particular, relevant to Medolution technologies in the areas of Internet of Things and Big Data systems, dependability, automated control management and device-based systems are discussed to identify the state of the art architecture, engineering, modelling and description techniques and to initially assess their relevance to Medolution project. Additionally, health data exploitation approaches are discussed, including such aspects as data integration, analysis and decision support. Data privacy and security standards and solutions relevant for Medolution are surveyed as well.

Conclusions provide an integrated vision on the way how technologies discussed topics in the main Chapters are brought together under Medolution based on the examined state of the art and enabling its foreseen advancement.

Two topics relevant for Medolution are presented in the Appendixes. Thus, relevant regulatory data privacy and security constraints are reviewed, which constitute an important aspect of market environment for application and advancement the relevant technical state of the art, and therefore present essential requirements to be considered by Medolution. Besides that, since Internet of Things (IoT) and Big Data application is an area of active research, results of relevant to Medolution European and national projects have also been surveyed.

1. Introduction

1.1. Aim of the activity

The aim of this deliverable is to depict the existing state of art on the technologies, trends and approaches relevant for the Medolution project, as they may deliver conceptions and solution elements input to the development of the Medolution architecture, model, platform and demonstrator applications.

Medolution project targets the current need to reduce the cost of healthcare while improving the quality of life of patients. The project aims to create smart environments that integrate professional and user-created data that would enable the relevant information to be used for the support of patients and healthcare professionals in their decision-making on diagnosis, treatment and further monitoring. The main technical challenges to address this goal are: 1) to deal with the enormous amounts of heterogeneous data and data sources (e.g. TB of data), 2) to integrate and combine heterogeneous (in terms of media modality, creation source and time processing constraints) data, and 3) to extract relevant information from them, while ensuring safety and reliability of the devices in the patient's environment that produces and consumes this data as well as ensuring security and privacy.

Medolution project intends to address the challenges by realising big healthcare data processing and analysis in the cloud leading to:

- Early and pro-active decision support for patients and healthcare professionals in the form of timely meaningful alerts and notifications
- The ability to generate healthcare predictions based on continuous trend analysis
- The ability to share healthcare data between devices and persons

These project objectives result in a wide range of topics, from abstract system properties to supporting implementation approaches, as well as engineering, modelling and description techniques to different technologies and conceptions of the Big Data Platforms, IoT and Data exploitation domains. In particular, the related issues of Big data management, cloud management, security, interoperability, analytics, decision support and system integration in Healthcare need to be examined.

The Medolution project builds upon the results of the Medusa project that provides collaborative cloud access to medical information relevant in critical situations, where security, latency and collaboration have been addressed. Medolution adds and focuses on mobile and long-term monitoring and decision support. Selected aspects of Medusa topics that still need innovation will be addressed in this context.

With this background in mind, Task 1.1 of the Working Package 1 of the Project is devoted to the corresponding state of the art analysis in all related to Medolution technical areas. This document, Deliverable 1.1, reports the results. Since the project schedule plans two iterations of the state of the art analysis, the Deliverable 1.1, is the first version conducted at the beginning of the project. The second iteration, Deliverable 1.5, is due at the end of the project.

In this first version, the selection and structuring of the topics, conducted analysis and depths of contributions cannot profit yet from rich project experience and is mainly based on the existing expertise of the project partners in their domains.

At the end of the project, in the second version, the state-of-the-art analysis will be re-visited and a closer look will be taken on the approaches actually been applied. It will also depict the innovation achieved during the Medolution project lifespan on top of the existing work in the field.

1.2. Overview of the Deliverable

The document is structured as follows:

- Chapter 1 provides an overview and a summary of the document.
- Chapter 2 introduces an overview of a Big Dependable Healthcare System (BDHS) in terms of its characteristics, attributes and key elements of its architecture.
- Chapter 3 examines technologies, modelling- and architecture- approaches for Big Data Systems in Healthcare domain.
- Chapter 4 discusses systems architecture frameworks and modelling languages for supporting dependability.
- Chapter 5 analyses Device-based and IoT systems, independent device integration and management systems in Healthcare.
- Chapter 6 provides an overview of automated technical management models for medical systems.
- Chapter 7 investigates Healthcare data exploitation, in particular techniques and approaches for integration, analysis of healthcare data, representation of the analysis results as well as technologies and methodologies for decision support systems.
- Chapter 8 examines security solutions for IoT and Big Data systems, in particular general cloud security, data security standards, as well as anonymization and user centric data privacy techniques.
- Chapter 9 presents conclusions by highlighting the complementarily between the main technologies, supporting implementation approaches and techniques in their current state of art for Medolution discussed in previous sections and planned advancements to be realized in Medolution innovations.
- Appendix A gives an overview of various regulatory constraints for healthcare systems to be considered, including international and national data privacy and data security regulations to be considered for Medolution. Additionally, aspects relevant to the project concerning resource ownership, SLA, as well as medical devices certification are discussed.
- Appendix B is devoted to the European (IST and ITEA) and national research projects relevant for Medolution.

2. Medolution – a Big Dependable Healthcare System

The landscape of care is changing rapidly: the population as a whole is ageing and at the same time people (patients) live longer, often with one or more chronical diseases, causing structural increased cost of healthcare. In combination with these trends, the virtualisation of care develops at high pace and intensity, which results in large amounts of heterogeneous, clinically relevant information, in great technical variety, from many sources, for health professionals and patients. All this information needs to be handled and managed promptly. These are the fundamental challenges for care providers, public authorities and also for patients, which Medolution addresses in an integrated way.

In general, a Healthcare System is a system that provides both professional caregivers as well as patients an entry point towards the history, actual status and next steps regarding the monitoring and/or treatment of a patient towards a single condition. Having this system as a dependable system adds the availability and reliability to such a system, providing both the professional caregivers as well as the patients the trust in this system. The addition of “Big”, resulting in a Big Dependable Healthcare System (BDHS) implies the system is not only targeting a single combination of professional caregivers and the respectful patients to the condition case, but targets many simultaneously professional caregivers and patients, in combination of many different conditions, treatments and monitoring needs.

Medolution will deliver this “Big Dependable Healthcare System” which brings the relevant medical information to health professionals and patients at the right time at the right place, in the most effective, and intelligent way, constructed from all these different sources. In contrast to the limitations of the existing point-to-point capabilities and applications, Medolution allows scaling to millions of patients in parallel, supporting information flows from a multitude of sensor devices to many specialised medical applications. Medolution will deliver the methods and systems to connect these medical applications, addressing many varying diseases in parallel to serve a large number of patients and clinicians.

Medolution must be a trustworthy, dependable infrastructural component, dedicated to the professional care and hospital sector. To realise this, Medolution will create so called ‘’Smart Patient Environments”. As such the scope of Medolution will be to provide patient and hospital “data” streams that span the entire ‘’continuum of care’’. See Figure 1 below.

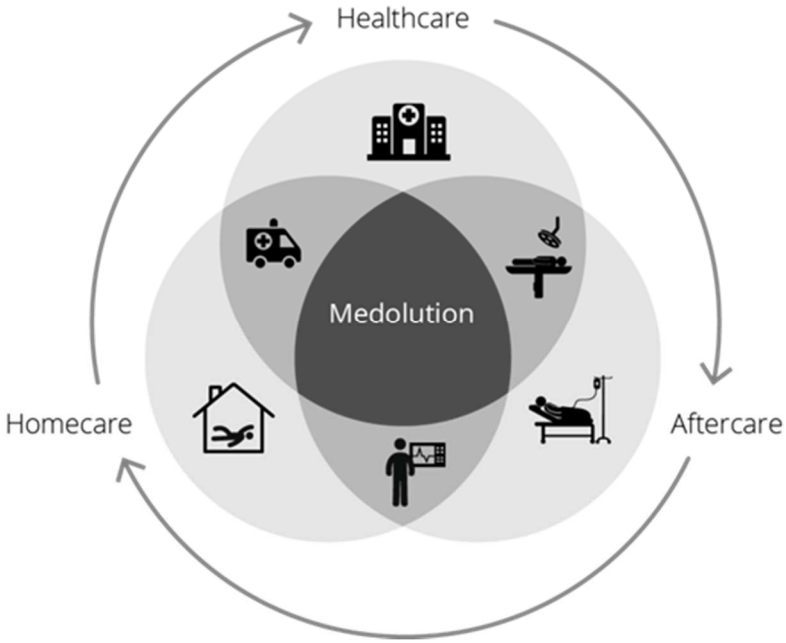

Figure 1 Continuum of care

One of the main Medolution objectives is to provide data transfer from various types of sources from medical sensor devices to hospital information systems, while integrating these sources at a "Big Dependable System" which will be stable, continuously accessible and will provide real-time services. These key characteristics and elements determine the key technological areas of development and innovation (including design standards) for Medolution.
The identified technological areas of development are:

- IoT and Big Data Solutions for Healthcare
- Dependability
- Devices and IoT Solutions for Healthcare
- Automated Technical Management for Healthcare Systems
- Healthcare data exploitation
- Privacy and security solutions for IoT and Big Data systems.

The picture in Figure 2 below displays a schematically overview of all the functional components that combined will allow for the functioning of the Medolution system as a Big Dependable Healthcare System.

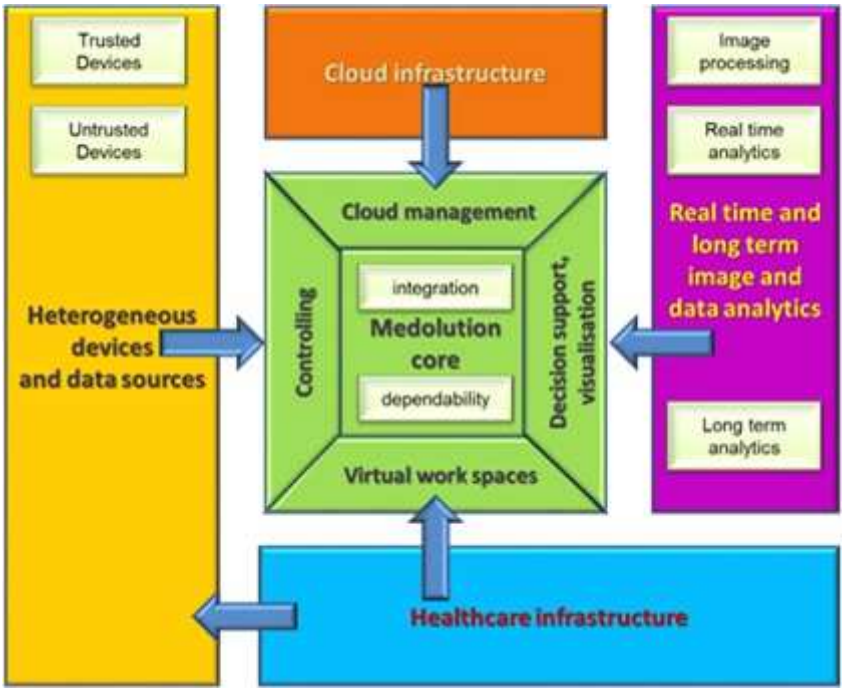

Figure 2 BDHS from the Medolution FPP

Consequently, the state of art in these topics is addressed in the upcoming sections of the document. In addition, *Relevant international and national data privacy constraints* and *Relevant European research projects in healthcare data processing* can be found in the Appendixes.

3. IoT and Big Data Solutions for Healthcare

Medolution will require Big Data capabilities to deal with the huge volume of health data made available through the use cases, and to make most of the analysis results available in real-time. Part of the data will come from connected devices; therefore, both Internet of Things (IoT) and Big Data architecture are to be considered. This Chapter provides an architectural overview of IoT and Big Data Systems. It will then present a panel of Big Data tools and platforms, indicating the difficulties related with their use for building applications. Limitations of these solutions will be highlighted as well as the way how they will be addressed in Medolution, in particular in the Medolution platform under WP4.

3.1. Standards in IoT and Big Data solutions for Healthcare

With regard to Big Data, a lot of efforts exist to create standards in this area. Since Big Data is a domain-neutral development most of the standards are generic to be applicable for all Big Data developments. The following standards shown in Table 1 have been selected because they facilitate creation of interfaces between components of the Big Data reference architecture (see [1]) and they are an abbreviated version of the information from [2]:

Table 1 Design Standards for IoT and Big Data Systems in Healthcare

Standard	Description
ISO/IEC 9075-*	This standard defines SQL. It contains the definition of data structure and operations on stored data.
ISO/IEC TR 9789	Guidelines for the Organization and Representation of Data Elements for Data Interchange
ISO/IEC 11179-*	Multipart Standard for the definition and implementation of metadata registries.
ISO/IE TR 19075-*	Technical Reports on SQL related technologies (Xquery, Java Bindings, etc.)
ISO/IEC 19763-*	Information Technology - Metamodel Framework for Interoperability (MFI)
ISO/IEC 10918:1994	Information Technology - Digital Compression and Coding of Continuous
ISO 6709:2008	Standard Representation of Geographic Point Location by Coordinates
ISO 19157	Geographic Information Data Quality
ISO/IEC 15408-2009	Information Technology - Security Techniques - Evaluation Criteria for IT Security
ISO/IEC 27033-1:2009	Network Security
ISO/IEC 29100:2011	Privacy Framework
ISO/IEC 27004	ISO/IEC 27004, Information security management - Measurement
W3C Platform for Privacy Preferences (P3P) 1.0	Standard format to express privacy practices of Web sites. Allows easy interpretation by software tools.
W3C Rule Interchange Format (RIF)	Standards for exchanging rules among rule systems.
OGC® Sensor Web Enablement (SWE)	This series of standards support interoperability interfaces and metadata encodings that enable real-time integration of heterogeneous sensor webs.

HL7 FHIR (Fast Healthcare Interoperability Resources)	FHIR leverages existing logical and theoretical models to provide a consistent, easy to implement, and rigorous mechanism for exchanging data between healthcare applications.

These examples show that there is already an abundance of standards relating or influencing Big Data and IoT even in healthcare, since the main difference here are the legal restrictions and constraints which further enforce such standards, or restrict specific operations between stakeholders. The usage of all these standards creates the framework for defining the system architecture as described in the following Section.

3.2. IoT and Big Data platforms

3.2.1. Systems architecture for IoT and Big Data Systems

IoT and Big Data Systems are different but still very compatible domains. When these domains are mixed together, one is considered as a data producer for the other which is then a massive data management system.

Building a system that integrates Big Data Analytics and IoT is complex work that requires rigorous architecture framework in order to build a system that fits the various requirements and constraints by relying on existing standard architecture of each domain, while taking to account many non-functional constraints such as heterogeneity of actors and components.

IoT - An architecture framework
There are several architecture frameworks for IoT domain. However, the utilization of the IoT-A Converged Reference Model [3] is often considered in the ITEA projects. This Converged Reference Model first defines a Reference Model for the IoT domain in order to promote a common understanding, then proposes a Reference Architecture that describes essential building blocks as well as design choices to deal with conflicting requirements regarding functionality, performance, deployment and security. The main aim of IoT-A can be explained using the pictorial representation as shown in Figure 3 below.

Figure 3 The IoT-A Tree [3]

The IoT-A Tree does not claim to be fully consistent in its depiction. It should therefore not be taken too strictly: on the one hand, the roots of this tree are spanning across a selected set of communication protocols (6lowpan, Zigbee, IPv6, etc.) and device technologies (sensors, actuators, tags, etc.) while on the other hand the flowers/leaves of the tree represent the whole set of IoT applications that can be built from the sap (information/knowledge) coming from the roots. The trunk of the tree is of the utmost importance here, beyond the fact that it represents the IoT-A project. This trunk represents the Architectural Reference Model (which means here Reference Model + Reference Architecture a.k.a. ARM), the set of models, guidelines, best practices, views and perspectives that can be used for building fully interoperable IoT concrete architectures (and therefore systems). In this tree, we aim at selecting a minimal set of interoperable technologies (the roots) and proposing the potentially necessary set of enablers or building blocks, etc. (the trunk) that enable the creation of a maximal set of interoperable IoT systems (the leaves).

Big Data architecture framework

The other hand of the integrated system architecture is the Big Data domain. Once again there are several reference architecture and models for Big Data, but, at the present time, one of the most prominent is undoubtedly the NIST Reference Architecture [1] for Big Data. The NIST Big Data Reference Architecture (NBDRA) is a high-level conceptual model crafted to serve as a tool to facilitate open discussion of the requirements, design structures, and operations inherent in Big Data. The NBDRA is intended to facilitate the understanding of the operational intricacies in Big Data. It does not represent the system architecture of a specific Big Data system, but rather is a tool for describing, discussing, and developing system-specific architectures using a common framework of reference. The model is not tied to any specific vendor products, services, or reference implementation, nor does it define prescriptive solutions that inhibit innovation.
The main aim of NBDRA can be explained using a pictorial representation. See Figure 4 below.

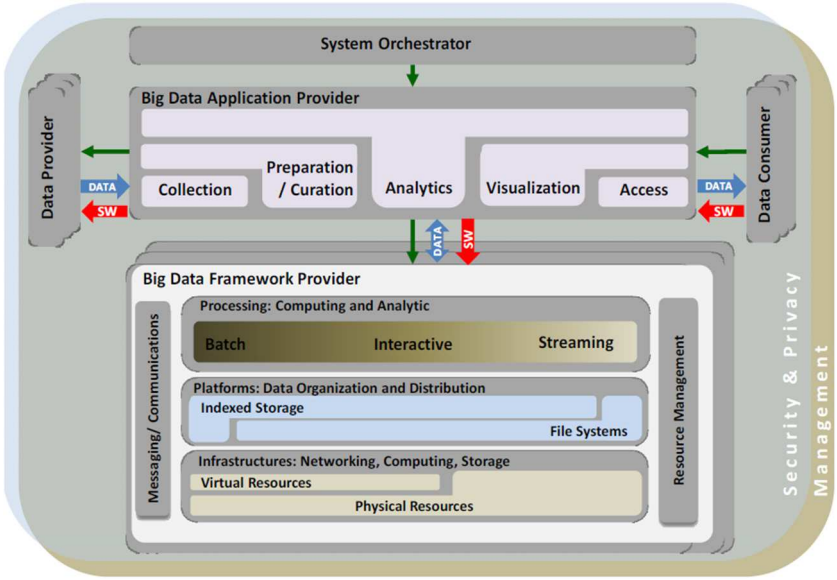

Figure 4 NIST Big Data Reference Architecture (NBDRA) [1]

The NBDRA is organized around two axes representing the two Big Data value chains: the information (horizontal axis) and the IT (vertical axis). Along the information axis, the value is created by data collection, integration, analysis, and applying the results following the value chain. Along the IT axis, the value is created by providing networking, infrastructure, platforms, application tools, and other IT services for hosting of and operating the Big Data in support of required data applications. At the intersection of both axes is the Big Data Application Provider component, indicating that data analytics and its implementation provide the value to Big Data stakeholders in both value chains. The names of the Big Data Application Provider and Big Data Framework Provider components contain "providers" to indicate that these components provide or implement a specific technical function within the system.

System integration
Assuming the IoT area of a project is compliant to IoT-A Reference Model, integrating Big Data and IoT is conceptually easier, because Big Data applications (including Analytics processing) would then be located at the flowers/leaves of the IoT-A Tree. From the Big Data Architecture point of view, the IoT area acts as a data provider.

3.2.2. Big Data platform

Big Data platforms are composed of all computer-related means to collect, process, analyse, and store large volume of data usually characterized by:

- High-Volume (amount of data),
- High-Velocity (speed of data in and out),
- And large Variety (range of data types and sources).

Today, these Big Data processing means can be classified as:

1. **Tools**, specialized in Data Integration and Processing (like Apache Flume, Hadoop Yarn & MapReduce, Storm, Apache Hive, etc.), data storage (like NoSql databases Cassandra, MongoDb, etc.), or data analysis and visualization (like R for analysis and Kibana for Visualization). Each tool deals with one aspect of Big Data. It is up to the user to configure and compose them in order to satisfy the data processing requirements of her/his Big Data application. To get rid of these complex tasks, integrated distributions are available.
2. **Integrated Distributions**, which are ready to use assemblies of tools such those described above, most of them around the Hadoop ecosystem: HortonWorks, Cloudera, MAPR,

Pivotal, HP HAVEn, IBM, WSO2, etc. These distributions are relatively frozen in terms of components choice, non-evolutionary, not specially adapted to Cloud deployment, and sometimes proprietary.

3. **Online Data Processing services**, which provide Big Data related functionalities as a service on the Cloud. Amazon RDS is a relational database service, Amazon EMR is a Hadoop based Big Data processing service, Amazon DynamoDB is a NoSQL database service. Each service deals with one aspect of Big Data processing, but there is no way to easily compose them, to enhance them (add new components), or to deal with data collection and mediation.

The goal in Medolution Platform will be to benefit from the Integrated Distributions advantages, e.g. reduced configuration steps and ready to use availability, while eliminating the related drawbacks, e.g. allowing components choice, evolutionary behaviour, and Cloud deployment. This last point will also permit to make such distributions usage available as online services. More details are presented below.

3.2.2.1. Tools

Storage components:

- Distributed File Systems like HDFS are the basis facilitating huge data processing through Data Distribution (fragmentation) approaches.
- Relational Databases like MySQL or PostgreSQL are still used in Big Data use cases, for structured, concurrently shared data.
- NoSQL databases are used for non or semi structured data. There are many of them among distinct categories: Document oriented (MongoDb), Graph oriented (Neo4j), column oriented (Cassandra, HBase), etc. NewSQL databases are new generation relational databases intending to provide the same power as NoSQL databases for online transactional data processing (read-write) while maintaining ACID properties of traditional relational databases (VoltDb, Nimbus Db).
- Time series database (TSDB): Druid and InfluxDB are examples of a software system that is optimized for handling time series data, arrays of numbers, indexed by time (a datetime or a datetime range) [4].

High Performance search components:

- ElasticSearch [5] is a search engine based on Apache Lucene. It provides a distributed, multitenant-capable full-text search engine with a RESTful Web interface and schema-free JSON documents. It is usually used with LogStash [6], a component for data collect and treatment, and with Kibana [7], a component for visualization of data stored in ElasticSearch. Both of these components will be provided in the Medolution Platform.

Data Analysis Components:

- Hadoop [8] is a Java framework from Apache for creating distributed and scalable applications. It allows working on thousands of nodes and petabytes of data. Main building blocks of Hadoop are HDFS for storage clusters, and MapReduce, a Java framework for parallel processing.
- Mahout [9] is another Apache project providing a Java functions library for modelling and Machine Learning.
- R [10] is a free software environment for statistical computing and graphics, mainly used for data analysis and data mining.
- Spark [11] is an Apache open source cluster computing framework providing implicit data parallelism and fault-tolerance. It intends to provide an alternative to the Hadoop batch processing mode through in-memory processing.

- Storm is an Apache open source distributed real-time computation system to easily and reliably process unbounded streams of data. It is another alternative to Hadoop for real-time processing while Hadoop provides only batch processing.
- Pandas [12] is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language [12].

All these Data Analysis components will be available in the Medolution Platform. Real-time and stream processing components are described in more details in the following Section 3.2.2.2 on Fast Data Processing.

3.2.2.2. Fast Data Processing (low latency)

One issue today in applying Big Data solutions to the healthcare domain is the big data processing latency and the capability to extract insights on-the-fly to deliver them at the right time for the healthcare staff. To solve this issue, one of the Medolution core platform goals is to leverage the latest advances in the stream processing field and to experiment with both the emerging in-memory processing capabilities and lambda architectures approach for deploying new reactive applications. This section describes the considered technologies that have been studied regarding this streaming layer in order to determine if we should integrate and/or adapt them into the Medolution platform.

Real-time Big Data Solutions
Apache Hadoop is one of the first open source projects addressing the Big Data paradigm. Introduced in 2006, it provides a fault tolerant architecture based on the HDFS storage system and MapReduce that allows efficiently parallelizing operations. Data is partitioned in data blocks among the nodes of a cluster, so that the same operations are performed in parallel on local blocks on each node. However, this is a batch-processing mode that results in high latency of applications, which does not satisfy the reactivity requirements of considered real-time applications. This is the reason why a new generation of tools has appeared, which is described below.

Spark & Spark Streaming
Spark [11] is a data processing framework developed by the Algorithms, Machines and People Lab (AMPLab) of Berkeley University since 2009. It has become a top level project of the Apache Foundation in 2013, the same year the Databricks Company is created to ensure its commercialization. Spark is not especially a real-time framework; it is positioned as the basis of a new generation of Big Data processing tools. Indeed, it provides some real-time dedicated modules that allow Spark to cover both batch and real-time modes. In a few words, Spark addresses the drawback of Hadoop that consists in reading & writing permanently in HDFS: Spark loads all data in memory. This allows solving Hadoop performance issues in three kinds of application: iterative problems, graph treatments and real-time. Each of them requires many read operations. Moreover, Spark proposes additional libraries as MLib (Machine Learning), GraphLib that provide distributed algorithms directly applicable. Spark Streaming [13] brings stream processing capabilities to Spark.

Storm
Storm [14] is a real-time analysis framework created in 2011 by BackType. The solution was bought the same year by Twitter and the project became open source. It distinguishes from other real-time solutions through its approach issued from the Complex Event Processing (CEP) domain. Storm does a tuple at a time processing (a tuple is the native data structure of Storm, a set of sorted elements) by processing the event as it comes. To ensure the guaranteed processing of an individual event, an acknowledgement event (ack) is emitted and Storm implements this with a clever mechanism that only requires few bytes of storage per source record to track the acknowledgements.

Samza

Samza [15] has been created by LinkedIn with the goal of handling in real-time the data flow generated by their platforms, in particular geo-localization data, services logs, etc. Since 2013 it has been an open source project hosted in the Apache incubator, and is now a top-level project. It is strongly coupled with Kafka, a message broker open source project (also initiated by LinkedIn), which is its unique direct data source. This does not prevent from using various data sources, since Kafka itself provides many connectors. Samza relies on YARN (Yet another Resource Negotiator: the Hadoop map has reduced resource manager since version 2 of Hadoop), which allows a good level of fault tolerance and resources cluster management. Data processing is performed within Samza Containers based on Linux Containers (LXC). The role of a Samza container is to manage execution of one or several tasks defined within a job. Operations are periodically executed on Kafka event detection.

Apache Apex

Apex [16] is an enterprise-grade unified Batch and Stream processing framework that was originally created by DataTorrent in 2012. It entered the Apache Incubator in August 2015 and has become a top-level project of the Apache Foundation since April 2016, 20th (by summer 2016 when this is written, only the "incubator web page" is available!). Apex is a native YARN framework, which allows a good level of resource management, multi-tenancy and security. While Spark Streaming processing is based on micro-batches and Storm does a Tuple at a time processing, Apex takes a different paradigm, windowed processing, to achieve lower latency. Moreover, The Apex platform comes with Malhar, a library of reusable connectors, and computing operators that can be used to quickly create applications.

Kafka Streams

Kafka Streams [17] is a Java library for building distributed stream processing applications using Apache Kafka, currently in Tech preview and will be part of Kafka 0.10. It is created by Confluent, a new company formed by several engineers who built Kafka and Samza at LinkedIn. In a similar way to Samza, it is strongly coupled with Kafka and leverages its capabilities to offer data parallelism, distributed coordination and fault tolerance. Kafka Streams addresses the drawbacks of Samza by offering more operational simplicity: Getting rid of the dependency on a Hadoop cluster and providing a convenient high-level API. Regarding processing, Kafka Streams API allows for event-at-a-time processing.

Apache Flink

Apache Flink [18] is a community-driven open source framework, originally developed by dataArtisans. It became an Apache Incubator project in March 2014 and was accepted as an Apache top-level project in December 2014. Flink's core is a streaming dataflow engine and it uses one runtime for streaming and batch processing. It comes with a rich set of features including support for event time and out-of-order events, highly flexible streaming windows, iterative computations, Storm compatibility and its own memory management. Moreover, Flink integrates with the rest of the data-processing ecosystem (YARN, HDFS, Kafka, etc.) and offers libraries with high-level APIs for Machine learning, Graph Analytics, and Relational Data Processing.

Apache Beam

Apache Beam [19] is an open source, unified programming model for defining and executing data processing workflows. It became an Apache Incubator project in February 2016. Beam is based on the research of Google on the Millwheel and FlumeJava papers, and is the resulting project of Google open sourcing its Cloud Dataflow SDK. Its goal is to unify all the data processing engines under one API: data processing pipelines defined with the Beam model can be run by any number of Beam processing engines. Several engines have been developed to run Beam pipelines, currently including a Beam runner for Apache Flink and Apache Spark.

Lambda Architecture

Originally introduced in 2011 by Nathan Marz (creator of Apache Storm) [20], the lambda architecture has been thoroughly researched and elaborated since then [21], [22]. The lambda architecture is a Big Data processing architecture combining the run of a streaming processing system (real time layer) alongside a batch processing system (batch layer), both performing essentially the same calculation on an input data stream. The idea is to provide low latency data analysis results through the real time layer, with the ability to consolidate these results using the batch layer, since the real time layer may provide inaccurate results (either because of the use of an approximation algorithm, or because the streaming system itself does not provide correctness). Figure 5 below illustrates this approach schematically.

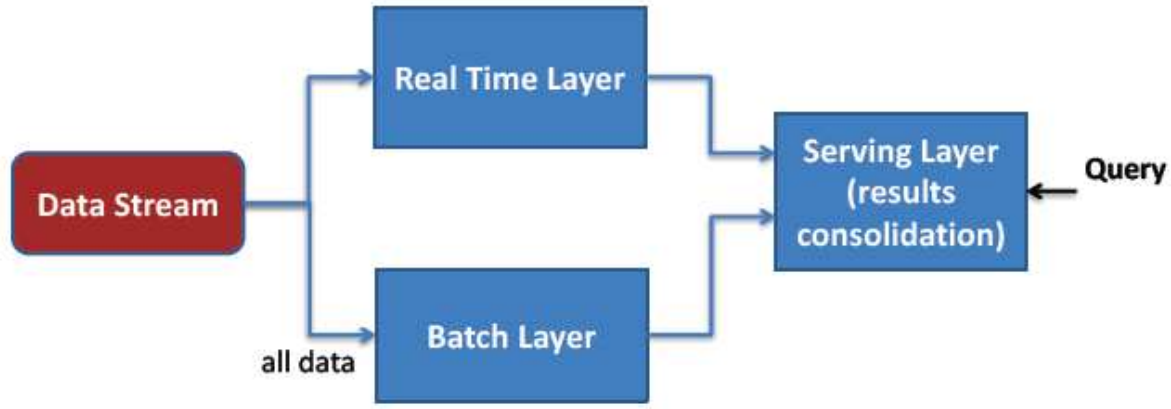

Figure 5 Lambda Architecture

The lambda architecture will be experimented within Medolution project, as well as alternatives to it, since new solutions for dealing with real time big data processing are emerging every day.

3.2.2.3. Integrated Distributions

HortonWorks is an Open Source Big Data platform featuring the Hadoop ecosystem (See Figure 6 below).

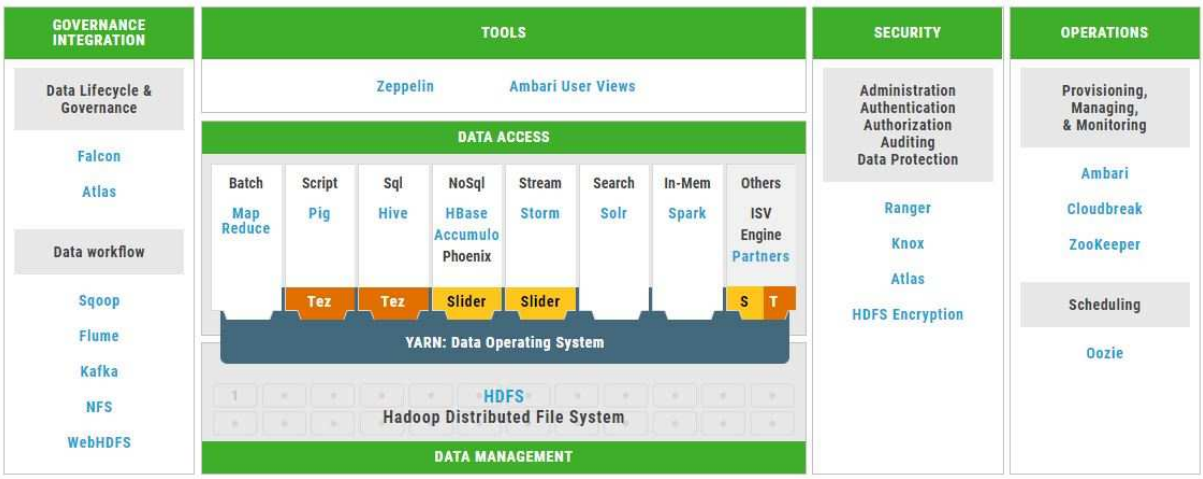

Figure 6 HortonWorks [23]

Cloudera [24] is an open source "Enterprise Data Hub" with a proprietary administration tooling (See Figure 7 below).

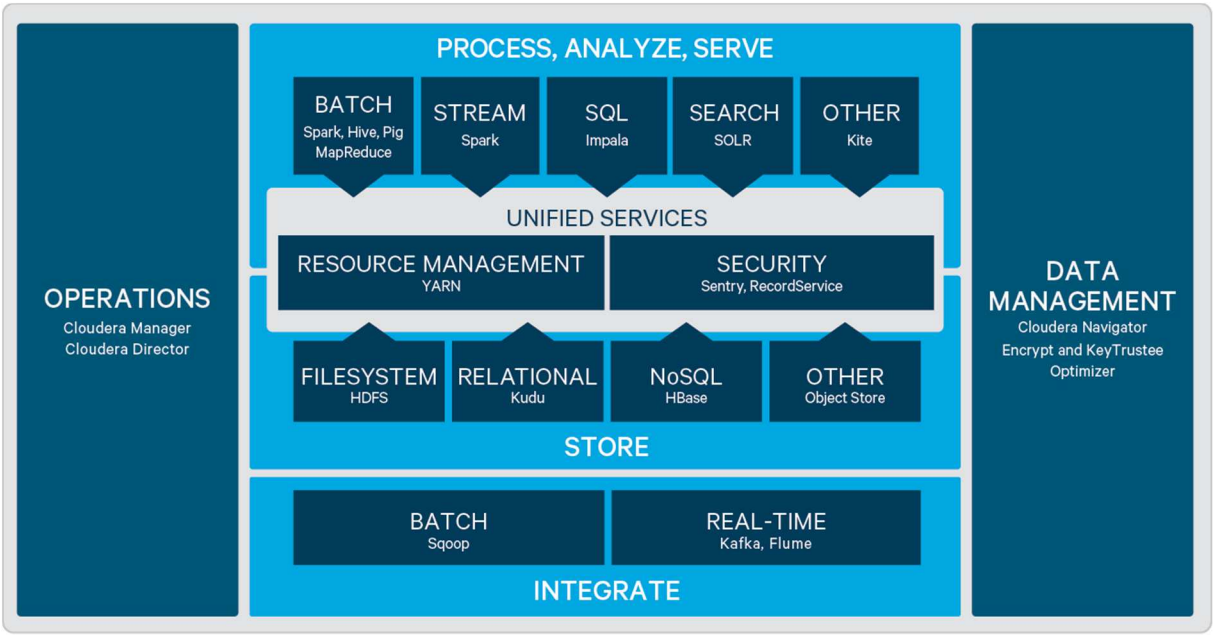

Figure 7 Cloudera [25]

MapR [26] is a Hadoop distribution based on a specific file system (MapRFS) and a proprietary administration tooling (See Figure 8 below).

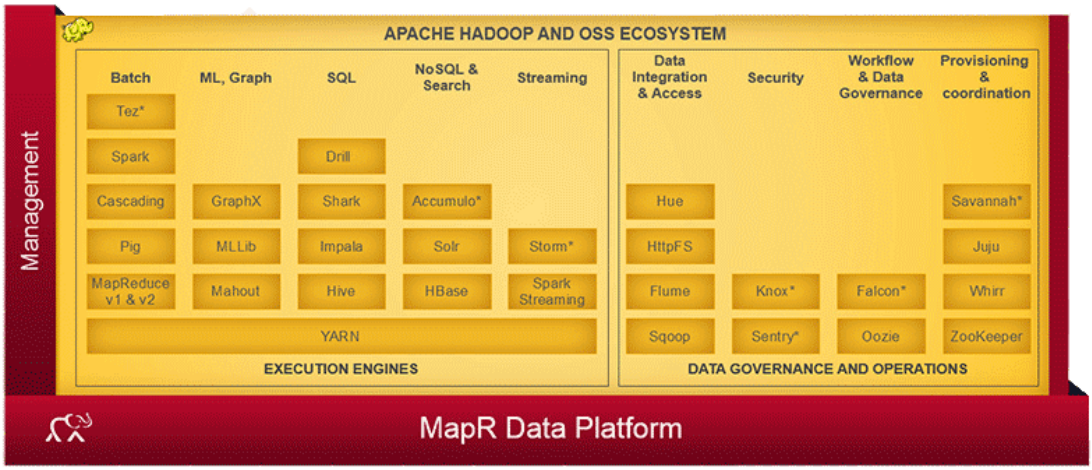

Figure 8 MapR [26]

MapR, HortonWorks, and Cloudera Hadoop distributions will be available in the Medolution Platform as they are considered as the most used Hadoop distributions; platforms described below for illustrative purposes are considered as more proprietary stacks and won't be addressed in the Medolution platform.

Pivotal [27] is a Big Data stack built on top of its core Hadoop based component Pivotal HD, HAWQ and GemFire HD. It targets Business Data Lakes approaches (See Figure 9 below).

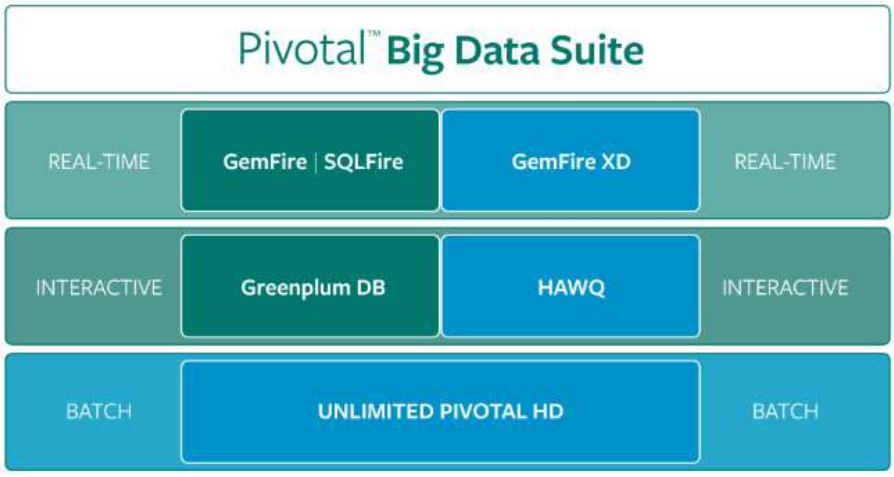

Figure 9 Pivotal Bid Data Suite [27]

HP HAVEn is a Big Data software suite composed of a Hadoop stack, a text oriented search component "Autonomy IDOL", and a column oriented database "Vertica" and a data access auditing tool "ArtSight Logger" (See Figure 10 below).

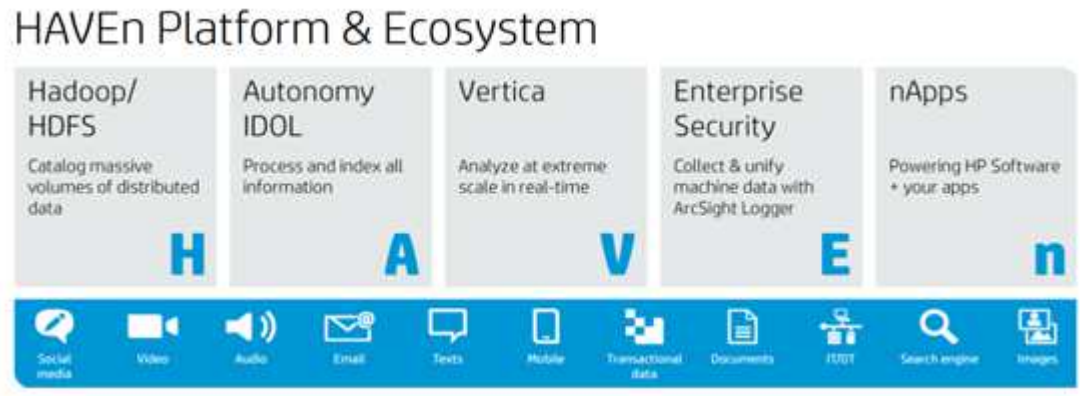

Figure 10 HP HAVEn [28]

3.2.2.4. Online Data Processing Services

A number of vendors currently provide "online Big Data Services" offerings on the market. Some examples of such online Big Data Service are discussed below for illustrative purposes.

Talend Big Data Sand Box is a ready to use environment combining the Talend Big Data Platform based on a Hadoop distribution (Cloudera, MapR or HortonWorks) and ready to run Big data scenarios. This allows users to install and configure a Big Data environment and to quickly build prototypes. The idea of providing an easy to configure virtual environment is interesting but for now this offer is "closed", "pre-packaged" and targeting prototypes only offered by Talend as an online service.

Amazon DynamoDB is a NoSQL database service fully managed and evolutionary. It automatically manages data throughput and storage capacity, as well as replication.

Amazon RDS (Relational Database Service) allows to easily configure, manage and scale a relational database among Amazon Aurora, Oracle, Microsoft SQL Server, PostgreSQL, MySQL and MariaDB.

Amazon EMR (Elastic Map Reduce) is an Amazon Web Service (AWS) for data processing and analysis. Amazon EMR processes data across a Hadoop cluster of virtual servers on the Amazon Elastic Compute Cloud (EC2). The elastic in EMR's name refers to its dynamic resizing ability.

Within Medolution, similar functionalities will be provided on premises on the hosting platform. Considering the data privacy concerns related to health data, it cannot be envisaged to process it on such hosted public online services. However, the capability to deploy some Medolution platform Big Data services on Amazon EC2, for e.g. some proofs of concepts or when data is not critical can be considered.

3.2.3. Big Data Infrastructure as a Service

Data streaming from billions of sources can provide predictive insights into customers, business risks and operational efficiencies. But cost-effective analysis of users' data from information silos and secure sharing of analytics across an organization is very complex.

For service providers, there are multiple ways to address the Big Data market with as-a-Service offerings. These can be roughly categorized by level of abstraction, from infrastructure to analytics software, as shown in Figure 11 below.

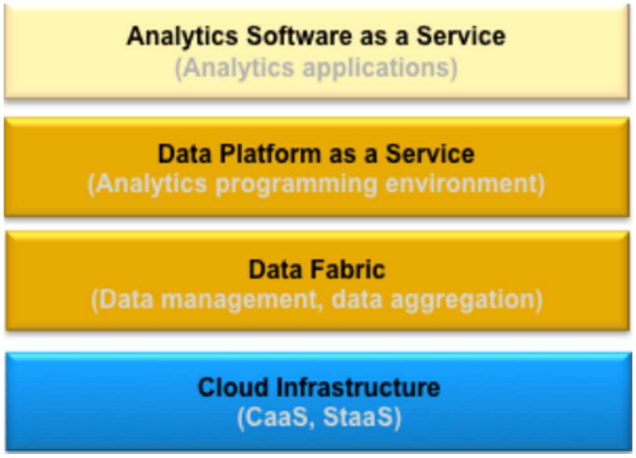

Figure 11 Big Data as a Service layer [1]

Starting from the bottom layer, any Big Data-as-a-Service infrastructure will usually leverage Infrastructure-as-a-Service components, particularly Compute-as-a-Service Cloud (CaaS) and Storage-as-a-Service resources and their management.

Also, a lot of data is actually generated by applications deployed in a service provider's Cloud infrastructure. Moving large amounts of data around can be prohibitive in some scenarios. Hence having the data that will be further processed already available in the service provider's infrastructure enables Big Data services to be a natural enhancement to a service provider's.

Most commonly, Big Data users are not seeking a standalone infrastructure as a Service but a whole Big Data Platform as a Service (BDPaaS) to help get value from their data much more quickly. Thus relying on the BDPaaS, users can rapidly develop, secure and deploy next-generation Big Data and analytics applications with a centralized, subscription-based platform that uses leading analytics tools, infrastructure and software.

An example of a BDPaaS offers an integrated as-a-service solution in the Cloud by combining:

- A Hadoop distribution (e.g. MapR) which allows Data Scientists users to store unstructured data in persistent Hadoop Distributed File System and processing it in MapReduce.
- A Cloud management system (e.g. Cloudify, and/or Alien4Cloud)
- A distributed Data Base solution based on SQL or NoSQL data base for Big Data (e.g. Apache Cassandra).
- Large Scale data processing engine (e.g. Spark or Spark Streaming)
- A data integration/ingestion tool offering Extraction, Transformations and Load functions (e.g. Pentaho).
- A data discovery tool (e.g. Qlik)
- A data visualization and comprehension tool (e.g. Tableau)
- Data analytics programming tools (e.g. pbdR: programming with Big Data in R).

3.2.4. Big Data Applications as a Service

Big Data Application as a Service points out the relationship between Big Data applications (software) and the Cloud, which in one word means providing a Big Data Software as a Cloud service. With a connected ecosystem of apps in the Cloud linking global and consolidated data stores, interesting and meaningful relationships are made possible. Not only can data be analysed on the same platform and across standardized structured and unstructured data, it can be linked to other applications in real-time. Users can also control their own access permissions to other applications for their data stored in the Cloud, bypassing potentially complex and often costly business process bottlenecks. Providing Big Data application in the SaaS manner creates a myriad of potential business opportunities in various areas such as healthcare, travel, finance, government [29], media, sale and marketing etc.

An example is the Top 3 Software-as-a-Service solution for Big Data analytics companies for 2015 provided by CEO WORLD Magazine [27].

- **1010data, New York, NY** – 1010data is one of the leading providers of Big Data discovery and data sharing solutions. It is used by hundreds of the world's largest retail, manufacturing, telecom, and financial services enterprises because of its proven ability to deliver actionable insight from very large amounts of data more quickly, easily and inexpensively than any other solution.
- **Alteryx, Inc., Irvine, CA** – Alteryx is the leader in data blending and advanced analytics software. Alteryx Analytics provides analysts with an intuitive workflow for data blending and advanced analytics that leads to deeper insights in hours, not the weeks typical of traditional approaches.
- **Ayasdi, Palo Alto, CA** – an advanced analytics company that provides Machine Learning software to Fortune 500 companies to solve their complex data challenges. Ayasdi pioneered the use of Topological Data Analysis (TDA), to simplify and accelerate complex data analysis.
- IBM and Apple and Google's Brillo [30] have been developing Big Data health platform, that apply Big Data techniques along with other fields like Statistical Modelling, Machine Learning, Data Mining etc.

A great example of interoperability efforts enabling provision of a Big Data Software as a Cloud in the Healthcare domain are Apple HealthKit and Google Fit. Apple HealthKit is a personal health database and tool kit that allows health applications and smart devices to pool all of a user's biometric information into one location. Apple HealthKit is designed to manage data from a wide range of sources, automatically merging the data from different sources based on user's preferences. Applications can also access the raw data for each source, merging the data themselves. Creating the modern health and fitness experience by Apple HealthKit involves many

different facilities, such as collecting and analysing data, providing actionable information and useful visualizations to the user, and fostering a social community. Apple HealthKit technology is being used to monitor diabetes or hypertension sufferers, and alerting doctors if data indicates possible problems. The physicians can thus intervene before the problem becomes acute, allowing hospitals to prevent repeat admissions. Apple's HealthKit also collects information from various sensors and devices such as glucose measurement tools, food and exercise-tracking apps and Wi-Fi connected scales. Google Fit is a health-tracking platform that uses sensors in a user's activity tracker or mobile device to record physical fitness activities (such as walking and cycling). Users can choose who their fitness data is shared with as well as delete this information at any time. Other than Apple HealthKit, Google Fit is also a health-tracking platform focussing on personal fitness rather than personal health. In Medolution adaptors to collect data from Apple HealthKit and Google Fit will be developed.

3.3. Limitations of existing solutions and relevance for Medolution

Medolution will rely on Big Data Platforms to deal with the huge volume of health data made available through the use cases, and to make most of the analysis results available in real-time. We have shown that Big Data related tools deal with one aspect of Big Data, and it is up to the user to configure and compose them in order to satisfy the data processing requirements of her/his Big Data application. To get rid of these complex tasks, integrated distributions are available. These distributions are relatively frozen in terms of components choice, non-evolutionary, not specially adapted to Cloud deployment, and sometimes proprietary. Online Big Data services deal each about one aspect of Big Data processing, however, there is no way to easily compose them, to enhance them (add new components), or to deal with data collection and mediation. All this makes it very difficult to use Big Data technologies for building applications, and the Medolution project has to cope with these issues to satisfy its objectives.

4. Dependability

In the Medolution project typically safety-critical medical systems are to be developed. Particular dependability requirements exist, thus the systems have to fulfil increased reliability, safety and security requirements, even in the presence of exceptions and faults. Therefore, Medolution has to develop dependable system architectures, which include appropriate fault-tolerance mechanisms and can support dependability when applying non-reliable devices or sensors. This section provides a general overview of the main aspects of dependability, which will be further examined and applied though out the project, in particular in the context of high risk medical devices (for example, left ventricular assist devices (VLAD)).

4.1. Fundamentals of dependability

In times of digitization and miniaturization of electronic devices and the interconnection of heterogeneous devices, the issue of dependability is more important than ever. Computer systems play an increasing role in today's society. In airplanes, chemical plants, medical devices or other safety critical applications, a system failure can cost people's lives. From this point of view, the definition of dependability is the ability of a system to deliver services to its users [31]. Figure 12 below shows the three fundamental properties of dependability, which will be explained in more detail below. The threats of dependability describe the things that can affect the dependability of a system. The attributes of dependability describe the characteristics required from a system. The last property is the means; these are techniques to improve the dependability during the development process.

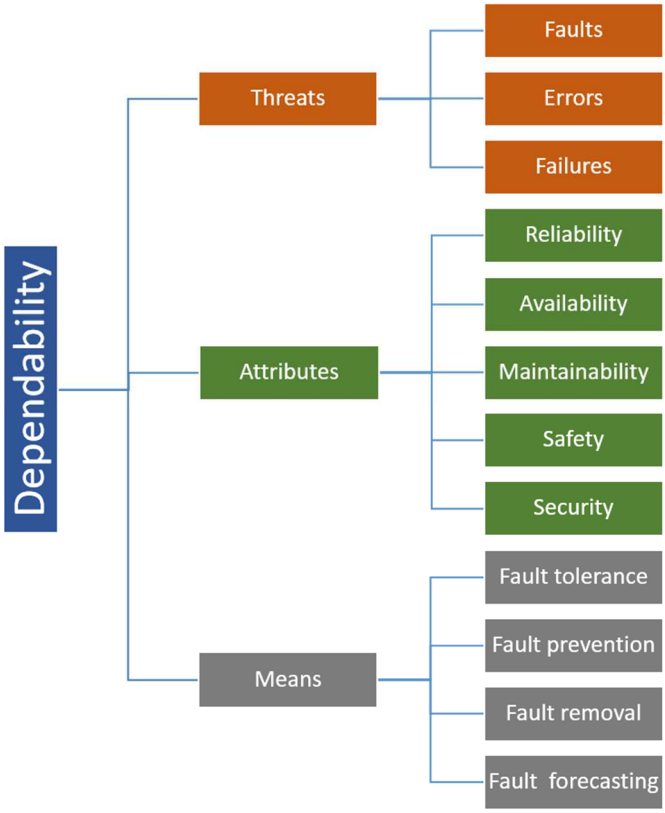

Figure 12 Dependability tree [31]

4.1.1. The threats of dependability

A system may fail because its operation does not comply with the specification, or because the specification has not been described sufficiently. An error is that part of the system state that may cause a subsequent failure: a failure occurs when an error reaches the service interface and changes the service. A fault is a defect and it is active when it produces an error, otherwise it is dormant [32].

Looking at the threats in more details, we observe that faults, errors and failures operate according to the chain as shown in Figure 13 below.

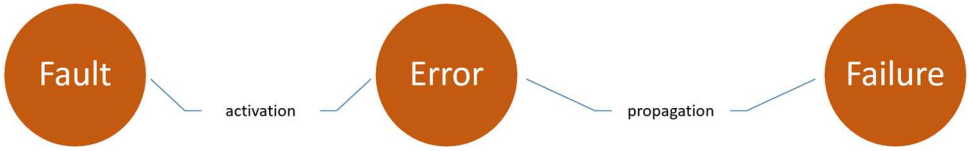

Figure 13 Fundamental chain of dependability threats [32]

1. A fault is active when it produces an error, otherwise it is dormant. An active fault is either an internal fault that has been activated by the computation process or environmental conditions, or an external fault. [32] Examples of fault activation are: execution of a buggy line of code, sending a signal via a damaged connector, or execution of a defective hardware
2. The presence of an error within a system can arise from:
    a. activation of an internal fault
    b. occurrence of a physical operational fault, either internal or external
    c. propagation of an error from another system → that is an input error
3. A failure occurs when an error is propagated to the service interface and unacceptably alters the service delivered by the system [32]. A failure of a component causes a permanent or transient fault in the system that contains that component. Failure of a system causes a permanent or transient external fault for the other system(s) that interact with the given system [32].

4.1.2. The attributes of dependability

The attributes of dependability describe the properties that are expected from a dependable system. The five attributes are:

- Reliability
- Availability
- Maintainability
- Safety
- Security

These are also called *RAMSS* properties. Depending on the application, one or more of these attributes are needed to appropriately evaluate the system behaviour.

Reliability: The reliability of a system is the probability that it will perform without deviations from agreed-upon behaviour for a specific period of time. Measures used to describe reliability are Mean Time To Failure (MTTF) and Mean Time To Repair (MTTR) [33].

- The MTTF is the average time from start of operation until the time when the first failure occurs.

- The MTTR is a measure of the average time required to restore a failing component to operation.

Availability: A systems availability is the percentage of time that it is able to perform its designed function. Uptime is when the system is available; downtime is when it is not. A common way to describe availability is in terms of number of nines.

Maintainability: As systems are used, new requirements emerge and it is important to maintain the usefulness of a system by changing it to adjust these new requirements. Maintainable software is software that can be adapted economically to cope with new requirements, and where there is a low probability that making changes will introduce new errors into the system [34].

Safety: The safety of a system is an evaluation of how likely it is that the system will cause damage to people or its environment. Safety-critical systems are systems where it is essential that system operation is always safe; that is, the system should never damage people or the environment even if the system fails [34]. Examples of safety-critical systems are: control and monitoring systems in aircraft, process control systems in chemical and pharmaceutical plants and medical device control system.

Security: Security is an attribute that reflects the ability of the system to protect itself from external attacks, which may be accidental or deliberate. These attacks are possible because most devices are now networked and are therefore accessible by outsiders [34]. Examples of attacks might be: viruses and Trojan horses, unauthorized use of system services, unauthorized modification of a system or its data.

4.1.3. The means of dependability

The development of dependable systems requires the use of four fault recovery techniques [32]:

1. Fault tolerance: how to execute correct services in presence of faults
2. Fault prevention: how to prevent faults or, at least, their propagation
3. Fault removal: how to reduce the number or severity of faults
4. Fault forecasting: how to estimate the future incidence and consequences of faults

These means improve the development process of dependable systems. This section describes the different techniques in more details, with the main focus on the fault tolerance, in more detail. An overview is shown in Figure 14 below.

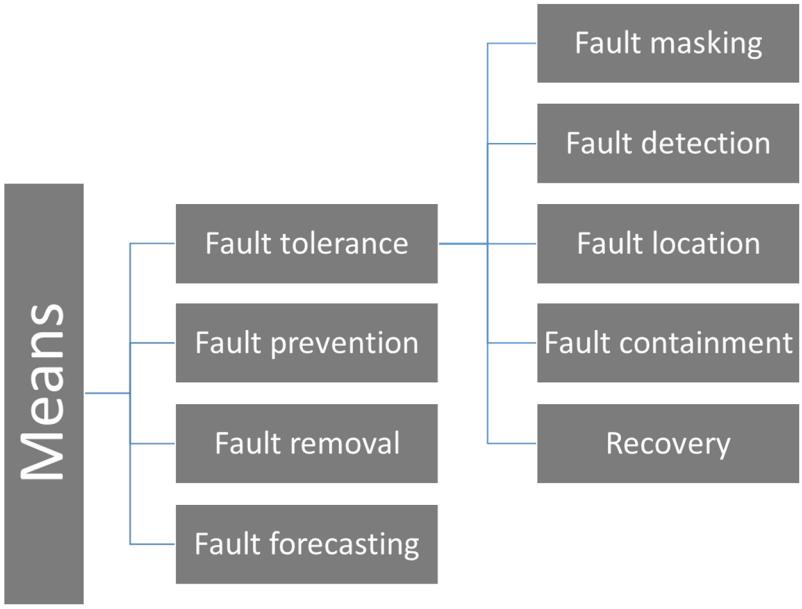

Figure 14 Classification of mean classes [32]

Fault tolerance

The main goal of fault tolerance mechanisms is the correct execution of services in presence of faults. For this purpose, the possibilities of *redundancy* and *diversity* are used on the hardware and software level. Redundancy is the provision of additional functional capabilities that would be unnecessary in a fault-free environment. The redundancy allows either to *mask a fault*, or to *detect a fault*, with the following characteristics of *location*, *containment* and *recovery*.

Fault masking is the process of ensuring that only correct results are delivered even though an error has occurred. This is done by preventing the system from being affected by errors, either by correcting that error, or by compensating it. The impact of the fault remains hidden to the end user [31]. Examples for fault masking are: memory protected by an error-correcting code corrects the faulty bits before the system uses the data, and triple modular redundancy (TMR) with majority voting. The fault masking technique is known as a *passive redundancy technology*.

Fault detection serves to detect faults that have occurred in a system [31]. Examples for fault detection are:

- Acceptance tests: The result of a program is subjected to a test. If the result passes the test, the program continues execution. A failed acceptance test implies a fault.
- Comparison/Duplication with Comparison: Comparison is used for systems with duplicated components. A disagreement in the results indicates the presence of a fault.

The fault detection technique is known as an active redundancy technology.

Fault location is used to determine at which point in a system a fault has occurred. An acceptance test, as it is used in the fault detection, cannot generally be used to locate a fault. It can only highlight that something has gone wrong. The same applies when a disagreement occurs during the comparison of two modules, as it used in the fault detection. It is not possible to tell which of the two modules are failed [31]. Examples for fault location are: standby sparing and pair-and-a-spare[1]. The fault location technique is known as an *active redundancy technology*.

[1] Concept of sparing relates to the situation when one component is active and operating and one or more components serves as standby. If an active component is affected by an error, this is replaced by a standby component.

Fault containment is the process of isolating a fault and preventing the propagation of the effect throughout the system. The aim is to limit the spread of the effects of a fault from one component of the system into another component [31]. This is achieved by frequent fault detection, multiple request/confirmation protocols and performing consistency checks between modules

System recovery is the process of the reconfiguration of affected and isolated components back to normal operation mode. This could be done by replacing an affected component by marking it off-line and by using a redundant system. Another option would be the system could switch it off and continue operation with an impaired capability [31].

Fault prevention
Fault prevention is guaranteed by quality controls during the specification-, implementation- and fabrication-process. A distinction is made between hardware and software quality controls [31]. For hardware, these are the following techniques: design reviews, component screening, testing.

For software, these are the following techniques: structural programming, modularization, formal verification. A careful design review prevents many specification faults. Through the efficient and accurate testing of a design, many errors and component defects can be avoided.

Fault removal
Fault removal can take place in two phases, on the one hand during the development phase as well as during as during the operational life. During the development phase, fault removal consists of three steps: 1) verification; 2) diagnosis; 3) correction.

Fault removal during the operational life of the system consists of 1) corrective maintenance and 2) preventive maintenance.

The latter faults include: 1) physical faults that have occurred since the last preventive maintenance actions and 2) design faults that have led to errors in other similar systems

Fault forecasting
Fault forecasting is managed by performing an analysis and evaluation of the system behaviour with respect to fault occurrence or activation. Evaluation has two aspects:

1. *Qualitative evaluation*, which intends to identify, classify, rank the failure modes, or the event combinations that would lead to system failures [32].
2. *Quantitative evaluation*, which intends to evaluate in terms of probabilities the extent to which some of the attributes of dependability are satisfied; those attributes are then viewed as measures of dependability [32].

4.2. Architectural elements and modelling languages for supporting dependability

Model-based system engineering (MBE) can substantially contribute to the efficient development of high-quality systems, particularly to their functional correctness. When using suitable architecture description languages, e.g., SFM, SysML or AADL discussed further in this Chapter, reliability and fault tolerance properties can be checked in early development phases. Moreover, the certification of the developed system is supported since the model-based safety analysis increases the evaluation assurance level. All these elements have direct relevance and application for addressing various aspects of the big dependable system to be developed under Medolution and are discussed in this section.

To evaluate the dependability of a system we need to first create a model of the system. This model of course has to approximate the real system. But we must also be able to analyse the model to

find out the dependability of the system. A number of different formalisms can be used to model dependable systems. We have divided the formalisms into three categories.

Dependability-specific models focus on modeling structures and phenomena that often appear in dependable models, such as the use of spare components and the propagation of faults through a system. The advantage of dependability specific models is that they are usually easy to use and compositional. Models are built by connecting predefined building blocks. The disadvantage is then that we are restricted to whatever building blocks are provided to us. Also dependability specific models can usually not be analyzed directly. Instead they are often transformed to a low-level mathematical model. Examples of dependability specific formalisms are:

- Continuous-time, exponential Markov chains (CTMC): A stochastic model based on labelled transition systems and exponential distributions.
- Generalized stochastic Petri nets (GSPN): A stochastic extension of Petri nets.
- Stochastic activity networks (SAN): A variation on stochastic Petri nets, which is geared towards dependability modelling.
- Stochastic process algebra (SPA): A process algebra with stochastic processes.
- Interactive Markov chains (IMC): A compositional stochastic extension to labelled transition systems.

Low-level models describe the behavior of systems in great detail. They are usually based on automata theory and Markovian stochastics. The advantage of this is that these models are usually very expressive and can be readily analysed. The disadvantage is that it can be difficult to create a low-level model of a large system because the model will also be very large.

Architectural approaches focus on the structure of the system being described. The idea is to start with a very abstract view of the system and then to refine this view to lower levels of abstraction. Analysis of architectural models is based on the one-model-multiple-analysis idea. The goal is to have a single model of a system on which we can perform many different types of analysis. The advantage of architectural models is that they are usually very well structured and can be used in sophisticated software/hardware development methods. The disadvantage is that it is often quite difficult to analyse architectural models, which is also very true for the dependability aspects of such models. Below a few architectural methodologies are listed which allow the specification of dependability features:

- System AVailability Estimator (SAVE): A dependability analysis tool which uses also uses architectural models to describe systems
- Unified Modelling Language (UML): A language that encompasses a great many aspects of computer systems. UML contains meta-languages which allow new (dependability) aspects to be added to existing models
- Architecture Analysis and Design Language (AADL): Another very broad architectural modelling language. A recent addition to AADL is the Error Annex which allows the modelling of dependability aspects [35].

To provide dependability throughout the whole System Development Life Cycle (SDLC) it is a benefit to evolve the product based on the MBE process. MBE is a software and system development paradigm. It aims to represent the whole System Development Life Cycle by architecture descriptions following visual modelling principles.

4.2.1. Architecture Description

The specification of A*rchitecture* D*escription* standardized by the ISO/IEC/IEEE 42010 [36] standard defines the practices, techniques and types of representations used by software architects to model,

analyse and represent a software architecture. This representation can take various forms to clarify specific concerns of interest to different stakeholders of the system.

There are several common mechanisms used for architecture description. These mechanisms facilitate reuse of best practice of description so that they may be applied to many systems:

- Architecture views and viewpoints
- Architecture frameworks
- Architecture description languages

Architecture views and viewpoints

Software architecture descriptions are commonly organized into views. An architecture view addresses a set of concerns held by the system's stakeholders. The ISO/IEC/IEEE 42010 standard defines the architecture views and viewpoints as:

> *"An architecture view expresses the architecture of the system-of-interest in accordance with an architecture viewpoint (or simply, viewpoint). There are two aspects to a viewpoint: the concerns it frames for stakeholders and the conventions it establishes on views.*
> *An architecture viewpoint frames one or more concerns. A concern can be framed by more than one viewpoint.*
> *A view is governed by its viewpoint: the viewpoint establishes the conventions for constructing, interpreting and analysing the view to address concerns framed by that viewpoint. Viewpoint conventions can include languages, notations, model kinds, design rules, and/or modelling methods, analysis techniques and other operations on views."* [36]

Architecture frameworks

The ISO/IEC/IEEE 42010 standard defines the a*rchitecture frameworks* as conventions, principles and practices for the description of architectures established within a specific domain of application and/or community of stakeholders.

Architecture frameworks of interest in software development include:

- "4+1" view model
- Reference Model for Open Distributed Processing (RM-ODP)
- The Open Group's Architecture Framework (TOGAF)
- Zachman's information systems architecture framework
- Generalized Enterprise Reference Architecture (GERA)

Architecture description languages

Architecture Description Languages (ADLs) are computer languages describing the software and hardware architecture of a system. The description contains hardware components like processors, devices and memory as well as software features like processes, data and threads. Often these languages allow a logical as well as a physical description of the connections between the components.

Examples of ADLs that could be relevant for the Medolution project are:

- Struktur-Funktions-Modell - is an abstract modelling language that supports analysis of the systemic interrelationships for the treatment of fault tolerance methods.
- Systems Modelling Language (SysML) - is a Domain-Specific modelling language for systems engineering that is defined as a UML profile.
- Architecture Analysis & Design Language (AADL) - is a modelling language that supports early and repeated analysis of a system's architecture.

4.2.2. Struktur-Funktions-Modell

The Struktur-Funktions-Model [37] (SFM) (Structure-function model) is a graphical model language that is defined by Klaus Echtle. On the one hand, the structural description aspect of SFM is a possibility to model error occurrence and propagation. On the other hand, the functional description characterizes the operation of fault-tolerance methods.

Elements of a SFM, as shown in Figure 15 below, are:

- *Nodes*: Components in a SFM are modelled by nodes. A component represents a hardware or a software element of a system.
- *Directed edges*: Directed edges symbolize the allocation of functions between components of a system. An edge can be understood in the following ways:
  - *K1 is a part of K2*
  - *K1 provides resources available for K2*
  - *K1 can be invoked from K2*
  - *K1 provides a service for K2*

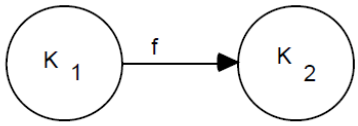

Figure 15 Elements of SFM [37]

The whole graph of a SFM represents a system. Systems can be structured hierarchically by the aggregation of subsets of nodes and edges to subsystems. A subsystem described by an inner and an outer specification. The edges connecting internal nodes of the subsystem with nodes of the systems or of other subsystems act as outer specification.

4.2.3. Systems Modelling Language

Systems Modelling Language (SysML) [38] is a general-purpose graphical modelling language that is defined and standardized by the Object Management Group (OMG). The goal of SysML is the design, analysis, specification, verification and validation of complex systems. These systems may include hardware, software, information, personnel, procedures and facilities. SysML can represent the following aspects of systems, components and other entities [39]:

- Structural composition, interconnection and classification
- Flow-based, message-based and state-based behaviour
- Constraints on the physical and performance properties
- Allocations between behaviour, structure and constraints
- Requirements and their relationship to other requirements, design elements and test cases

Diagram taxonomy

SysML reuses a subset of UML 2.0 and provides additional extensions needed to address requirements. To transpose these extensions, SysML adds to UML new diagrams and modifies others. The goal of the OMG was to use UML 2.0 as much as possible while avoiding changes, unless absolutely necessary. Finally, SysML includes nine diagrams, as shown and summarised in Figure 16 below:

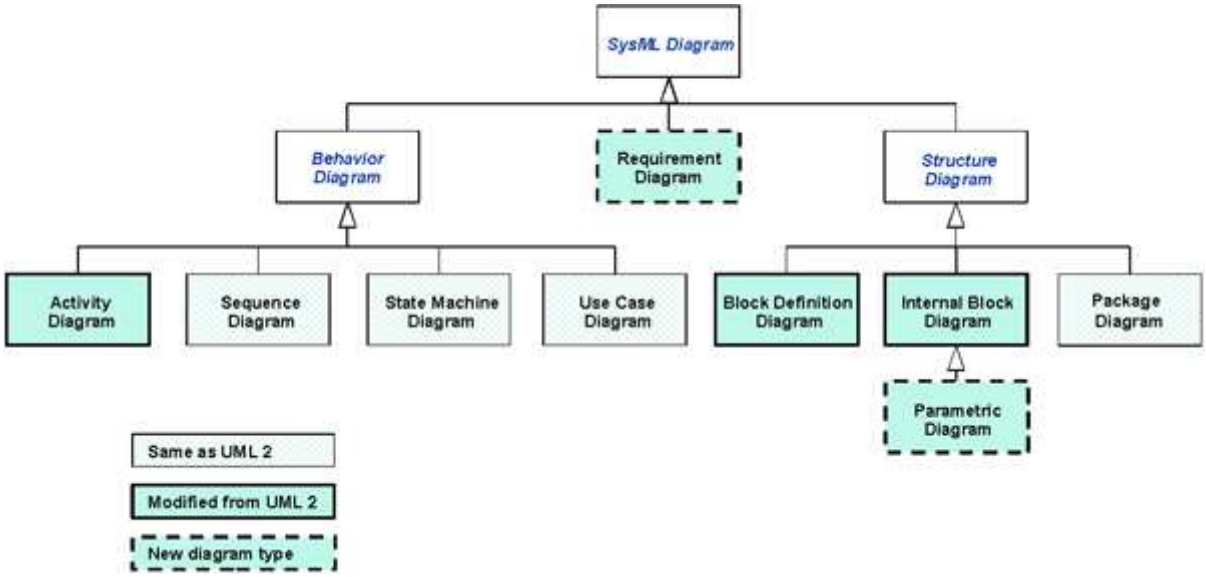

Figure 16 SysML Diagram Taxonomy [40]

This list of diagrams is not exclusive. A benefit of SysML is the possibility to extend it with other UML 2.0 diagrams if necessary. Additionally, it is recommended to combine SysML with other architecture and model languages.

Diagram elements
As described above, SysML consists of nine basic diagram types: four are the same as in UML 2.0, three are modified and two new types are added. To describe the diagram elements, we only concentrate on the modified and new diagram types that extends UML 2.0.

Activity Diagram
The Activity Diagram (AD) is the first modified diagram and belongs to the class of behaviour diagrams of SysML. The AD shows the flows of data and control between actions. It was already presented in UML 2.0 and is slightly modified in SysML. The most important modification is the adding of the possibility to model continuous flows. The basic elements of the AD are actions, controls, decisions, entry and exit points.

To enable the modelling of continuous systems, SysML added the possibility of characterizing the nature of the rate at which the flow circulates:

- Continuous (e.g. electric current fluid) or
- Discrete (e.g. events requests).

Block Definition Diagram
The Block Definition Diagram (BDD) is the second modified diagram and belongs to the class of structure diagrams of SysML. The BDD replacing the classic UML class diagram. The diagram is used to represent blocks, their properties and their inter-relations. Properties are the basic structural characteristics of blocks and these properties may be of several types:

- *Value* properties describe quantifiable characteristics in terms of value types.
- *Part* properties describe the decomposition hierarchy of the block in terms of other blocks.
- *Reference* properties describe relations of association or simple aggregation with other blocks.

The extensions of SysML for the BDD are the flow ports. These ports represent which data entering or leaving a block.

Internal block diagram

The Internal Block Diagram (IBD) is the third modified diagram and belongs to the class of structure diagrams of SysML. This diagram describes the internal structure of a block in terms of *parts*, *ports* and *connectors*. The IBD provide the internal view of a BDD.

- *Part*: Each end of the composition relationship that exists in the BDD is presented as a block (part) in the internal block diagram.
- *Connector*: The connector is a structural concept used to connect two parts and to provide them with the opportunity to interact.
- *Port*: The port concept in the internal block diagram also enables descriptions of the logic behind the connection, services and flows between blocks. Every part can contain several ports. There are two types of ports:
  - Standard port: descripts logical services between blocks, through interfaces as known from UML2.
  - Flow port: this type of port is new to SysML and can be used to produce a representation of the physical flows between blocks.

Requirement diagram

The first new diagram to SysML is the Requirement Diagram (RD). This diagram offers a graphical representation of requirements. The two basic properties of a RD are, first, a unique identifier, and second, a textual description of the requirement. Other properties could also be defined. The requirements can be connected to one another by relations of containment, refinement or derivation relations. An example of a requirement diagram is represented in Figure 17 below.

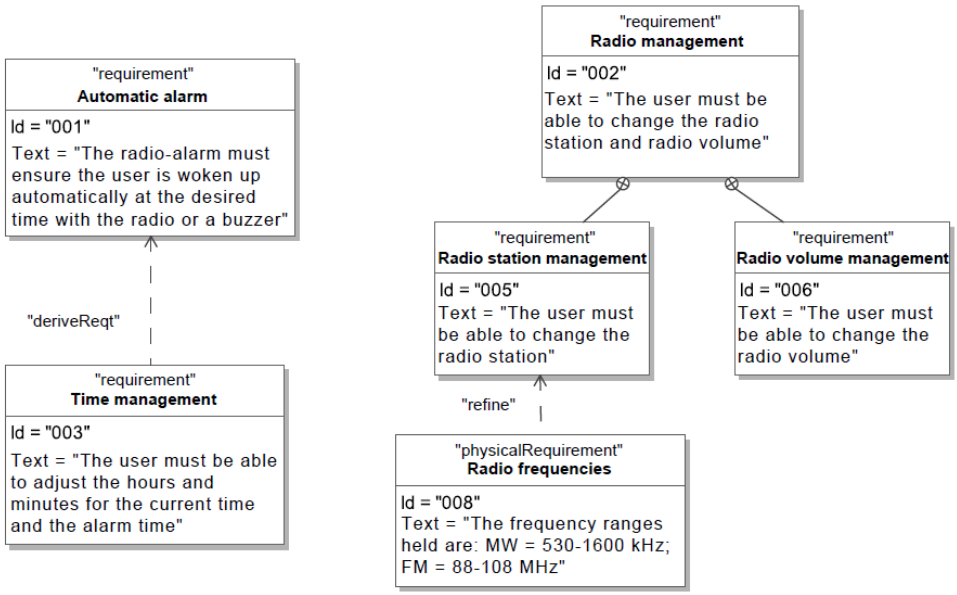

Figure 17 Example of a requirement diagram [41]

Parametric diagram

The parametric diagram enables constraints on system parameter values to be represented, such as performance, reliability and mass. Thereby each constraint is defined by parameters and a rule. This rule describes the relationship of the parameters with regard to one another. As shown in Figure 18 below, constrains must be declared in a block definition diagram.

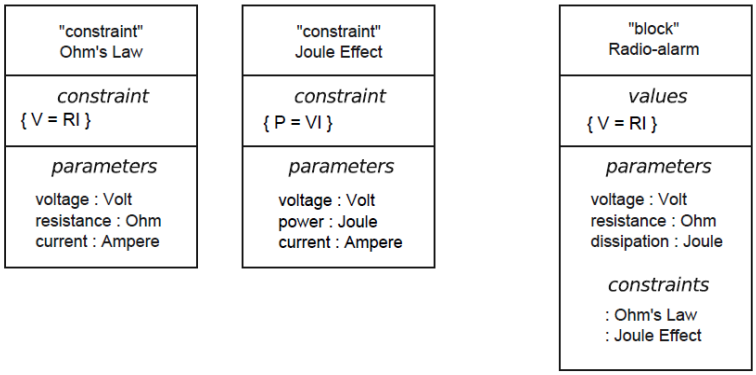

Figure 18 Example of constraints declaration [41]

These constraints are then applied to the parametric diagram, which is a specialized internal block diagram, in order to gather them together and to draw the 66 Embedded Systems connections between them. The constraint properties are represented by rounded rectangles in an internal block diagram. The parameters of the constraint are represented by ports and can be connected to one another. An example is represented in Figure 19 below.

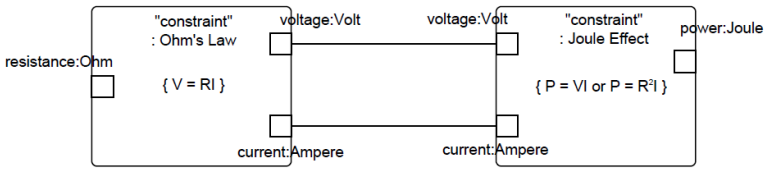

Figure 19 Example of a constraints network [41]

4.2.4. Architecture Analysis & Design Language

The Architecture Analysis & Design Language (AADL) is a SAE International standard [42], that provides a means for the formal specification of hardware and software architecture, which are important for developing safety critical systems like medical control systems.

The first version of the AADL standard was published in 2004 under the leadership of Peter Feiler. At this time, the standard of AADL describes the syntax and semantics of the core-language. After the language became more popular, a variety of annexes has been published in 2006 to the core-language of AADL [43], as for example:

- **Graphical AADL Notation Annex** defines graphical symbols for the modelling of architectures with AADL
- **AADL Meta-model and XML/XMI Interchange Format Annex** defines the abstract representation and interchange format for AADL models
- **Programming Language and Application Programming Interface Annex** defines a mapping of AADL to programming languages
- **UML Profile** for AADL facilitates UML-based tool support for AADL
- Behaviour **Annex** for detailed concurrency behaviour modelling of components
- **Error Model Annex** supports reliability and fault modelling through AADL

The Language

The AADL language offers components with precise semantics to describe system architectures. These components have a type and one or more implementations. A type represents the functional interface of the component; this means what is visible by other components. The implementation describes the contents of the component. Components can be divided into four categories [44]:

- Application software components
- Execution platform components
- Composite components
- Generic components

When working with AADL, in order to create system models, one can describe AADL components, *textually* and *graphically*. Figure 20 below presents the textual and graphical representation of the same "*thread*" component. The *parallelogram* **data_processing** represents the thread. The incoming (**raw_speed**) and outgoing (**speed_out**) **data pots** are represented as *triangles*. The last information of a thread, in this example, is the **period** of the execution which is represented as an *ellipse*.

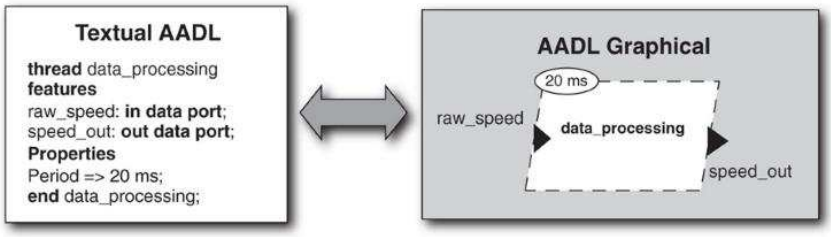

Figure 20 Graphical and textual representation of an AADL component [44]

"A component type specifies a functional interface in terms of *features*, *flow specifications*, and *properties*. It represents a specification of the component against which other components can operate." [42]

Figure 21 below shows a simple cruise control with two devices and one control system. The **flow source** starts at the brake pedal device at the **event data port output**. The flow goes into the control system through the **event data port input** and passed there to the **data port output**. This is called the **path flow** which was defined in the cruise control system. The flow sink ends at the throttle actuator device at the **data port input**.

```
device brake_pedal
features
        break_event: out event data port float_typ;
flow
        Flow1: flow source break_event;
end brake_pedal;

system cruise_control
features
        break_event: in event data port;
        throttle_setting: out data port float_typ;
flow
        break_flow: flow path break_event -> throttle_setting;
end cruise_control;

device throttle_actuator
features
        throttle_setting: in data port float_typ;
flow
        Flow1: flow sink throttle_setting;
end throttle_actuator;
```

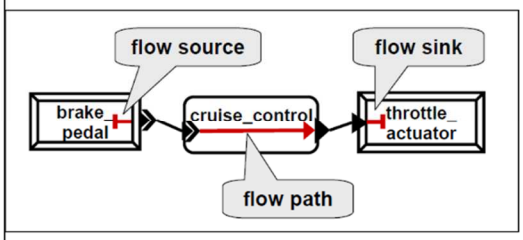

Figure 21 Flow Declarations within a Component Type Declaration [45]

Error-Model-Annex

The purpose of the Error-Model-Annex (EMV2) is to assess the dependability of a system with respect to the qualitative and quantitative aspects. In fault tolerant, safety critical systems error modelling is an important aspect of architectural design and should be integrated into the architecture specification. This Annex extends AADL to support safety and dependability analysis through error models that are attached to architectural components [43].

The EMV2 supports fault modelling at three different ways [46]:

- Modelling of fault sources in a system and their impact on other components or the operational environment through propagation → Error propagation between system components and the environment.
- Modelling of fault occurrences within a component, resulting fault behaviour in terms of failure modes, effects on other components, the effect of incoming propagations on the component, and the ability of the component to recover or be repaired → Component faults, failure modes, and fault handling.
- Focus on compositional abstraction of system error behaviour in terms of its subsystems → It allows for scalable compositional safety analysis.

The EMV2 sublanguage supports the declaration of collections of *error types* and their use in specifying *error propagations*. These error types are associated with interactions point of components to represent incoming and outgoing error propagations with related components. For each component we can also specify an error flow, i.e., whether a component is the source or sink of an error propagation [47].

Error types

The EMV2 provides the ability to declare error types that represent a categorization of faults and errors which can be relevant for a system. An error type can represent a category of fault arising in a certain component. Error types can be organized into different type hierarchies as an example types relating to service errors and/or types relating to timing errors [48]. A small selection of error types is described below[47]:

- *"**Service errors** are errors with respect to the action sequence as a whole rather than individual actions."*
- *"**Value Error** represents any kind of erroneous value."*
- *"**Timing Error** represents a service item being delivered outside its expected time range. This applies to a single communication step or to an end-to-end flow."*

Error propagations
Error propagation occurs when an affected component fails and passes on the error into the system. As an example, a component is the source of the error and transmits it to an adjacent component. AADL and the Error-Model-Annex offer the opportunity to specify several types of errors for each component and their propagation over the bindings. Similar to the flow of a component in the core language of AADL, we can specify the flow of error propagation. The flow has an error source, error sink or error path [47].

Figure 22 below shows how the error flow can be represented within a component by the EMV2. In the example, the **BadValue error** is an *error source* which is distributed the error through the P2 output port. In contrast, the **NoData error** is a *propagation error*. The *error path* is defined as follows: the error occurs at the input port P1 and is distributed in the component C where the error is passed through the output port P2 into the whole system [48].

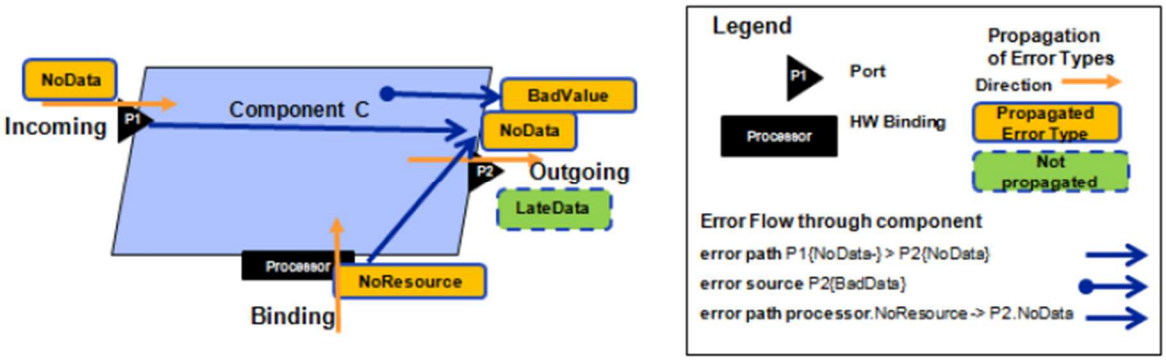

Figure 22 AADL EMV2 error propagation [48]

The advantage of this Annex is to have an accurate view of where risks are and what effect they have on the whole system. Error types and error propagation provide an approach to evaluate the system by means of error analysis and, if necessary, to extend and improve the architecture by suitable fault tolerance mechanisms.

4.2.5. Operation and management approaches supporting dependability design and integration

In production environment, it is not possible to tolerate faults. In these cases, one view is to obtain the best of the following:

- Reliability of command and features.
- Availability of services.
- Maintainability of production tools.
- Personnel and invested capital safety.

These characteristics, known under the general term of DEPENDABILITY, as described in Section 5.1.2, are related to the concept of reliance (to depend upon something). They are quantified in relation to a goal. Then, they are computed in terms of a probability and are obtained by the choice

of an architecture and its components. Finally, they could be verified by suitable tests or by experience. Therefore, because of the IoT being a fully distributed computing network as it is discussed in Sections in 4 and 6, it is important to mitigate any negative effects resulting from faults occurring in its components and to provide sustainable services. Many operation and management approaches supporting dependability design have been proposed or developed. In this section, we will focus on self-organizing software platform (SoSp) which is an IoT platform that allows patients or elderly users to be cared for remotely by their family doctors under normal circumstances or during emergencies. A fault management scheme enabling SoSp to provide situation-aware IoT services without loss of data and state is described in [49]. In fact, the proposed scheme focuses on how to cope with the faults of IoT services. The services would be fault tolerant by substituting the abnormal things with the normal ones with equivalent capabilities in near space as soon as possible. The implementation will be concentrated on how to find the substitutes quickly and to continue providing the services transparently in the IoT environments.

With simple state management of IoT services by the primary service group, the proposed scheme does not require any amount of memory or computing resources. Because the IoT services such as printing in a unit space have a short time from a request to its response, there is not much memory to keep the printing states to handle any faulty situation. After the request is served, the states saved are deleted because they are useless. The primary group can be set up at the time of IoT service deployment and can be activated or deactivated according to the service requests. There is no need to set up a primary service group and to disband it at every request/response.

Another approach which could be mentioned is the N-version programming defined as the independent generation of N >= 2 functionally equivalent programs from the same initial specification. This approach is detailed in [50]. The methodology for implementing N-version programming is relatively simple and can be generalized to other similar applications. In some cases, version programming has been effective in preventing failure due to defects localized in one version of code. N-version programming can be a practical approach if it is selectively applied at subroutine level. On the other hand, there are some negative aspects: Firstly, in the environment of some operating systems, certain implementation defects of a version may cause its associated 3-version program to be aborted by the operating system. Secondly, if missing program functions are the predominant software defects, then N-version programming may not be an effective approach.

4.3. Limitations of existing solutions and relevance to Medolution

Medolution will have to cope with challenges of integration between IoT and Big-Data platforms to build a dependable system. Dependable systems are usually implemented using dedicated embedded computer systems in a defined static configuration of reliable components to support features such as:

- System predictability,
- Real-time guarantees,
- Validate dependability.

Especially in systems with critical functionality, such as medical devices and in particular high risk medical devices, clear system structures are required to support modelling and risk analysis. On the other hand, networked device systems are usually implemented flexible by universal devices such as personal computers, smartphones and fitness sensors. These applications are delivered in dynamic adaptive configurations and of low dependability. As an example, vital values are monitored during fitness training so that influences on the health of the trainees can arise. Against this background, approaches that can contribute to a validation of dependability of applications that integrate flexible universal devices into dynamic configured networks are of particular interest.

5. Devices and IoT Solutions for Healthcare

The big challenge in Medolution is to make sensor data available for physicians or medical devices in a dependable and secure way. This would allow for combining heterogeneous devices and information to a new kind of data source available in the cloud. A lot of research has already been performed to integrate devices or to collect data in the cloud. The following section gives a brief overview of interoperability standards, communication technologies or device related approaches.

5.1. Heterogeneous independent devices integration approaches

According to a report published in 2014 [51], by 2018 an estimated 75.7 million consumer health and sponsored content fitness devices with integrated wireless connectivity will ship, compared to 23 million such devices in 2011. The research companies point out that Bluetooth smart-connected devices are the most popular, but devices that make use of the fitness and health-focused standards also have a foothold. Another important metric says that over the next five years, it is expected that 100 million wearable remote patient monitoring devices will ship. This growth is in part a result of providers who are more aware of the benefits that remote patient monitoring wearable devices can provide to patients outside of the hospital. Because of the growing interest in these devices, there's a bigger opportunity for platforms that collect data from several devices and apps, for example Apple's HealthKit and Google Fit, as it was discussed in Section 3.2.4. on Big Data Applications as a Service.

In today's medical arena, most of the medical devices in the market send their measurements to computer systems through custom, proprietary protocols by implementing the sending and receiving software on both ends. Especially, medical devices which are used inside the hospitals and clinical centres that ship with their managing software and where the medical professionals use the proprietary software tools to interact with the devices and see measured data. While each manufacturer creates and implements its own protocol, it is also the case that the same manufacturer creates different protocols for different device types or different versions of same device types. However, this is starting to change because of the radical increase of personal use of medical devices and patient empowerment in the healthcare process. One example can be seen at [52] where they build their custom protocols for interaction with biomedical devices.

Continua Health Alliance
With an enormous increase of the wearable sensors use in the healthcare sector, companies are getting together to build standards for device integration. Continua Health Alliance is one such major organization dedicating a lot of effort towards this integration. This is a non-profit, open industry coalition of healthcare and technology companies joining together to improve the quality of personal healthcare. With more than two hundred member companies around the world, Continua is dedicated to establishing a system of interoperable personal health solutions with the knowledge that extending those solutions into the home, fosters independence, while empowers individuals and provides the opportunity for truly personalized health and wellness management.

The Continua Health Alliance's Design Guidelines [53] contains references to the standards and specifications that Continua selected for ensuring interoperability of devices. It also contains additional interoperability design guidelines that further clarify these standards and specifications by reducing options in the underlying standard or specification or by adding a feature missing in the underlying standard or specification.

Bluetooth Special Interest Group
The Bluetooth Special Interest Group (SIG) is a privately held, not-for-profit trade association founded in September 1998. The Bluetooth SIG itself does not make, manufacture, or sell Bluetooth

enabled products. The SIG member companies are leaders in the telecommunications, computing, automotive, music, apparel, industrial automation, and network industries. SIG members drive development of Bluetooth wireless technology, and implement and market the technology in their products. The main tasks for the Bluetooth SIG are to publish Bluetooth specifications, administer the qualification program, protect the Bluetooth trademarks and evangelize Bluetooth wireless technology. A Bluetooth profile describes how devices communicating over Bluetooth interact, by specifying the configuration of the channel and the sequence of data exchange needed to establish the channel. It specifies the dependencies on other protocols and profiles, and the manner in which connection is established and configured.

Two specifications highly related to Medolution from this group are the Health Device Profile and Bluetooth Low Energy (also known as Bluetooth Smart) Profiles for each device.

Health Device Profile (HDP) is used to describe how health devices interact over Bluetooth. This profile uses the Multi-Channel Adaptation Protocol (MCAP) to establish communication channels. A control channel is used to establish and manage data channels. The data channels can be paused and restarted with minimal overhead and delay, by retaining the state of the connection before pausing it. This fast reconnection of the data channels allows power saving by allowing the controller to be placed longer in a low power mode. Authentication and encryption of the channels are mandatory. HDP also specifies the L2CAP modes as either Enhanced Retransmission or Streaming. The payload data carried is conformant to IEEE 11073-20601.

Bluetooth Low Energy (LE) Profiles. When the Bluetooth SIG announced the formal adoption of Bluetooth® Core Specification version 4.0, it included the hallmark Bluetooth Smart (low energy) feature. This final step in the adoption process opened the door for qualification of all Bluetooth product types to version 4.0 and higher. Bluetooth Smart (low energy) wireless technology features: ultra-low peak, average and idle mode power consumption, ability to run for years on standard coin-cell batteries, low cost, multi-vendor interoperability, enhanced range. The first specification of Bluetooth low energy wireless technology included two profiles to optimize its functionality for a specific group of products: remote display profile and a sensor profile. On top of these, several profiles have been published for medical device integration with Bluetooth LE.

CEN ISO/IEEE 11073 Health informatics - Medical / health device communication standards
ISO/IEEE 11073 Medical/Health Device Communication Standards [54] are a set of joint ISO, IEEE, and CEN standards for medical device interoperability. In this context, medical devices include primarily personnel, or end user, health devices such as blood glucose monitors, blood pressure monitors, thermometers, pulse oximeters, etc., that patients use in their own homes or other end points to monitor existing medical conditions. The ISO/IEEE 11073 (formerly called IEEE 1073) standards define messaging structures but not the transport layer upon which messages are transmitted. The transport layer messages can be carried through Bluetooth, Zigbee or USB, this standard only defines the data modal in the payload. For each device, there is a specialization of the general standard.

5.1.1. Communication between sensors networks, medical devices and Cloud based systems

Nowadays, many IoT platforms and IoT middleware are hosted on the Cloud. An issue that could be clearly presented is the connection of heterogeneous devices and actuators. This interoperability can be seen as the incompatibilities in terms of data files, semantics, or file sharing protocols and data sets. In addition, making IoT devices interact with existing medical systems, especially inside controlled health institutions, presents big challenges. There is the fact of heterogeneity of devices, protocols, and programming interfaces on one hand, and the requirement to have flexible, scalable deployment and keeping the system easy to configure and to manage, if not self-adjusting, on the other hand. Although some standards have emerged in this area, a common problem is that there

are many vendors who do not support these standards in their products, which increases interoperability issues and system integration costs. Furthermore, medical sensors have become increasingly interconnected with other devices and with computer resources available in the cloud. These resources are configured to receive, store, process, and distribute the information originating from sensor data.

Using a wireless solution on a Cloud storage system helps with connectivity issues and makes it easier to communicate across different information regimes. This remains nonetheless a main research area for addressing the challenges arising from using wireless technologies in medical environments, namely the different types of network communication infrastructures, fault-tolerance, data integrity, low-power consumption, transmission delay, node failure, etc. Connectivity to cloud computing resources can be intermittent, which may require sensors to record measurements in non-volatile memory for uploading at a later time.

An example of a Cloud-based solution is the Electronic Medical Information Exchange (known as eMix) [55]. This system allows physicians and patients to access medical reports from wherever they are. In fact, people can check their medical imaging reports, lab tests, and medical background in a secure distribution system that helps patients to access their records regardless of their location.

Another system that could be mentioned here is the 2net Platform manufactured by Qualcomm Life [56]. This system transfers, stores, and helps convert and display electronic medical device data. It is a Cloud based system designed to be interoperable with different kinds of medical devices and applications [57]. Patients as well as care providers have access to their information around the clock.

An artificial division can be created for the existing interoperability approaches: one side focuses on the technologies, built to work inside the hospitals while the other side is more focused on remote-monitoring, assisted living and patient empowerment outside the hospitals. Although legacy systems inside the hospitals are not much interoperable, latest efforts try to bring together data from different sources/devices to deduce mission critical information and also to build interoperable data acquisition and ingestion platforms collecting and analysing data from different medical devices. Intensive Care Units (ICU) are the leading places inside the hospitals for such platforms. There are a number of approaches which built custom interoperability platforms, but not without adopting any messaging or data model standards [58], [59]. This is what Medolution will try to address by incorporating international standards while realizing interoperability with medical devices.

Sensor communication networks
While early medical sensors had integrated user interfaces for displaying their measurements in an isolated manner, newer generations of sensors acquired the capability of interfacing with external devices using RS 232, USB, and Ethernet. More recently, medical sensors have come to include wireless connection capabilities, both short-range (e.g., Bluetooth, Zigbee, and near-field radios) and long-range (e.g., WiFi, cellular communications). Therefore, theses sensors can communicate wirelessly with nearby computers, PDAs, or smartphones, and/or with cloud computing services.

Persons wearing sensors can go everywhere. They can stay at home, but also travel to any locations. During travelling, and at other locations, the connection to the Cloud can be lost. Networks in hospitals are controlled, but if a patient travels around, network availability cannot be guaranteed. Thus, people can use a XDSL or fiber network at home, or a wireless network when travelling. Wireless communication networks are available in different forms: 3G/4G and even 5G, WiFi, bluetooth, etc. For sensors generating a small amount of data, LoRaWAN™ (Low Power Wide Area Network (LPWAN) [60] can be an alternative.

Reliability is a very important factor in healthcare systems. Therefore, different network communications infrastructures should be considered according to the situation and context, such as adopting services with higher QoS when dealing with high-risk patients.

Devices in Medolution

The following profiles have been considered for adoption in Medolution:

- Device Information Service
- Health Thermometer Service & Profile
- Heart Rate Service & Profile
- Glucose Service & Profile
- Weight Scale Service & Profile.

The following specialization of the general standards are planned to be adopted in Medolution:

- Device specialization – Blood pressure monitor [61]
- Device specialization – Thermometer [62]
- Device specialization – Weighing scale [63]
- Device specialization – Glucose meter [64]

5.2. Middleware platforms to deliver on-demand access to IoT services from multiple infrastructure providers

Over the last decade, developing a middleware platform that is interoperable with multiple infrastructure providers like smart cities, digital agriculture and smart enterprises has been a big challenge. In fact, this middleware should have the ability to connect almost every infrastructure provider and in the same way delivering on-demand access to IoT services such as Big Data resource, storage, rules engine, data processing and data analytics. Moreover, Cloud Computing technologies bring a lot of benefits to enhance the on-demand fact for accessing IoT services. Benefits are mainly related to scalability, high availability, high computing capability and auto scaling. However, legacy IoT solution which is in general based on a single physical programmable board could not offer a higher quality service and an affordable latency for executing analytics to treat enormous feed of heterogeneous data received from multiple sensors. So far, Cloud layer is charging on collecting data from physical sensors or physical gateways hosted on infrastructure providers. Then, a Cloud-based middleware treats this incoming data using on-demand Cloud-based IoT services. Nowadays, we distinguish a lot of middleware platforms that are further divided into two types: proprietary and open-source.

Firstly, there is a large number of proprietary platforms currently offered at the market. For example, we distinguish Amazon IoT platform that was launched at the end of 2015. AWS IoT [65] offers the capability to connect to smart systems that can filter, transform, and act upon separate device data in real-time, based on the business rules defined by the customer. It also offers a capability to route generated data flow or messages to other AWS services such as Kinesis, Dynamo DB, Lambda, S3 and Amazon Machine learning. The drawbacks of AWS IoT platform is that it is more expensive compared to the fact that the customer runs the app in his own local servers [66]. Moreover, there are still some limitations compared to the way that a customer creates an EC2 cluster to handle incoming messages. Additionally, AWS IoT middleware is a newer service that always releases with bare-bones functionality. Consequently, we can consider that this middleware can't cover almost complex data analytics especially in medical healthcare and medical data sets and is designed to the use of AWS IoT services.

Additionally, we can find the IBM Watson IoT Platform [67] that is a fully managed, Cloud-hosted service that makes it simple to derive value from IoT devices. When connected with the IBM Bluemix

platform, Watson IoT middleware provides simple and powerful application access to IoT services and data. It allows for easily and rapidly composing analytics applications, visualization dashboards, and mobile IoT applications. This platform has a set of capabilities built around cognitive computing being better suited to large-scale IoT applications than traditional programmable computing. Cognitive IoT applies Machine Learning to experiences with the environment, to interactions with people and to the data from the devices.

In summary, each proprietary technology is specific to its software vendor who can leverage cross-selling opportunities in terms of other existing platforms.

Secondly, the open source platforms should be mentioned. In fact, OpenIoT project, described in Annex B, proposes a new open source platform focusing on providing the dynamic formulation of self-managed Cloud environments for IoT applications. Therefore, it will serve as a blueprint for non-trivial applications that will be delivered in an autonomic fashion and according to a utility-model based on Cloud infrastructures by means of interoperability and semantic annotation. Indeed, OpenIoT infrastructure provides the ability for composing and delivering IoT services that comprise data from multiple sensors.

Other open-source initiatives are also found for IoT platform implementation, such as Eclipse Kura [68] and OM2M [69] projects, PlatformIO [70], etc. requiring different levels of effort for adaptation to any specific project. In addition, and in a similar but more restricted way, some open platforms exist, where the underlying code is proprietary but they offer an open API for all the tools to connect and operate IoT solutions in a fast way, as for example KAA [71], Compose [72], etc.

5.3. Device control

With the advance of the technology and the diminishing size of electronic components more and more devices from basic sensors and actuators to complex digital equipment and controllers become intelligent and able to communicate with other devices and systems. Such devices and systems have been traditionally the domain of embedded systems developers, but as the number is of devices in a network is growing and the devices are mostly communicating by using IoT-technologies a shift from devices to service-technologies has been performed.

5.3.1. SOA for Management of Devices

There is an increasing demand and opportunity to establish flexible binding between the physical world of sensors and actuators and the software world of IT systems. The pace with which the communication technologies are embedded in more and more types of such devices and connect them to some network, as well as cost reduction of related production and deployment is amazing
To date the integration of devices, sensors and actuators is not just an increasingly important requirement for home networks, it is also required for building more effective business applications, e.g. real-time monitoring of the location of freight in logistics applications, or in building a cost effective health care infrastructure to e.g. remotely monitor vital parameters from patients. However, the device protocols and standards are multiplying at a comparable rate, and IT experts are confronted with evermore increasing number of integration problems, resulting in reduced system stability and performance and increased overall costs. The Service-Oriented Device Architecture (SODA) aims to eliminate much of the complexity and cost by leveraging existing and emerging standards from both the embedded-device and IT domains.

Modelling devices as services
SODA is an adaptation of a SOA, focusing on the boundary layer between the physical and digital realms. In other words, a sensor, such as an ECG monitor, can translate a physical phenomenon into corresponding digital data. Or, an actuator, such as an alarm beacon or exercise bicycle, can

translate a digital signal into a physical phenomenon. Sensors and actuators combine either physically or conceptually to create complex devices and services—such as a health medical exercise module.
In the past years the devices were associated only within a single well-defined system. With the Internet, IT systems can now access signals from numerous devices on an ad-hoc basis. The ability to access and control aspects of the physical realm, which are critical to an enterprise, opens new opportunities and advantages but can be self-limiting. The protocols, connections, and interfaces to devices are extremely diverse, and programming with them is often unfamiliar territory to system and Internet developers. Ancillary device interface software often is as critical to the completion of the IT project, with limited reuse and relatively high maintenance and support costs.

How a service-oriented architecture can help
SODA aims to

- provide higher-level abstractions of the physical realm,
- insulate enterprise system developers from the ever-expanding number of standard device interfaces, and
- bridge the physical and digital realms with a known service or set of services [73].

SODA implementations can use existing and emerging device standards (like UPnP, MDS, Bluetooth etc.) and SOA standards (like JAX-WS, JAX-RPC, WS-RM, WS-Addressing, WS-Eventing, REST etc.). SODA should be straightforward to implement, unlocking the potential to enable a new level of Internet-based enterprise systems.

A device integration developer would be responsible for encapsulating devices as services, dealing with the device-specific connections and protocols as well as with network interfaces needed to publish the data over a standardised SOA protocol. A standard specified device service can have a wide variety of underlying hardware, firmware, software, and networking implementations that don't affect the consumer of the service.

The overall system design would specify the required service interfaces. Suppliers would be responsible for the device adapters and service logic required to provide the specified service for their devices. Other developers could build more complex or composite services from lower-level device services. The system integrator could bid out components from multiple suppliers and avoid maintaining multiple versions of device-specific interfaces in the application code. Enterprise developers could code to a common or even standard set of services. They could not only build this application with device interfaces but also build and integrate future applications and system enhancements reusing these same device services. The enterprise could upgrade device hardware, firmware, and even the lower-level device interfaces with little or no impact on the consuming applications.

5.3.2. The Architecture

Conventional approaches to device integration often focus on custom interface software communicating to enterprise applications through a variety of IT middleware and API technologies (e.g. CORBA, EJB, Jini, Web Services, REST or SOAP). However, SOA, standards, and open software initiatives are moving beyond this middleware architecture. Although IT applications are being adapted to a SOA, standards for defining the low-level device interfaces are still emerging. However, technology exists today to leverage an SOA across the entire spectrum of critical events and data originating from devices.

Mechanisms for building and sharing service interfaces, capabilities for remote software maintenance, and loosely coupled messaging models present highly effective technologies for SODA's implementation. SODA requirements include

- using a device adapter model to encapsulate device-specific programming interfaces;
- employing loosely coupled messaging on the services side, capable of supporting multiple streaming services commonly used in SOA enterprise systems, such as an Enterprise Service Bus;
- using open standards, where available, at the device- and services-interface level;
- providing a means to present standard or open service interfaces to devices that have proprietary protocols or where it might not be practical to drive standards into the low-level device interface;
- supporting the implementation of a spectrum of device adapters—from simple, low-cost sensor data to complex device protocols;
- supporting loading of remotely configurable logic components in device adapters for maintenance, upgrade, and extended functionality; and
- adapting security mechanisms as required for the domain.

A SODA implementation comprises of three main components (See Figure 23 below). Device adapters talk to device interfaces, protocols, and connections on one side and present an abstract services model of the device on the other. The bus adapter moves device data over network protocols by mapping a device service's abstract model to the specific SOA binding mechanism used by the enterprise. The device service registry provides for the discovery and access of SODA services.

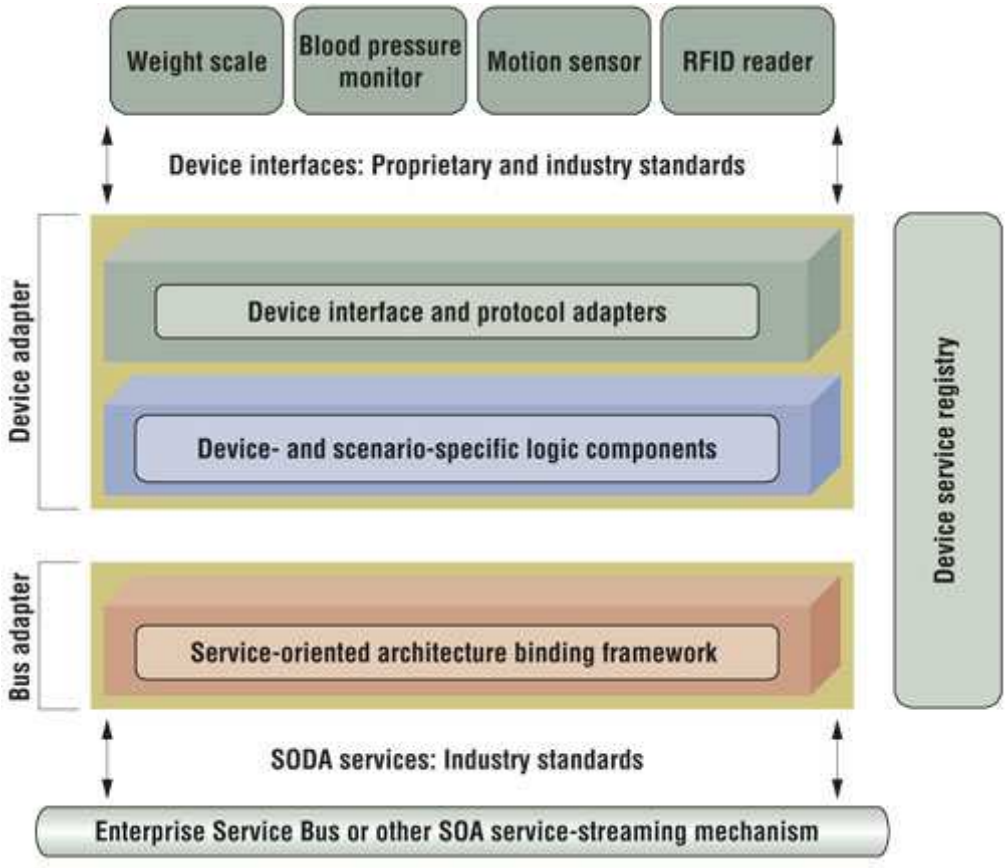

Figure 23 Service-Oriented Device Architecture Model [74]

While there is still no standard agreement, device interface and protocol adapters within SODA implementations provide a common model of devices to the software used to create service

interfaces. Forthcoming widespread adoption of standards at the device-interface level will reduce the development and maintenance costs of device adapters and their corresponding SODA services. However, standards at the services layer can provide the largest leverage for both the device and enterprise markets.

Rapid standardization of device services, device-services transport mechanisms, and tools will let device manufacturers develop their interfaces and provide SODA services to the enterprise, shifting development responsibility for device adapters and services to the appropriate point in the supply chain rather than forcing enterprise developers to deal with thousands of APIs. For this adoption to take place, the SODA model must evolve with open and accessible standards, which must cover the specific services used within and across enterprises. Reducing the barriers to acceptance requires that the standards be open and part of a community and that samples, examples, frameworks, and tooling be made available through reference implementations.

5.4. Limitations of existing solutions and relevance for Medolution

The technologies described in the Chapter 5 show that there is a trend to harmonize the interoperability between data suppliers (sensors) and service suppliers (e.g. Google Fit, Apple Healthkit), so that these environments can work together with many types of sensors. However, the interoperability between these toolkits seems to be very limited as the current publications suggest. This would also diminish the interoperability options with the Medolution big dependable system that is envisioned.

In the area of IoT, emerging frameworks and standards build an important foundation for the development of dependable systems and devices; however, the integration of such frameworks into a development process directly addressing the dependability requirements is missing.

For Medolution, patients can be at any location, not only within the walls of a hospital. Network connectivity cannot be guaranteed at any patient's location. For example, a wireless connection can be of bad quality when a patient enters a shopping mall. Solutions shall be sought to optimize the available Internet connection bandwidth under all circumstances.

6. Automated Technical Management for Healthcare Systems

This chapter introduces the state-of-the-art approaches to the field of the automated technical management. In particular, it presents the fundamentals of the automated technical management and covers its main functional areas. The well-established management infrastructure as well as the corresponding OSI/ISO standards are explained and figuratively represented. The chapter continues with the widely accepted administrative approach to the automated technical management, the policy-based management, explaining its paradigm and presenting its main structure. Further, the use of models is addressed, which enriches the management approaches and has proved to be successful during the last decades. The model-based management as a paradigm is introduced. In order to demonstrate the wide acceptance of the approaches in the field of automated management of the medical devices and systems, several recent research projects are also briefly presented in the Sections 13-16 of the Appendix B.

6.1. System Management Solutions and BDHS

The sensitivity of the medical domain implies that deployed devices and systems shall fulfill advanced and strict requirements of the application domain. As it is discussed in Appendix A, the scope and content of laws, regulations and norms describing these requirements is very wide. The following specific issues are addressed most often:

- **Safeguard clause**: Patient's as well as other persons' clinical condition and safety are not be compromised. Associated risks are to be eliminated or reduced as far as possible. If necessary, adequate protection measures should be taken, these can include alarms, notifications, and warnings in any form. In case any shortcomings of the protection measures exist, users must be informed of the residual risks.
- **Patient orientation**: The aim to primary treat the patient, not the disease demands primarily for patient-centered approaches. This concerns the whole complex of healthcare provision as well as its singular sectors, fields of actions, measures, and instrumentals, like medication, interventions, devices, etc. Thus, ability to adapt to the dedicated patient's needs is inevitable.
- **Standard conformity**: Modern healthcare delivery supposes strict adherence to medical standards, regulations, laws, and other norms. The same applies to medical sensors, devices as well as medical services. In order to enable their interoperability at the highest level, it should be resorted to existing technical standards. Among other things, they address personal health information exchange issues.
- **Information security**: One of the main principles in medical ethics concerns patient's confidentiality. According to it any information revealed by a patient to a healthcare provider is strictly private; unless the patient gives consent to disclose it to a third party or it can be justified by law.
- **Accountability**: Accountability supposes obligation of the parties to justify and take responsibilities for their activities. With respect to the medical application domain, accountability entails processes and procedures which provide for professional competence, legal and ethical conduct, financial performance, adequacy of access, public health promotion, and community benefits.
- **Quality Assurance**: To attain the highest performance and safety level is essential for the healthcare services. Thus, comprehensive quality management systems for medical devices and systems are important. Market entering for soft- and hardware providers implies a set of regulatory procedures which refer to the quality assurance also. The manufacturers

are obliged to demonstrate that their product does what it is supposed to do and is able to demonstrably meet the medical claim.

The above listed application domain specific requirements establish a basis for the concrete technical requirements of a deployed Big Dependable Health System. In order to meet the requirements, comprehensive system management solutions are needed. Thereby, tailoring a management solution for the particular use case is not enough. Advanced architecture patterns and elements are targeted which go beyond the traditional device and system management and address complex specifics and aspects of the medical domain.

6.2. Fundamentals of Automated Technical Management

In order to provide a review of the research field, we need to introduce the main paradigm and fundamentals of the complex of management concept. According to [75],
"The management of networked systems comprises all the measures necessary to ensure the effective and efficient operation of a system and its resources pursuant to an organization's goal".

Ensuring operation implies equipping the systems with configuring, reconfiguring, tuning, protecting, and recovering capabilities. At the same time, the complexity of these tasks is to be hidden from users and administrators. Essentially, the management performs an intelligent control loop: automated methods collect the needed details from the system; these details are analysed in order to determine if something needs to be changed; a plan and/or sequence of actions which specify the required changes is worked out; and finally the plan is executed. A common knowledge in form of management information is a basis for the whole management process.

6.2.1. Management Functional Areas

The ISO/ITU-T joint committee has worked out a Recommendation ITU-T X.700 [76] which turned into a widely accepted standard for network and systems management. According to it, the OSI Management Architecture defines in its functional model the following five functional areas:

- **Fault:** Fault management involves reactive and proactive measures in order to detect, isolate, eliminate as well as prevent faults in the behavior of a system.
- **Configuration**: Configuration management focuses upon ways and mechanisms to identify, control, collect data from and provide data to the system. It implies preparing, initializing, starting, providing for the continuous operation of as well as terminating services of the system.
- **Accounting:** Accounting management provides a way to identify costs arising from the usage of system's resources and to establish charges for it.
- **Performance**: Performance management defines quality of services provided by the system. It enforces required measures in order to guarantee and provide evidence that the quality of service complies with the contracted agreements.
- **Security**: Security management evolves all sorts of measures and mechanisms in order to guarantee and support application of defined security targets. Operating on different abstraction levels, it defines and controls security services and mechanisms, distributes and stores security-relevant information as well as reports and warns on security-relevant events.

6.2.2. Management Infrastructure

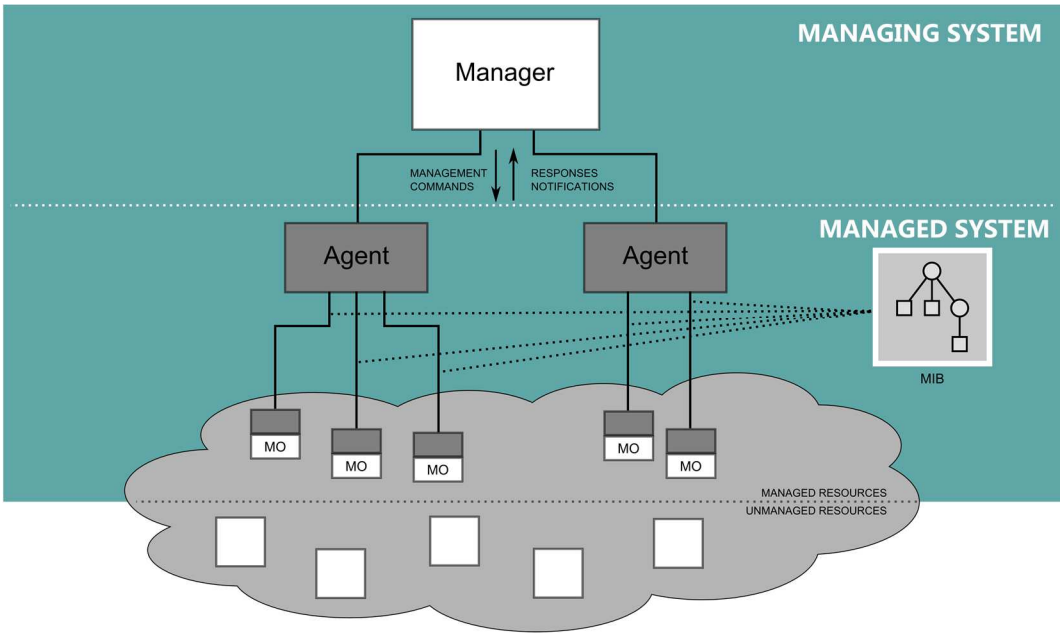

Figure 24 Management infrastructure

The management infrastructure is depicted in Figure 24 above, which distinguishes between the managing and the managed system. The managing system hosts managing processes, managers, while the managed system is resided by management agents. A manager is the part of the management process that takes decisions based on collected management information. The manager monitors and configures the managed system by communicating with the management agent residing there. Managers can be arranged in a hierarchical structure when required by the system structure.

Thus, in context of OSI/ISO Network management, objects in the Management Information Base (MIB) are defined using a subset of Abstract Syntax Notation One (ASN.1) called "Structure of Management Information Version 2 (SMIv2)" RFC 2578 [77].

A large-scaled distributed system may contain hundreds of managed objects. In order to make management practically feasible, the concept of a management domain is introduced. Management domains provide the means of partitioning management responsibility by grouping objects accordingly to common characteristics. Common management operations or actions can be applied to designated sets of management objects by providing domain-specific management rules. Thus, the policy concept is brought in [78], [79]. The following section addresses the use of policies in order to facilitate the automated technical management.

6.2.3. Policy-based Management

Policy-based management is a widely accepted administrative approach to the automated technical management. Policies represent logic that determines the behavior of the managed system. Operations and actions on resources are conducted according to predefined rules that express this desired behavior. In order to support system reactivity or even proactivity, constant changes in requirements and conditions at runtime are to be considered and the system operation is to be adapted to them. Thus, policies allow a new level of autonomic computing to arise. Thereby the variety of policy-based management approaches is extremely great starting from deeply tied to the application domain and dependent on the field of operation to those completely generic and suitable to be applied in multiple operation environments.

The Internet Engineering Task Force (IETF) Policy Framework Working Group and the Distributed Management Task Force (DMTF) have come up with a policy to represent, manage, share, and enforce policies in a vendor-independent, interoperable, scalable manner [80]–[82]. The basic elements of the framework are presented in Figure 25 below: the policy management tool, the policy repository, the policy decision point (PDP), and the policy enforcement point (PEP).

The **policy management tool** is a framework component used by the administrator to define, generate, and manage policies that are to be applied to the system. The policy management tool is not standardized so far. Generally, it should support such important aspects like centralization and business-level abstraction. Thus, it is supposed to allow a single point of configuration and provisioning for the system components, which simplifies the work of the administrator, who can specify the required policies and provisioning details altogether. It also supports the management of large-scaled systems, which can comprise a great number of policies and configurations at a single point for the sake of overview and conflict avoidance. Business-level abstraction caters for the convenience of the administrator by means of using domain-specific abstractions and language for system and policy definitions.

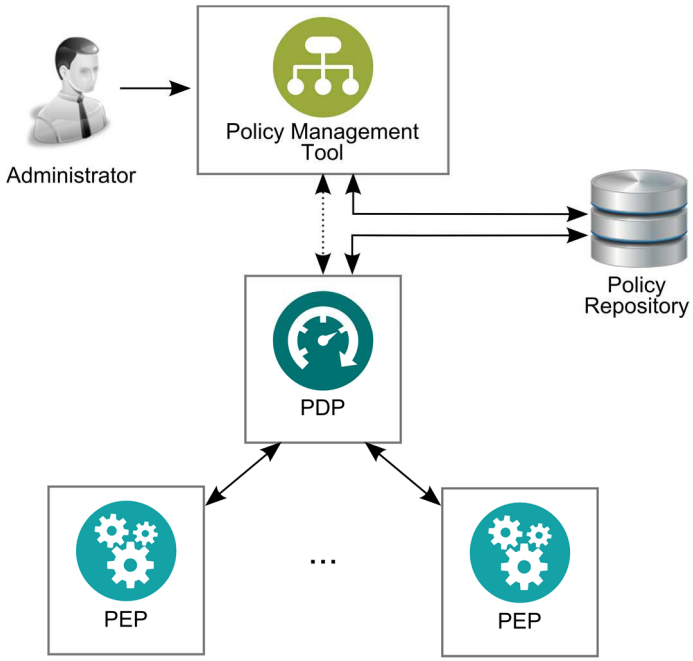

Figure 25 IETF Policy Management Framework

The generated policies are stored in the **policy repository**. For the sake of interoperability, the format of the stored policies has been specified within the Policy Core Information Model (PCIM) (RFC3060) [81]. According to it, policies are represented as a set of policy rules, which in turn consist of a set of conditions and actions. A Boolean logic is assumed: if condition clause is true, the actions clause is to be executed. Policy rules are reusable and may be prioritized. A time schedule indicating activity time periods can be associated with a policy rule. Moreover, policy rules can be aggregated into groups (optionally nested) enabling hierarchical representation. The PCIM can be refined in accordance with the application domain in order to define a domain-specific policy information model, like e.g. in [83]–[85].

PDP is a logical entity that interprets policies, makes policy decisions, and communicates them to the PEPs [82]. The actual policy enforcement is a task of the **PEP**s. They are responsible for starting the interaction between the managed and the management system. On a predefined event, PEP generates a request for a policy decision and sends it to the PDP. As soon as the PDP returns the policy decision, the PEP enforces it while executing the policy actions on the corresponding management objects.

The framework covers the whole lifecycle of the automated management starting from system planning and initial configuring (policy management tool) at the design phase to monitoring and control functions which ensure the dependable system behavior enforcing the predefined policy rules (PDP/PEP) at runtime phase. In this context the notion of system model appears. During the design phase a detailed system model is elaborated which serves as a basis for defining policies for the management system. Thus, model-based management paradigm is of interest.

6.2.4. Model-based Management

Using a system model in order to support technical management of devices and systems has proven to be successful during the last years. For instance, in [86] the authors present a model-based approach to network management. A reactive self-configuring model-based hybrid hard- and software system is presented in [87], whereas [88], [89] address model-based management of services-oriented systems.

The model is an abstract formal representation of a system [90]. A given system may have plenty of different models. Each of them represents a particular aspect of the system and only this aspect. Each model has a specific purpose and is described in the language of its unique metamodel. The metamodel defines how elements of a system are to be chosen in order to generate a given model. Thus, a metamodel, which the model is conformant to, specifies what aspect of the system the model represents (See Figure 26 below).

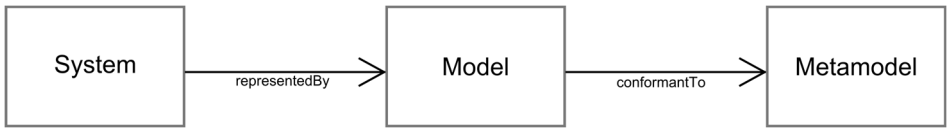

Figure 26 Correspondence between a System, Model and a Metamodel [90]

In November 2000, the Object Management Group (OMG) introduced the Model Driven Architecture (MDA™), adhering to the global trend of and realizing Model Driven Engineering (MDE) principles [91]. Based on the established standards like MOF™ (Meta Object Facility™), XMI® (XML Metadata Interchange™), OCL™ (Open Constraint Language™), UML® (Unified Modeling Language™), CWM™ (Common Warehouse Metamodel™), SPEM™ (Software & Systems Process Engineering Metamodel™), the MDA separates business and application logic from the underlying platform technologies by providing platform-independent models and leverages them to “enhance the agility of planning, design, and other lifecycle processes, and improve the quality and maintainability of the resulting products”.

In the field of automated technical management, the usage of models brings similar benefits and gains in importance. Several steps towards standardization have been taken lately. Thus, the DMTF’s Common Information Model (CIM) provides a common definition of management information for systems, networks, applications and services, and allows vendor- and domain-specific extensions. It is an information model, a conceptual view of the managed environment, which unifies and extends the existing instrumentation and management standards using object-oriented constructs and design. The CIM standard includes the CIM Metamodel [92], the CIM Schema [93] and a set of relevant specifications [94].

Relying on the OMG’s UML specification [93], the CIM Metamodel is the basis on which CIM schemas are defined. It defines the semantics for the construction of new conformant models and comprises common basic elements for representing models (e.g. object classes, properties, methods and associations) [92]. The actual models are described by the CIM schemas representing the resources of a managed system, including their attributes, behaviors, and relationships. The

CIM schema is structured into the distinct layers: core model (applying to all areas of management), common model (applying to the common areas like systems, applications, networks, and devices but independent of a particular technology or implementation), and extension schemas (technology-specific extensions to the common model) [95]. The CIM specifications define the management infrastructure, the details for integration with other management models, the syntax, semantics, naming conventions [96] as well as the use of the Managed Object Format (MOF) language [97] for specifying CIM models.

DMTF's Web-Based Enterprise Management (WBEM) comprises a set of specifications that cover discovery, access, and manipulation of resources modeled using the CIM [98].

6.3. Model-based Management of Medical Systems

Usage of models in order to support medical systems varies for each application area. For example, models are suitable and can be used for supporting the workflow management system in health care [99], [100]. In [100] a model-supported process management of medical systems is presented. The authors claim that now the main tool "to reflect the arrangements of the clinical pathways and to support and standardize the decision-making of the physician as well as the (planning of treatment)" is the hospital information system (HIS). Thus, they propose that covering these tasks, in particular automated support at the organizational level, should be done by a dedicated management system. The long-term quality of care and therefore the patient satisfaction are to be achieved by integrating elaborately modelled and planned patient's pathway models. The management process involves modelling, planning and execution management phases. The modelling phase covers the basic work on the analysis and picturing the treatment processes. The organization specific parameters are considered during the planning phase, whereas instantiation of pathways models for individual patients is done during the execution phase. Moreover, a management cockpit is supposed to give an opportunity to query the pathway instances in real-time as well as to keep track of the patient's individual ways hereinafter (See Figure 27 below).

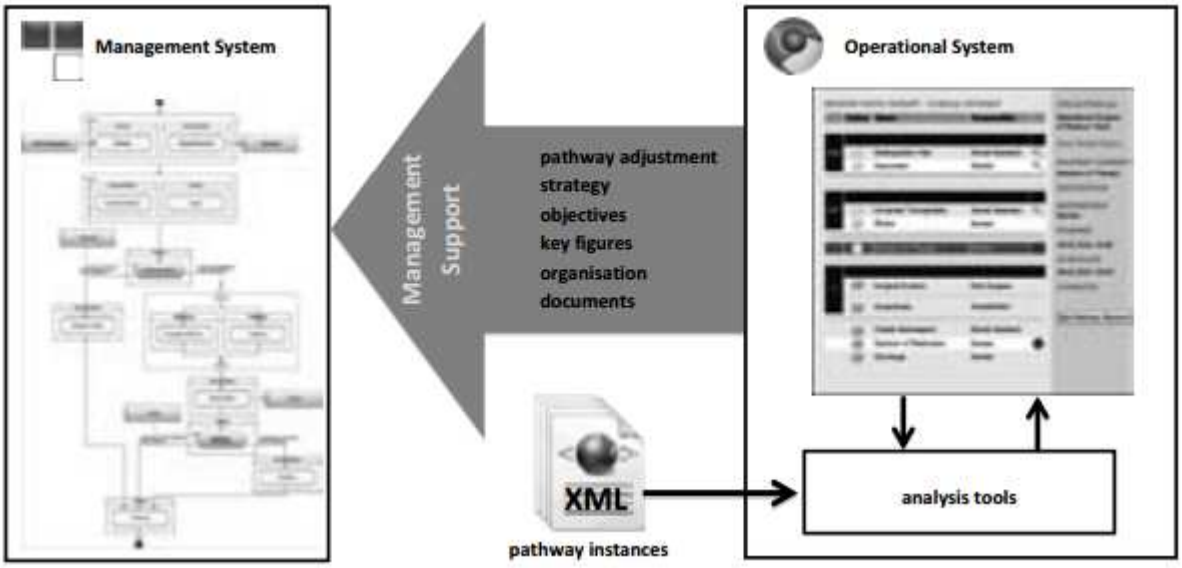

Figure 27 Model-supported Treatment Workflow Management [100]

Another example of application area is model-based medical decision support for diagnosis and/or prognosis assisting the medical staff in their work [101]. Models can also be used for engineering of medical systems [102]–[104] providing a basis for fast prototyping, testing, safety verification.

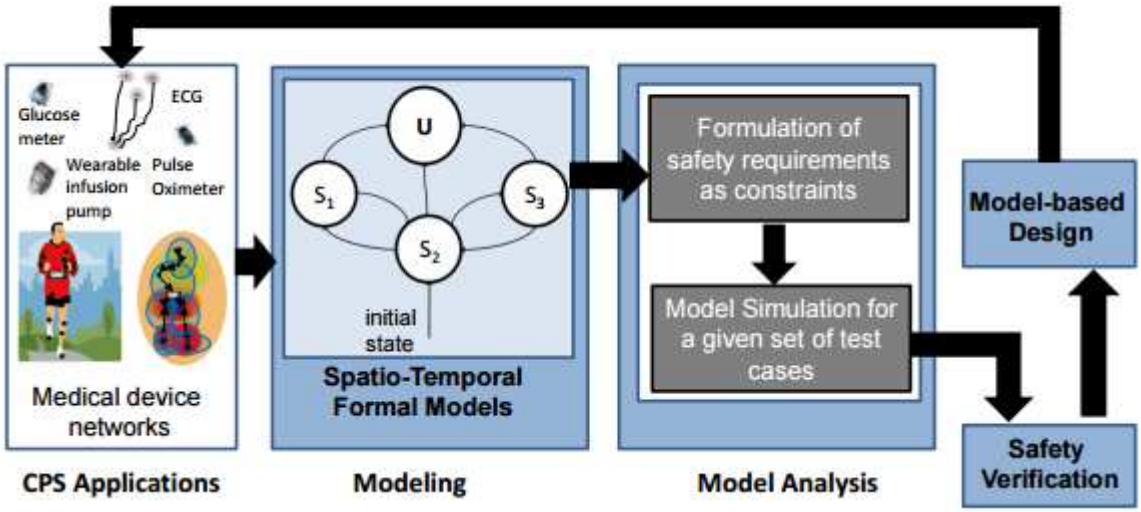

Figure 28 Modelling to Support Engineering of Medical Systems [103]

In [103], an approach to design cyber-physical medical systems (CPMSes) has been elaborated (See Figure 28). The proposed framework uses abstract models that "consider the operation of the CPMS as definite steps in an algorithm and simulate them as a state machine." The models are then analysed considering given current operating conditions, so that system properties represented as model parameters could be obtained. Subsequently, the model parameters and the safety requirements, represented as constraints, are compared with each other in order to support the safety verification. The authors claim that CPMSes cover discrete models of the computing systems and continuous dynamical models of the physical environment that exchange data. Hence, both discrete and continuous elements are used for modelling of CPMSes.

In [102] the authors report on dramatically increased verification and validation efforts in the field of medical systems over the last years. Within German national SPES2020 project [105] run from 2009 to 2012, the seamless model-based development of safety-critical systems was researched in order to provide for validation and verification of requirements, simulation, verification, as well as virtual integration testing. The proposed SPES Modelling Framework (See Figure 29 below) adheres strictly to the principles of stakeholder concerns, hierarchical decomposition, seamless model-based engineering, separation between problem and solution as well as logical and technical solution, and consideration of crosscutting system properties during the development process. Two fundamental concepts of **viewpoints** and **abstraction layers** forming a two-dimensional design space are promoted. A viewpoint in follows the notion of the IEEE Standard 1471 "Recommended Practice for Architectural Description of Software-Intensive Systems" [106] and is regarded as s a template or pattern for the development of individual views on the system (Requirements Viewpoint, Functional Viewpoint, Logical Viewpoint, and Technical Viewpoint). Abstraction Layers are user defined, i.e. application domain specific ("Supersystem", "System", "Subsystem", and "Hardware/-Software Component"). As a proof of concept a study case from the healthcare domain has been described demonstrating the application of the approach to engineering of an extended care system comprising body area network devices, a VAD, and a telematics system.

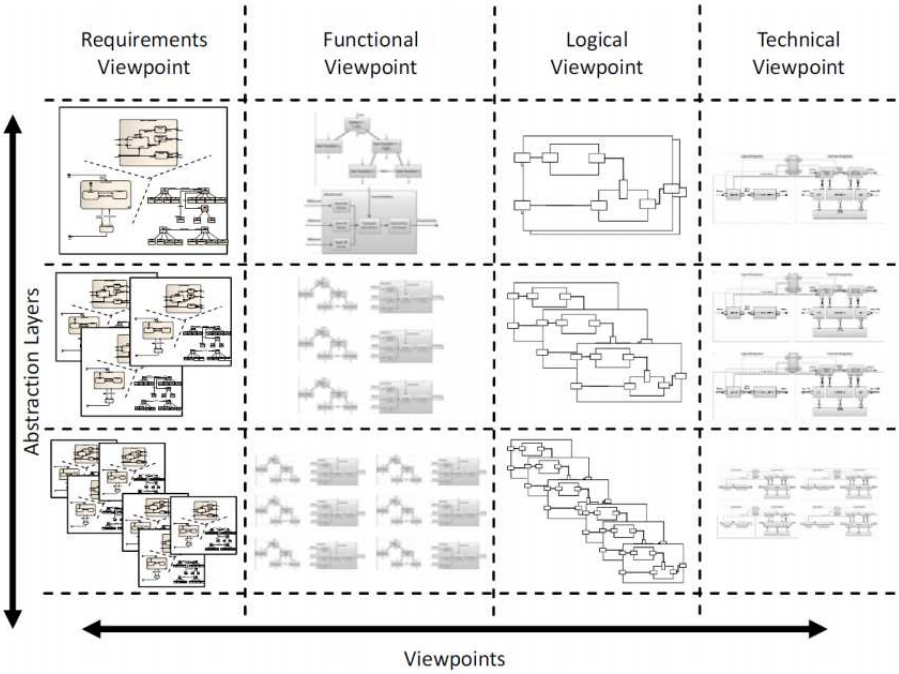

Figure 29 SPES Modelling Framework [102]

In [107], the model-based design is claimed to support the development of such critical systems like medical. Its well-founded methodology is reported to provide a solid basis for the tool support.

6.4. Limitations of existing solutions and relevance for Medolution

Automated technical management is a widely accepted solution, which aims to enrich and support applications and systems. Within the Medolution project it is planned to resort to the well-established management paradigms which incorporate policy and model support.

Particularly, the usage of policies for management of medical systems has proved to be a feasible approach in order to provide for an adaptable and flexible autonomous behavior. The usage of policies in the reviewed projects, however, goes very often far beyond the management's scope of responsibility. Sometimes they are tightly embedded into the application logic and are used to steer the application workflow. This can indeed turn out to be a limitation for Medolution, since it aims for software engineering procedures for certification support. The usage of policy definition languages (like Ponder [108], Ponder2 [109], APPEL [110], etc.) requires interpretation at runtime as well as affords additional complexity. The similar issue concerns the usage of ontologies in order to provide a common definition of the types, properties, and interrelationships of the system entities (like FIPA [111], CC/PP [112]).

Though models have been used in order to support medical devices and applications, we are not aware of any works on model-supported technical management solutions for medicals systems. The introduced approaches to the model-supported process management of medical systems veers toward application workflow steering. Taking into consideration the sensitivity of the application domain as well as a need for further certification support this turns out to be a considerable limitation of the solution.

7. Healthcare data exploitation

This chapter deals with the exchange, analysis and application of healthcare data. One of the main challenges in Medolution is to turn a vast amount of data coming from a variety of sources into consolidated, usable information for medical professionals, but also usable for the patient. This requires:

- healthcare data integration: integration of the data captured by sensors, imaging devices, other medical devices, EHR systems.
- healthcare analytics: the development of analytics components to explore and enrich health care data.
- healthcare decision support: to combine and interpret the captured data by applying rules defined and approved by healthcare professionals.
- interactive user interfaces for the medical professionals and patients.

7.1. Healthcare Data Integration

Data integration is about how to combine data from a large variety of heterogeneous sources into meaningful and valuable information. Data from different systems need to be integrated technically and semantically.
To achieve semantic interoperability in the healthcare domain numerous standardization efforts are in place in order to define common information models or common data elements such that all systems can operate with data on the same knowledge level.

7.1.1. Integrated approaches for heterogeneous sources of healthcare information

Many standardization efforts focus on Electronic Health Records (EHRs) in order to facilitate integration of electronic health data accumulating in healthcare facilities (hospitals, clinics, regional data warehouses etc.). The most important initiatives are:

- openEHR – an open standard for health data based on a complete separation between software and clinical models, thus ensuring universal interoperability [113].
- Health Level Seven (HL7) – a set of standards for transfer of clinical and administrative data between software applications [114].
- Integrating the Healthcare Enterprise (IHE) – an initiative by healthcare professionals and industry to improve the way computer systems in healthcare share information. IHE promotes the coordinated use of established standards such as DICOM and HL7 to address specific clinical needs in support of optimal patient care [115].

The following Standard Development Organizations (SDO) are active in the field of healthcare data integration:

- ISO/TC 215 [116]. The ISO/TC 215 is the International Organization for Standardization's (ISO) Technical Committee (TC) on Health informatics. TC 125 works on the standardization of Health Information and Communications Technology (ICT), to allow for compatibility and interoperability between independent systems. E.g. ISO 13606.
- Health Informatics committee of the European comity for standardization (CEN/TC 251) [117].
- American Society for Testing and Materials (ASTM) E31, the technical committee responsible for development and maintenance of the Continuity of Care Record (CCR) standard [118].

7.1.1.1. CEN/ISO 13606 EHR Communication

CEN TC 251 is the preeminent healthcare information technology standards developing organization in Europe. They issued the important and relevant standard CEN/ISO 13606 [119] designed to achieve semantic interoperability in the electronic health record communication. This European standard also has been approved as an international ISO standard. The five part standard defines the logical models and interfaces required to support the generic communication of EHR data and archetypes between heterogeneous EHR systems:

- Part 1 – Reference model: The content of any EHR system can be mapped onto this Reference Model. In addition, it specifies functionality such as: attestation, versioning, audit trail, signatures.
- Part 2 – Archetypes interchange specification: Defines the Archetype Object Model and suggests the Archetype Description Language (ADL 1.4)
- Part 3 – Reference archetypes and term lists: Lists State models, and needed vocabularies. In addition, it specifies mappings to openEHR specific constructs
- Part 4 – Security requirements and distribution rules: a profile of the ISO 22600 Privilege and Access Control standard that allows the specification of Access Control Lists (the Patient Mandate) as a profile
- Part 5 – Messages for exchange: an interface specification for EHR-Extracts that are exchanged between systems

7.1.1.2. HL7 CDA, CCD and FHIR:

The most important HL7 standards are:

- Messaging Standards - interoperability specifications for health and medical transactions
- HL7 Clinical Document Architecture (CDA)
- HL7 Continuity of Care Document (CCD)
- HL7 Fast Healthcare Interoperability Resources (FHIR)

The HL7 Clinical Document Architecture, previously called Patient Record Architecture (PRA), defines the structure and semantics of medical documents for the purpose of exchange [120]. CDA documents are encoded in Extensible Markup Language (XML). They derive their meaning from the HL7 Reference Information Model (RIM) and use the HL7 Version 3 Data Types, which are part of the HL7 RIM.

Many national and international pilot projects use HL7 CDA Release One as a format for clinical documents [121]. Commercial products implementing CDA are also starting to become available. Since medical documents are currently mainly stored in clinical information systems that already use the HL7 standard, vendors have experience with HL7 and are likely to be aware of the opportunities offered by CDA. Strictly speaking, the HL7 Clinical Document Architecture (CDA) is not an EHR standard since it only defines parts of an EHR architecture. However, the CDA forms an important component of an EHR and is currently being harmonized with the equivalent structures in EN 13606 and openEHR.

The HL7 Continuity of Care Document (CCD) [122] specification restricts the Clinical Document Architecture Release 2 (CDA) to meet the requirements of the ASTM E2369-05 Standard Specification for Continuity of Care Record (CCR). This restriction therefore supports a CDA representation and transformation of a CCR document by using CDA templates for transformation specification.
The Continuity of Care Record (CCR) was developed to cover information regarding pertinent clinical, demographic and administrative data for a specific patient as a snapshot in time. The CCR

allows a healthcare practitioner or systems to collect and aggregate data about a patient and to forward it to other healthcare practitioners to support continuity of care.

HL7 CDA is a document mark-up language that specifies the structure and semantics of clinical data for exchange purposes. From a CDA point of view, CCR is a standardized data set (template) by which CDA can be restricted to specify CDA for summary documents. Therefore, CCD is an alternate implementation of the ASTM ADJE2369 Standard of CCR using CDA syntax and format of the proprietary CCR format.

Recently HL7 introduced Fast Healthcare Interoperability Resources (FHIR) - a draft standard for the exchange of resources through restful interfaces. FHIR defines a set of "Resources" that represent granular clinical concepts, such as Observation, Encounter, Condition. The resources can be managed in isolation, or aggregated into complex documents. Resource description are based on simple XML or JSON structures. FHIR leverages these resources models to provide a consistent, easy to implement, and rigorous mechanism for exchanging data between healthcare applications, through http-based RESTful protocol where each resource has predictable URL.

7.1.1.3. IHE Patient Care Coordination (PCC) Templates:

Since broad-based, scalable computable semantic interoperability across multiple domains requires the integration of multiple standards, the international initiative Integrating the Healthcare Enterprise (IHE) plays the key role of “integration organization” involving multiple stakeholders (including both vendor and provider organizations). IHE promotes the coordinated use of established standards such as DICOM (Digital Imaging and Communications in Medicine) and HL7 to address specific clinical need in support of optimal patient care. For that purpose, IHE develops integration profiles that provide precise definitions of how standards can be implemented to meet specific clinical needs.

IHE focuses on several clinical domains such as cardiology, radiology, laboratory, pharmacy, IT infrastructure, patient care coordination, etc. Within the scope of the Patient Care Coordination (PCC), IHE has defined some EHR content templates as well. In general, the EHR/PHR content templates are built on top of the well-accepted content standards such as HL7 CDA and CEN 13606 to further refine these standards by: (i) restricting the alternative hierarchical structures to be used within the instances, (ii) constraining optionality and cardinality of some elements, (iii) defining the code systems and codes used to classify parts of the document and also (iv) describing the specific data elements that are included.

7.1.2. Semantic interoperability approaches for health data integration

Semantic interoperability can be defined as the ability of two or more computer systems to exchange information in such a way that the “meaning” of that information can be automatically interpreted by the receiving system accurately enough to produce useful results to the end users of both systems [123].

7.1.2.1. Semantic Interoperability Paradigms

There are two main paradigms for the semantic interoperability between EHR-systems: Messaging Paradigm and the Two-Level-Modelling Paradigm. The SemanticHealthNet EU-project [124] addressed this problem quite deeply by involving the major stakeholders to solve this problem, such as the EN13606 Association, IHTSDO and the WHO.

- In the messaging paradigm a maximal data set is defined to be used in a generic context using a use case. The structure of the message and all its needed vocabularies is agreed upon in a consensus process. IT-vendors produce software that implement the standardized

message and use for this purpose most often a profile defined by IHE. The IHE-profiles are mostly based on HL7 message standards. In an IHE Connectathon various vendors test their applications that implemented the specific IHE-profile against each other and publish the results. Any change in the data set or any other part of the profile needs a new IHE-process including the Connectathon.

- The two-level-modelling paradigm is based on the production of a library of Archetypes. Archetypes are data/information components defined as constraints on a Reference Model. This particular model is a generic model that specifies how data/information is documented, archived, versioned, attested, etc. It is a generic model of any EHR. CEN/ISO 13606 EHR is an open international standard that is based on this Two-Level-Modelling Paradigm.

There are many efforts adopting the messaging paradigm for semantic interoperability of healthcare data. The efforts which try to facilitate the exchange of EHRs for better care of the have been developing Common Data Element (CDE) models. A few examples can be summarized as follows:

- Health Information Technology Standards Panel (HITSP) has defined the C154: Data Dictionary Component [125] as a library of the HITSP defined data elements to facilitate the consistent use of these data elements across various HITSP selected standards. These data elements are served through PDF documents and spreadsheets. For example, HITSP C32 [126] which describes the HL7/ASTM Continuity of Care Document (CCD) [122] content for the purpose of health information exchange, marks the elements in CCD document with the corresponding HITSP C154 data elements to establish common understanding of the meaning of the CCD elements.
- The Federal Health Information Model (FHIM) [127] develops a common computationally independent model for EHRs.
- The Transitions of Care Initiative (ToC) [128] maintains the S&I Clinical Element Data Dictionary (CEDD) [129] as a repository of data elements to improve the electronic exchange of core clinical information among authorized entities in support of meaningful use and improvement in the quality of care. The Query Health [130] initiative extends this data dictionary, and establishes Query Health CEDD to enable an architecture for querying distributed EHRs in order to aggregate healthcare data for collecting quality measures and monitoring disease outbreaks.
- The Clinical Data Interchange Standards Consortium (CDISC) [131] provides common dataset definitions in (a) Study Data Tabulation Model (SDTM) [132] for enabling the submission of the result datasets of regulated clinical research studies to the FDA and in (b) Clinical Data Acquisition Standards Harmonization (CDASH) [133] for integrating SDTM data requirements into the Case Report Forms.
- The Biomedical Research Integrated Domain Group (BRIDG) [134] developed the Domain Analysis Model (DAM), which harmonizes CDISC data standards with the HL7 Reference Information Model (RIM) [135]. The BRIDG model unifies the concepts in the clinical care and research domains and creates a shared generic representation for each data element.
- Mini-Sentinel [136] is a pilot project to create an active surveillance system to monitor the safety of FDA-regulated medical products by accessing pre-existing electronic healthcare records. It proposes a Common Data Model (CDM) so that analytic applications can run on a uniform model. This model is maintained in a PDF document and partner EHR Systems are expected to translate the EHR data to this common model.

There are other similar efforts to define CDEs and accompanying data models like Observational Medical Outcomes Project (OMOP) [137] and I2B2 [138] data model.

7.1.2.2. Semantic Interoperability based on Standard Terminologies and Ontologies

Medical terminologies are used simply to ensure standardization of data entry and interoperability of healthcare information systems. Some advocate the use of standardized medical terminology and associated data formats by healthcare practitioners during the initial recording of data. However, many clinicians share the view that faithful recording of patient data can only be achieved by using natural language. The best known reference terminology for healthcare is SNOMED CT, a clinically validated, semantically rich, controlled vocabulary. SNOMED CT provides a standardized way to represent clinical phrases captured by the clinician and enables automatic interpretation of these [139].

With the emergence of the openEHR approach and its archetype reference model, there has been an increasing interest in recent years in exploring how semantic web technologies in general, and ontologies in particular, can facilitate the representation and management of archetypes. Archetypes are expressed in the Archetype Definition Language (ADL), which structures the content in four main sections: header, description, definition and ontology. Archetypes are used to specify clinical recording scenarios such as a laboratory test, a blood pressure measurement, a medication order, etc. Archetypes also define the clinical knowledge in the EHR by representing clinical concepts in the form of structured and constrained combinations of entities contained in the reference model.

Several attempts have been made to use ontologies for expressing the semantics of EHR data and consequently for encoding archetypes. An ontology is an explicit specification of a conceptualization shared by a community [140]. It is usually and information artefact, written in some sort of formal language, that describe concepts and relation in a given domain. Ontologies can be specified in any formal language. First-order or modal logics are frequently employed given their high expressivity. However, one of the main aspects of the evolution of ontologies is the use of specific ontology representation languages. These are languages that include special constructs that enhance or facilitates the task of representation and management of different ontologies. Today, the Web Ontology Language (OWL) is the de facto standard for specifying ontologies in the Web and in other systems [141]. It is a semantic mark-up language designed to represent rich and complex knowledge about things and relations between them. OWL has a formally defined meaning and it can be considered a general-purpose modelling language. OWL modelling results are called ontologies. In the medical domain OWL representations have been proposed for clinical information and clinical models from different EHR standards such as ISO 13606, openEHR, HL7 or Clinical Element Models (CEM). The OWL representations support the transformation of clinical models and clinical data between different EHR standards and OWL reasoning has been used for validating and checking the consistency of clinical models [142]. Since OWL uses first-order logic, the models, and description of data in these models, can be formally verified. Thus, inconsistencies in the model can be detected, and new information can also be inferred, by machine reasoning.

7.1.3. Middleware platforms to ingest health data to Big Data architectures

Ingesting health data with the use of Big Data technologies is a rather new technology for the e-Health domain. Recent technological developments and approaches in Big Data can be summarized as following.

There are still a lot of issues that need to be resolved before efficient health data analytics can be performed. One of the most important issues is the binding of data (to patients, situations, sensor devices). Since data are coming from different devices, this metadata has to be bound to patient IDs. The wide variation of data models and data warehouses with their own data binding is also a current problem. Figure 30 below shows an example of the versatility of types of data for a single patient.

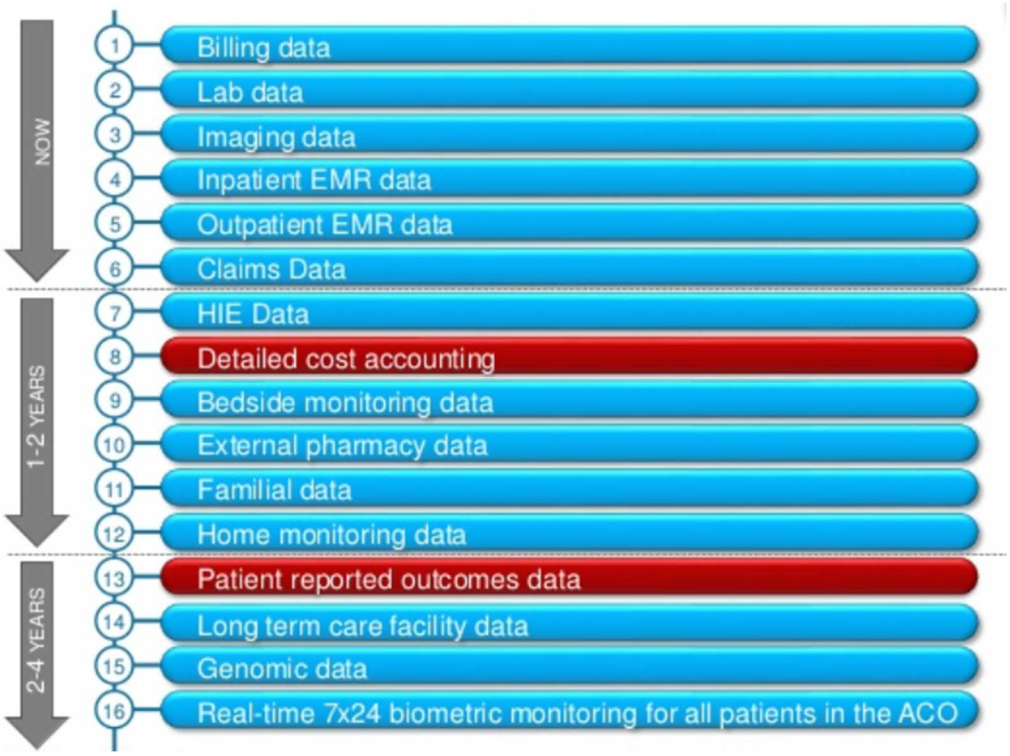

Figure 30 Example of the versatility of medical data for a single patient

The concept of data warehousing dates back to the late 1980s when IBM researchers Barry Devlin and Paul Murphy developed the "business data warehouse". From that point of time, many implementations have been successfully developed in different domains including healthcare. However, until recently, the approach was building proprietary enterprise data models whether it is dimensional approach (star-schema) or normalized approach (3NF models). These approaches, now called early binding architectures, force early data bindings to build one-size-fits-all solutions compromised, least-common-denominator, warehouses. Time has proven early binding architectures to be inflexible, and thus unsuitable for the today's high volume and high velocity of data and rapidly changing data analytics requirements of domain experts.

7.1.3.1. Late Binding Data Lake Architecture

A new approach, called late-binding architectures, delays data binding until the proper time and context, and retains the collected data its original, undiluted value. In these new Big Data architectures, the repository for structured, unstructured and semi-structured data in its original format is generally called "Data Lake". Having healthcare as one of the popular use cases for Big Data and analytics, we are recently observing several implementations of these architectures and concepts in healthcare [143], [144]. A data lake acts as the technology enabler to capture maximum value from all of the data being created across the continuum of care. Data lake architectures meet rapidly evolving business and clinical requirements by quickly and efficiently analysing new combinations of data from multiple sources across the health system and other systems that can be related. Traditionally, healthcare organizations have invested substantial time and effort to extract, transform, and load (ETL) data from its original format into data warehouses purpose-built for business intelligence and scientific analytics. A data lake strategy simplifies storage, management, and analysis of Big Data by consolidating data in real-time, near real-time or in batch from disparate sources and across multiple protocols. In this way, it can unify all data drawn from traditional databases and unstructured data, such as patient images, lab reports, pathology, genomics, clinical notes, and social media activities, clinical trial results, medical sensors,

wearables, home care IOT appliances, etc. Over the data lake, providers start to build analytic "sandbox" environments to execute predictive care analytics at scale and in near real-time [145]. The data lake architecture opens opportunities to find correlations across vast stores of data they previously were not able to query or examine.

7.1.4. Limitations of existing solution and relevancy for Medolution

There are several shortcomings of the current approaches in Healthcare data integration and interoperability. The first major issue is that e-Health domain is not still interoperable. The common data exchange layers are defined either as data dictionaries or through abstract data models that try to ensure interoperability within the boundaries of the associated initiatives. For instance, the query services, analysis methods or data exchange protocols envisioned by these initiatives can seamlessly run on top of the agreed common data element models which are set of core data elements. However, when it comes to achieving a broader range of interoperability, these efforts fall short: proliferation of common data element models does not help to solve the interoperability problem [146].

Regarding the data ingestion perspective, current solutions are not capable of processing data from different sources in different formats. While this problem is trying to be solved, new technology advancements are taking place, and new forms of data collection mechanisms for the patients (i.e. IoT architectures, new wearable devices etc.) are coming up together with new generations of hospital information systems and EHR systems. All these data should be ingested and analysed together in order to deduce meaningful results through data analytics. That is, processing EHR data only for data exchange is not enough for exploiting the power of existing data. The semantics of EHR data should be linked with the data coming from other sources also.

Existing solutions for EHR integration do not incorporate the Big Data architecture, since it was not needed up to now. Accumulating data in data warehouses and steaming data from different sources challenge the existing architectures and existing approaches fall short to meet the requirements in this Big Data world. For example, the challenges now are linking a diagnosis from a patient's EHR with the data coming from wearables on that patient and performing data analytics to help physicians for predictive medicine or even to help the patient take actions against bad conditions. Medolution will build the necessary Big Data platform, tools and APIs to ease this analytics using data from very different sources.

7.2. Healthcare Data Analytics

Analytics is "how you make sense of your data and uncover meaningful trends" [147]. It is often unclear how analytics is different from analysis and the words are used interchangeably. In the dictionary analysis is defined as "separation of a whole into its components parts", whereas analytics is defined as the method of "logical analysis". A method of logical analysis is commonly performed using algorithms. This applied logic produces a model in which the parts are related with statistical relevance. Often, analytics is future-oriented, predicting relations, whereas analysis is associated with what is or has been [148] .

Integrating Data Analytics in operational Healthcare Information systems requires [149] the use of full range and huge amount of heterogeneous information including electronic medical records, images and sensors that we refer as big data. The extraordinary potential to the exploitation of these amount of valuable information by using a combination of machine learning and data mining tools will improve patient care process and patient life quality [150]. According to [151], current health care systems under development or in production are lacking the potential benefits of big data analytics, see also [152]. A number of tools have been proposed to enable this potential:

- The first tool is a set of predictive models that can be obtained from big data sets covering variety of populations. These models will be combined with clinical practice guideline as decision support items to healthcare providers or clinical researchers in order to leverage personalized care in real time. These solutions may offer early detection and diagnosis before a patient develops disease symptoms.
- The second tool is a system for IoT based remote monitoring of vital signs, which can capture in real-time from wearable medical devices healthcare data and analyse in continuous manner these data to provide clues or indictors for caregivers. The monitoring can be done in hospitals or at home. The direct benefits of monitoring are ensuring the safety of patients and enabling the prediction of adverse events [153].
- The third tool will be the undertaken of comparative study to develop better ways for diagnosing and treating patients, such as mining large amounts of historical and unstructured data, looking for patterns, and model various scenarios to predict events before they actually happen [152].
- The fourth tool is a public health repository on a national level that turns patient and treatment information into actionable knowledge that allows timely detection and prevention of infectious diseases and outbreaks, thus benefiting the whole population [153].

The main benefits of using data analytics components in healthcare is to identify patterns of care and discover associations from massive healthcare records, thus providing a broader view for evidence-based clinical practice, identify previously unnoticed patterns in patients related to hospital readmissions and support a better balance between capacity and cost. In general the data analytical process starts by acquiring data from several sources, filtering and extraction features from it according to specific criteria, and then applying machine-learning algorithms [154].

7.2.1. Big Data Analytics for Healthcare

The main steps involved in analyzing healthcare data are data pre-processing, feature extraction/selection and machine learning. According to [155], the extraction and selection of a subset of important and relevant features from a large set of measured data is called feature selection (or attribute selection, or variable selection). This is especially important when working with complex and large medical datasets as these tend to contain a lot of redundant features. By applying dimensional reduction algorithms, the redundant information can be transformed into a small set of relevant features.

Machine learning is used to automatically learn general rules for prediction or classification by using two approaches: supervised learning (or predictive learning) and unsupervised learning (or descriptive learning). Supervised Machine learning applies when data instances are provided with labels explaining the ground truth. Unsupervised learning on the other hand, operates on unlabeled data in order get clusters of features allowing to discover meaningful classes. A special case of machine learning is deep learning. This approach is gaining huge successes and is bringing nice opportunities and potential in particular for health-care such as in image based pattern recognition where learning is done on multiple levels of representations in deep architectures. Deep learning is suitable for large un-labeled data-sets, such as patient personal information. A deep learning architecture is used for learning representations from both the labeled and unlabeled data thanks to the combination of unsupervised pre-training and supervised fine-tuning strategies to construct the models. One of the examples of deep learning architecture are composed of Restricted Boltzmann Machines, [156], which are probabilistic generative models that learn a joint probability distribution of observed (training) data without

using data labels. Restricted Boltzmann Machines use large amounts of unlabeled data for exploiting complex data structures. In order to create deep belief network, the learning is used to obtain the weights (and biases) between layers. Deep learning algorithms have been used in neuroimaging where restricted Boltzmann machine are used to get functional and structural MRI data for examining the depth parameter in the deep learning analysis for this specific medical data, and determining if the proposed methods can discover the unclear structure of large datasets. One of the unique characteristics of deep learning algorithms is their ability for learning data distribution without using label information. However, advanced deep learning methods are required to deal with noisy data and incomplete data sets by using for instance semi-supervised learning.

A recent survey of current Big Data methods and techniques applied to healthcare data, the problems and future possibilities, is found in [155]. It lists the challenges related to the well-known 4Vs characterizing Big Data: Volume, Variety, Velocity and Veracity, and adds 2 other Vs relevant for healthcare: Validity (are the data correct) and Volatility (how long do the data remain valid). It also presents a detailed overview of all steps involved in analyzing healthcare data, from preprocessing to feature extraction/selection and machine learning.

7.2.2. Medical Image Analytics

Image analytics is not a commonly used term [157]. Xerox PARC has defined this as extracting information from images and video, which is a quite similar definition as used for “computer vision” [158]. Three subfields of their *Image* Analytics are identified: 1) *fundamentals,* which is very similar to old-fashioned image processing; 2) *scene understanding:* in which logics objects and people are tracked and the patterns of these movements are interpreted in terms of behaviour, such as vandalism; 3) *people behaviour and group dynamics,* in which human activity and behaviour is recognized.

Venter and Stein have defined Image Analytics as the automatic algorithmic extraction and logical analysis of information found in image data [159]. However, they also give barcode recognition and facial recognition as examples, topics that have been part of image analysis already. In the medical field, their examples are image-based clinical decision support systems, also a topic that is commonly considered a medical image analysis approach.

A division has been made between *prescriptive* analytics and *predictive* analytics. Predictive analytics produce models that describe relations between metrics and variables. Prescriptive analytics has the ability to learn and adapt during the processing of images, video, text, and sound [159] and generates a set of prescriptions (suggested future actions) based on these learnings. According to IBM [160], Image Analytics is about spotting “non-obvious patterns” in images.

Alternatively, non-traditional image analysis methods like convolutional neural networks, and random forest decision tree classifiers have been called Image Analytics tools. Their application is quite popular in digital pathology images [161], and “nomics” data.

Medical Image Analytics has been defined as “adding medical images” to traditional healthcare analytics by IBM [162]. One of the Medical Image Analytics projects that IBM is working on is “Medical Sieve” [163], in which they use Watson [164] to generate a decision support system for radiologists exploiting its deep analytics. However, if we critically look at what Medical Sieve has produced [165], no new concepts or methodologies on image analytics have been introduced.

Medolution wants to explore a possibility to include contextual data (from a variety of other data sources) in the analysis of large sets of medical images. As medical image analytics is in its infancy, Medolution will also invest in developing a suitable platform to support the development of new image analytics algorithms.

7.2.3. Prototype Development Platforms for Medical Image Analytics

This section describes the result of a survey for prototype development platforms designed to perform medical image analytics.

VTK/ITK (originated from GE)
The Visualization Toolkit (VTK) is an open-source, freely available software system for 3D computer graphics, modelling, image processing, volume rendering, scientific visualization, and information visualization [166].
ITK is an open-source, cross-platform system that provides developers with an extensive suite of software tools for image analysis. Developed through extreme programming methodologies, ITK employs leading-edge algorithms for registering and segmenting multidimensional data [167].

MATLAB (MathWorks)
The MATLAB platform is optimized for solving engineering and scientific problems. The matrix-based MATLAB language is the world's most natural way to express computational mathematics. Built-in graphics make it easy to visualize and gain insights from data. A vast library of prebuilt toolboxes lets you get started right away with algorithms essential to your domain. The desktop environment invites experimentation, exploration, and discovery. These MATLAB tools and capabilities are all rigorously tested and designed to work together [168].
MeVisLab (MeVis Medical Solutions AG)
MeVisLab represents a powerful modular framework for image processing research and development with a special focus on medical imaging. It allows fast integration and testing of new algorithms and the development of clinical application prototypes [169].

Syngo.via Frontier (Siemens)
Syngo.via is the universal imaging software for 3D reading and advanced visualization. Multi-user, multi-modality, and multi-disciplinary, it streamlines radiology and Molecular Imaging, links departments, and connects sites. It brings a level of quality and efficiency to your viewing, reading, and reporting processes. Syngo.via is a smart, scalable imaging software that works: from practices to clinics to hospital chains [170].

IntelliSpace Discovery (Philips)
Clinicians performing research need one consistent environment where they can integrate new analysis tools and validate new research workflows. IntelliSpace Discovery offers an innovative way to evaluate the latest analysis methods at the forefront of medical imaging [171].

Open Innovation (Philips)
The Open Innovation Platform is a rapid prototyping environment to support the development, verification and clinical validation of algorithms. An important characteristic of this platform is that it enables deployment of prototypes based on CE labelled platforms[2] to be validated at clinical sites. This means that after the prototyping phase a development effort must be invested only once for the realization of the final product.

7.2.4. Data visualization in healthcare context

Data visualization refers to the techniques used to communicate data or information by encoding it as visual objects (e.g., points, lines or bars) contained in graphics [172]. Visualizations such as box

[2] Clinical users define the requirements for a new product. As a result, demonstrators / prototypes and products are built and need to be clinically evaluated. A demonstrator / prototype as well as a product need to comply with "essential requirements" as described in Annex I of Directive 93/42/EEC. According to this, medical devices must not only be safe but also function in a medical-technical way as described in the manufacturer's "intended purpose". Compliance with these requirements is proved within a certified quality management system according to EN ISO 13485. The prototyping environment will comply with the mentioned medical device directives.

plots and correlation matrices help quickly to understand the composition and relationships in the data. Which visualization is the most appropriate depends on the nature of data and its composition, what information is to be conveyed visually to the target audience, and how viewers process visual information. The 3 Vs of Big Data bring new visualization challenges. A white paper from SaS [173] presents an overview of common visualization techniques and tools, including those that are relevant for Big Data. Other big players as IBM, Google and Tibco also develop this subdomain.
When applied to healthcare data, visualization methods can help to

- classify findings – e.g. IBM Watson tumor detection;
- give insight in repetitive patterns – e.g. establish cause-effect relation to analyse epidemics;
- predict disease progression – e.g. extrapolate aneurism growth over time, etc.

Medolution doesn't aim to develop a new visualization technology as such but to investigate which type of visualization is most appropriate for healthcare professionals to explore big healthcare data. In this regard, it seems worthwhile to follow the developments in the field of visual analytics. This sub-domain of big data visualization has a strong focus on interaction and the combination of human and computational analytic capabilities.

7.2.5. Limitations of existing solution and relevancy for Medolution

Data Analytics
With respect to Data analytics algorithms, Medolution must solve important limitations related to the trustworthiness and heterogeneity of the sources. In Medolution, the data will be obtained from different sources such as medical devices and wearable sensors or from the application forms that are used to collect inputs from caregivers and patients. The trustworthiness of these sources may vary from low to high and depends on the context. To cope with the sheer size of the datasets Medolution must use sophisticated statistical techniques [174]. The trustworthiness of sources has a direct relation with the quality and complexity of the data sets used to implement the machine learning algorithms. In fact, it is common that the healthcare data contains biases, noise, and abnormalities. High-quality data can not only ensure the correctness of information but also reduce the cost of data processing. It is highly desirable to clean data in advance of analysing it and using it to make life-or-death decisions. However, the variety and velocity of healthcare data raise difficulties in generating trusted information. In addition, optimized methods are currently missing to deal with data quality subsequent problems before implementing classification, regression or clustering algorithms such as dimensionality, data over-fitting, repeated measures, missing values and missing variables, data redundancy and incidental endogeneity.

Medical Image Analytics
The current medical analytics prototype development environment is meant to be used on a so-called standalone device. This is insufficient for Medolution as the project will combine imaging based diagnostic results with other data sources like e.g. lab values, pathology results, etc. Most likely it should be possible to run the prototype applications on a client–server configuration and use a web browser to show the prototype side-by-side with a lab view application and include the pathology results in this same browser window as well (gallery application). This was more-or-less realized in the Medusa project. The solution was sub-optimal (rigid) and a more flexible solution might be requested.

Data visualization
At this point in time it is hard to tell whether the data visualization components currently available on the market fully support the decision making based on the available input sources. It will depend on the requirements for the different use-cases in Medolution.

7.3. Healthcare Decision Support Systems

There is a longstanding history of decision support in the medical sector. Most research focuses on decision support in clinical settings (of which Clinical Decision Support (CDS systems) are the best known). The aim of this type of decision support is to help medical professionals in diagnostics or selecting the right treatment. Some systems deliver real-time decision support during medical interventions. These are in fact expert systems defined and managed by medical professionals, based on a relatively limited set of data.

Healthcare Decision Support Systems or more commonly named “Clinical Decision Support Systems” (CDSS) are recognized for their ability to reduce healthcare costs and to improve healthcare quality. A clinical decision support system (CDSS) is a health information technology system that is designed to provide physicians and other health professionals with clinical decision support (CDS), that is, assistance with clinical decision-making tasks.

A useful overview of the SotA in CDSS is provided in the report published by the Agency for Healthcare Research and Quality of the U.S. Department of Health and Human Services in 2009 [175].

7.3.1. Types of CDSS

Clinical decision support systems (CDSS) are “typically designed to integrate a medical knowledge base, patient data and an inference engine to generate case specific advice” [176]. This type of CDSS is knowledge-based. It requires a (non-trivial) set of rules and an inference engine which can combines the rules from the knowledge base with the patient data resulting in a proposal (or simply relevant information) for a decision or diagnosis. A knowledge-based CDSS requires explicit knowledge from experts to-define the algorithms to be applied. It fails when hard data is not available as basis for the inference algorithms, but the big advantage is that this type of CDSS can explain why a certain decision has been supported. In this context, CDSS are said to be very helpful to assist in the clinicians’ work, still they are reluctant to adopt CDSS because of unnecessary workflow disruptions. Integrating a CDSS with a clinical workflow is of paramount importance to assist care professionals in reducing likelihood of errors and improving care quality. Thanks to use of Computer Interpretable Guidelines (CIGs).

The other type of CDSS is non-knowledge-based and takes a machine learning approach: it creates its own logic by detecting patterns and drawing conclusions on the basis of these patterns. The strength of this approach is that the statistical or machine learning algorithms used can uncover patterns or dependencies from huge volumes of data which are invisible to the human eye.

An overview of algorithms and methods that are relevant for DSSs and as well as a detailed discussion on data mining techniques are presented in [177].

7.3.2. Relevant Trends

The attention in decision support is shifting to data analytics for the mass market of consumers and patients. As a result of technological developments in the field of mobile computing and more recently the Internet of Things, a continuous stream of data can be available from a broad variety of data sources. This leads to new challenges with regard to the selection, interpretation and presentation of data, and at technological level with regard to scalability. Systems need to be capable to process billions of data streams, continuously and in parallel, secured per individual user.
Some applications of predictive analytics for decision support systems in the medical field are the following:

- *Emergency care* could benefit from clinical predictions built using data science tools with abundant potential input variables available in electronic medical records. Patients' risks could be stratified more precisely with large pools of data and lower resource requirements for comparing each clinical encounter to those that came before it, benefiting clinical decision making and health systems operations. The largest value of predictive analytics comes early in the clinical encounter, in which diagnostic and prognostic uncertainty are high and resource-committing decisions need to be made [149].
- With respect to *clinical care planning*, a patient not attending an appointment, a no-show, is disruptive to a clinic, may cause access and scheduling issues because of its effect on clinic capacity, and may increase the cost of clinic operation. While the importance of identifying individual patient no-shows is recognized, scheduling models that incorporate the presence of no-shows typically use an average no-show rate for all scheduled appointments. CDSSs can be developed based on machine learning models that uses past sequences of successes and failures, over a limited historical horizon, in a regression-like approach, to predict the probability of a success on the next occurrence [178].
- CDSS can be used also *for chronic disease management* to support clinical investigation. The best example is the prediction of diabetes and comorbidities. CDSSs can rely on the predictive analysis of diabetic treatment using regression based data mining techniques to discover patterns using classification algorithms that identify the best mode of treatment for diabetes across different ages [179].

Other developments that have an impact on CDSSs:

- *Data mining and storage*: the ever-increasing amounts of data that can be processed allow to provide a (near) real-time stream of data including context information of what patients are doing. It is expected that the assessment of the medical data will improve through the availability of a multitude of data streams – even from non-medical devices such as fitness trackers. Another new line of research attempts to develop a smart CDSS, by taking into account social and emotional parameters, exp. for remote patients. These are important steps towards more individualized treatments.
- *Data integration:* the integration of CDSSs with electronic health records and computerized physician order entry systems can reduce healthcare costs [180].
- *Cloud computing model*: processing massive amounts of data is enabled by the provisioning of storage and computational capacity.

7.3.3. Limitations of existing solutions and relevancy for Medolution

Existing decision support systems differ from the decision support needed in Medolution in the sense that:

- Existing decision support systems are often limited to diagnosis, and when they support treatment it is to warn clinicians about potential problems with drug dosage and/or drug interaction. No system has been designed yet for real-time decision support that makes use of multiple data streams, including highly sophisticated imaging data.
- Existing decision support systems are not linked to a hospital's medical protocol or guideline management system and medical procedures appear to have a relatively low degree of standardization, therefore it requires substantial time investments of an organization to adapt a DSS to their own needs.
- Existing decision support systems lack the flexibility required to accept massive amounts of data from heterogeneous sources on the one hand and to use these data for personalised decision support on the other.

- Even if there are CDSSs available in the market that are integrated with Electronic Health Record (EHR) systems, there is a need for an interface that allows for natural interactions using the Human Computer Interaction (HCI). Ideally, the decision support rules are to be defined by the medical professionals themselves. This requires that the rules can be defined in human language and converted into executive code.

7.4. Interactive user interfaces in healthcare applications

In this section we will examine the existing healthcare applications to understand state of art for their UI. Afterwards, we will explain state of the art frameworks and approaches for implementing generic and model based UIs.

There are various healthcare applications on the market that provide real-time data to healthcare actors such as patient, patient's family or caretakers, doctor / clinician. These applications mostly offer a presentation layer to provide the user with health device data and monitoring services. To get a better understanding of existing UI technologies and approaches relevant for Medolution, some examples with definitions and screenshots are given below.

- **iWander** [181] is an android application for patients suffering from Alzheimer's disease or dementia. It makes use of the GPS function of smart phones to track patient's location. These interfaces are created statically at design time and cannot be extended or modified easily. (See Figure 31 below).

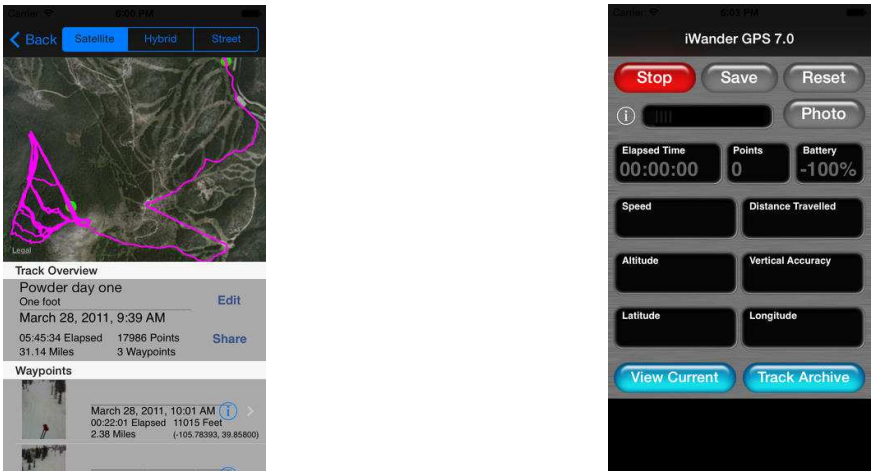

Figure 31 Screenshots from iWander which are statically created at design time [181]

- **Siren ePCR™** Suite [182] is a secure electronic patient care reporting system that improves the speed and accuracy with which paramedics can record patient's information. Designed for use on-scene or in ambulances en route to the hospital, the Siren ePCR™ software and complementary hardware employ an easy-to-use, touch-screen interface to provide paramedics with more efficient data capture tools. (See Figure 32 below).

Figure 32 Screenshots from ePCR™ [182]

A common feature of these applications is that they provide user with data from specific devices or provides user to enter specific input to system. So, they require no generalization and model base approach and instead implement application-specific GUIs.

7.4.1. Frameworks and approaches for implementing generic and model based UIs

Developing UI is a tedious task and is prone to errors in the same way as any other domain of application development. Model-based UI architectures aim to make the process less tedious and erroneous and at the same time more reusable and cross-platform friendly. Available e-Health/m-Health applications mostly do not utilize Model-Driven architectures.

In recent years, the UI paradigm began shifting to model-based frameworks, which put domain model at the heart of their design. The UI can be generated from domain model and it keeps in sync with domain model data objects (data binding). A model-based UI framework utilizes traditional UI frameworks instead of replacing them completely. Swing [183], SWT [184] are traditional Java UI frameworks. Since traditional frameworks are low- level APIs relative to model-based ones, they have more flexibility, but model-based APIs generally produce more rapid and robust results. They also provide multi-platform support. When switched to another platform and UI rendering framework, a UI which that is functionally identical but different in look and feel is generated through the same domain model data. Look and feel can still be modified if required, while platform and UI framework are kept the same by providing a different rendering and layout implementation. Model-based frameworks generally make use of DSLs (domain specific languages), which are generally used to define metadata at different levels of frameworks. For instance [185] presents a DSL such as below that defines role-based UI metadata where access level, data detail and visibility to access data from UI are assigned for each user role.

There are several model-based frameworks. Two of them are listed and explained below.

The CAMELEON reference framework: This model was developed from 2001 to 2004 under the CAMELEON Shared-Costs RTD IST Project (Contex Aware Modelling for Enabling and Leveraging Effective interaction) [186]**,** which aimed to build methods and environments, supporting the design and development of highly usable context-sensitive interactive software systems. This framework proposes four different abstraction layers for adaptive interfaces for a step-by-step generation of a user interface. The user interface presented to the user is referred to as the Final User Interface (FUI), which is rendered by the UI toolkit of the given platform, for example GTK+ or Java Swing. FUI is derived from a Concrete User Interface (CUI). This model is basically the same as the FUI, but independent of the toolkit description language. The CUI is generated from the Abstract User Interface (AUI). The AUI describes the interface independent of interaction modalities and devices. The AUI is derived from a domain model (the task and concepts model combined). The following section discusses the models applied in the interface adaption process (i.e. task model, concepts

model, context model and UI model) followed by a description of the automated adaption process. (See Figure 33 below).

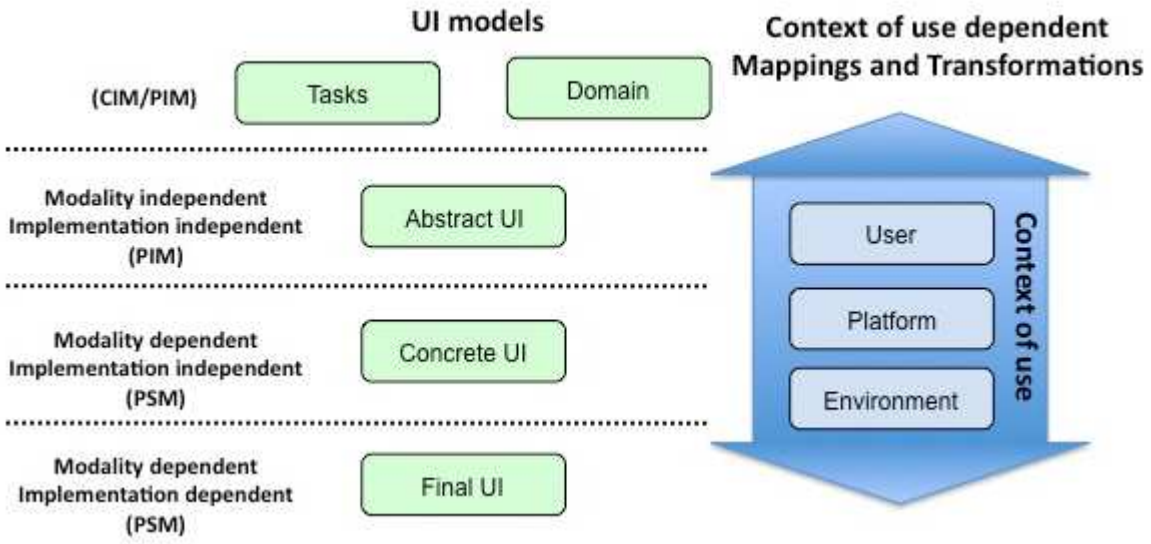

Figure 33 A simplified version of the Cameleon Reference Framework (CRF) [186]

Mappings and transformations between levels of abstraction depend on the context of use

EMF (Eclipse modelling framework) [187]: EMF can already generate Swing code from a given model data for PC standalone platforms. Work is in progress for integrating Web-based (with AngularJS SDK), iOS and Android platforms. EMF also has Eclipse IDE integrated implementation that makes it even more productive UI development candidate. Basic features of EMF are listed below:

- A software architecture proposed by the OMG (Object Management Group).
- Application specified in high-level, Platform Independent Model (PIM).
- Transformation technologies used to convert PIM to Platform Specific Model (PSM), implementation code.
- Includes several open modelling standards:
  - UML™ (Unified Modelling Language)
  - MOF (Meta-Object Facility)
  - XMI (XML Metadata Interchange)
  - CWM (Common Warehouse Model)

EMF's components can be summarized as follows:

- Core Runtime: Notification framework, Ecore meta-model, Persistence (XML/XMI), validation, change model
- EMF.Edit: Support for model-based editors and viewers, Default reflective editor
- Codegen: Code generator for application models and editors, Extensible model importer/exporter framework

Figure 34 below describes the relation between the basic components of EMF.

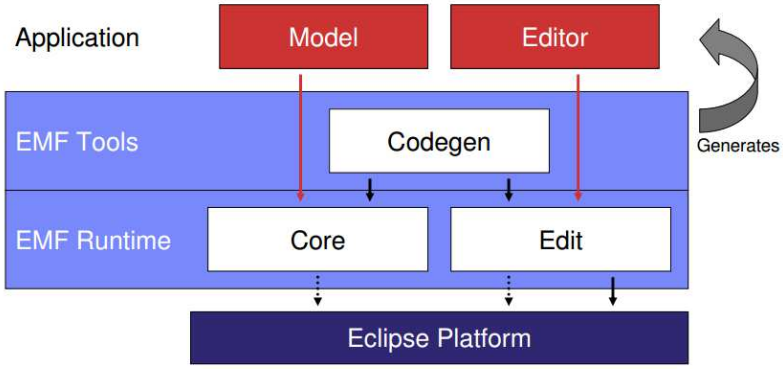

Figure 34 Fundamentals of EMF [187]

Finally EMF's model import and generation system operates as stated below to generate UI from domain model files. (See Figure 35 below).

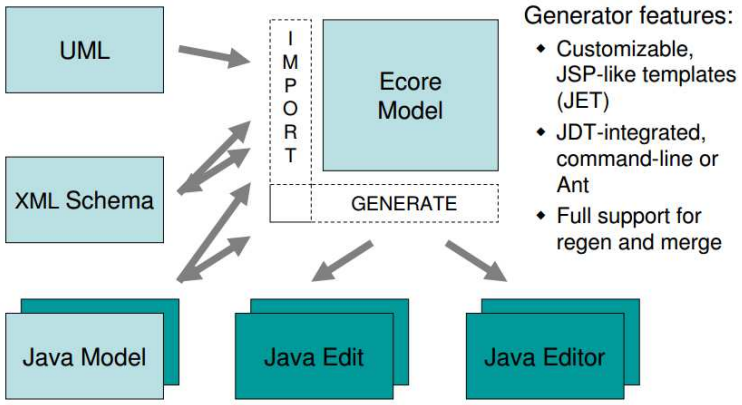

Figure 35 EMF's model import and generation system [187]

7.4.2. Limitations of existing solution and relevancy for Medolution

Existing solutions such as Siren ePCR™ implement specific and custom UIs. There are two reasons for that. Firstly, UI screens such as shown in Figure 31 can only be implemented in a custom fashion and not by a model-driven UI architecture. Secondly, most of the applications do not aim to utilize data from various devices, which makes model-driven architecture unnecessary. As a result, such applications should either overcome fundamental design changes if they are trying to support inputs from various devices or continue with tedious UI implementations for a new device screen to be added.
In Medolution, an extensible UI architecture should be used to integrate every new incoming device without much effort and problem. Architecture should also make use of custom made screens such as shown in Figure 32, since such UI cannot be generated. Model-driven architecture also allows to use other custom designed DSLs. For example, it is possible to design supportive DSL to define availability of domain data fields for each actor role that uses the application.

7.5. Limitations of existing solution and relevancy for Medolution

When it comes to the exchange, analysis and application of healthcare data, the challenges are manifold.

- Interoperability of healthcare data exists to some extent but the proliferation of common data element models does not help to solve the interoperability problem. Current solutions are not capable of processing data from different sources in different formats. The challenge for Medolution is to achieve a form of automatic data format harmonisation such that data from new applications and devices can easily be added independent of the origin of the data.
- With respect to data analytics, Medolution must solve important limitations related to the trustworthiness and heterogeneity of the sources. It is highly desirable to clean data in advance of analysing it and using it to make life-or-death decisions. Moreover, it is important to develop optimized methods for dealing with data quality issues before implementing classification, regression or clustering algorithms. Medical image analytics is in its infancy. Medolution wants to contribute to this field by developing a prototyping platform and new algorithms that allow to use (big) data from other sources in the analysis and interpretation of images.
- When it comes to using the collected and analysed data, current decision support systems lack the flexibility required to accept massive amounts of data from heterogeneous sources on the one hand and to use these data for personalised decision support on the other. Related to this is the need to enable medical professionals to define and tune the decision support rules themselves in human language without the intervention of programmers or an information specialist.
- For the development of user interfaces for medical professionals and patients, custom UIs are not feasible. Medolution needs an extensible UI architecture.

8. Privacy and security solutions for IoT and Big Data systems in Healthcare

One of the paramount challenges in Medolution is ensuring privacy and security while dealing with a huge volume of health data from various heterogeneous data sources and making the analysis results available for health professionals and potentially patients. This Chapter deals with the Privacy and Security solutions that are relevant to the project, while a detailed overview of the relevant privacy and security regulatory constrains to be considered is presented in the Appendix A. Since Medolution builds upon the results of the Medusa project, which provides a collaborative cloud access to medical information and addresses security, latency and collaboration related aspects, this Chapter addressed only selected topics of Privacy and Security in the context of Medolution. Thus, this Chapter first introduces general cloud security aspects relevant to BDHS. It follows with an overview of the security standards that can be applied for managing privacy and security on all stages of working with health data. In addition, encryption and anonymization as well as user-centric data privacy are then surveyed to introduce privacy strategies and technologies relevant for health data analytics that can be applied and enhanced to meet Medolution objectives.

8.1. General Cloud Security

Because of the obvious scalability, flexibility and availability at low cost of Cloud services, there is a rapid trend of adopting Cloud computing among enterprises, but in e-Health there is strong resistance to integrate Cloud services/computing with medical data (including the ones collected by medical devices) mostly because of security and privacy vulnerabilities on the sensitive information carried with medical data. Keeping security and privacy issues as the main concern, however, it is hard to resist the advancements in Cloud computing technologies in e-Health arena as well.

Securing a Cloud system, as other IT systems, has three major challenges:

- Integrity: means that the various elements of the system (data, applications, services, infrastructure, etc.) cannot be modified without their owner's acknowledgment.
- Confidentiality: means that these elements can be accessed only by authorized entities (people or software).
- Availability: means that these elements can be accessed when they are needed.

A Cloud system can expose its services at different layers [188]:

- At Infrastructure as a Service layer (IaaS), organizations use the infrastructure (virtual machines, network, data storage) proposed and managed by the Cloud provider. They bring their applications and services to this infrastructure.
- At Platform as a Service layer (PaaS), organizations build applications upon the service APIs (databases), tools and infrastructures proposed and managed by the Cloud provider.
- At Software as a Service layer (SaaS), organizations subscribe directly to the applications they need to treat their data. All the elements (applications, services, infrastructures) are managed by the Cloud provider.

Potential security threats against the Cloud system exist at each of these layers. But, according to the considered layer, providing adequate protections will be the responsibility of the Cloud consumers (organizations) or the Cloud providers.

Threats will depend on the type of the Cloud system: *public* (the Cloud provider is external to the organization and the services proposed are shared with several organizations) or *private* (the Cloud

provider is internal to the organization). Although a private Cloud system can be considered safer than a public one, threats still exist. Organizations can also use a hybrid Cloud system, which means that they mix services and resources from public and private Clouds. In this case, the threats are cumulative.

8.1.1. Cloud Security Levels

Security should be ensured at different levels to keep the Cloud system running [189], which are discussed in more details below:

- User level
- Software level
- Virtualisation level
- Network level
- Data Storage level

1. User level security

The User level applies to end users of organizations (who run applications) and administrators of organizations and Cloud providers (who manage different parts of the Cloud system, according to the considered service layer).

At this level, security is mainly achieved by strong authorization and authentication mechanisms enforced to access to applications, services APIs and tools. Here, a well-known implementation of authorization and authentication scheme is better than a custom one because it will be already tested at a large scale.

Recent authentication mechanisms make use of multi-factor authentication technique. To the classically unique credentials for each user, a second factor is added. This new factor, as for example Amazon Web Services (AWS) implements it, can be based on a unique number that the user receives by SMS during the connection process and sends to the system to confirm the authentication.

In case of connection to an identity management system, this should be done by using published APIs, as for example Lightweight Directory Access Protocol (LDAP).

A strong authorization and authentication mechanism is the first security barrier against external threats. But this is not sufficient against threats from internal staff of organization or Cloud provider, so security policies have to be defined and enforced. For example, in order to maintain data confidentiality and integrity, the Cloud provider's administrators should just have possibility to manage data without being able to see what exactly the data is. The Cloud provider's administrators should also regularly check and update these security policies.

2. Software level security

The software level applies to applications and database, web or development-services and their APIs. Some examples of the threats resulting from the unauthorized usage of web applications are:

- Cross Site Scripting (XSS) attacks, which will inject malicious scripts into web contents.
- Backdoor and debug procedures let into the code, intentionally or not, which may allow intruders to bypass security controls and access confidential data.
- CAPTCHA Breaking, which can result in DoS (Denial of Service) attacks from software robots or in data pollution.
- SQL Injections (SQLi), which result in execution of malicious SQL statements (also commonly referred to as a malicious *payload*) that control a web application's database server.

So, organizations and SaaS providers must implement known application security techniques during the development process of their applications, to protect them from the common vulnerabilities associated with the web. Technologies as Active Content Filtering, Content Based Data Leakage Prevention and Web Application Vulnerability Detection have for example been proposed to prevent XSS attacks.

SQL Database services can be the target of SQL injection attacks, which aim to insert a malicious code into a standard SQL code. Thus intruders gain unauthorized access to a database and are able to access confidential data [190]. To prevent such attacks, developers adopt some techniques such as not using dynamically generated SQL in the code, validating all the request parameters entered by the user, etc.

APIs of web services can be used by organizations to build applications to access the Cloud system. Access is generally made via HTTPS requests and users have to use a secret access key to calculate a signature, which will be included in their session requests for authentication. Via different kinds of attacks and automated tools, intruders will try to break encryption or to get signatures and user credentials and so, may expose the whole Cloud system. However, with well-secured web services and tools like web application firewalls (WAF), these credentials and signatures expire after the user's session so the risk is not very high.

Development services can be proposed by Cloud providers to organizations to create their own Cloud services. Malicious codes can be inserted by attackers in such services and can be executed with development services. This may expose the whole Cloud system and so, disable the availability of applications even to the authorized users.

3. Virtualization level security
Software in the Cloud (applications and services) are running on guest virtual machines (VMs) that are executed on hypervisors on the physical computing resources of the Cloud provider (internal to the organization, in case of private Cloud, or external, in case of public Cloud). The Cloud provider operates, manages and controls these various components.

Physical host security is typically handled by the Cloud provider.

Hypervisor security mechanisms exist to help the system integrity with a strong isolation between VMs, particularly in a public Cloud environment where physical and virtual resources are dynamically shared between multiple tenants (organizations). Using authentication and encryption techniques such as IPsec, VMs are authorized to only communicate with the ones they are supposed to [191]. Hypervisor security mechanisms also bring dedicated physical interfaces to VMs for communicating with the host operating system.

Due to the isolation between guest VMs, applications and services deployed on different VMs are also isolated from each other and the data belonging to one organization are inaccessible to others. However, software deployed on one guest VM operating system remains sensitive to attacks. A malicious code inserted in the operating system may for example interfere with the hypervisor or other VMs. Another example is an attack on Secure Shell (SSH), the basic way to connect to the operating system, in order to get API keys or user credentials.

Thus, users who have the management responsibility of the guest operating system must apply the main security measures such as installing virus checker, configuring the VM security group firewall, enforcing system updates and security patches, etc. They should also enhance security by installing specialized tools (host based firewalls, host based intrusion detector etc.).

4. Network level security

At the network level, Cloud security concerns mainly the integrity and confidentiality of data, which transit on the network, particularly in case of public Cloud. Availability of the whole system may also be affected in case of an attack at this level. Here are examples of attacks, which can be launched at the network level:

- Denial of Service (DoS) attack: it aims to make a service unavailable or highly degraded by saturating the network bandwidth and the computer capacity with a huge amount of service requests. We talk about Distributed Denial of Service (DDoS) when the attack is launched from many computers at the same time.
- IP spoofing attack: here the attacker masquerades as a trusted host to conceal his identity or gain access to a network in order to steal or corrupt user's data.
- Network sniffing attack: it is a passive attack that aims to capture network packets and read data inside them.

These kinds of attacks are not specific to Cloud environments. They may occur as soon as the considered network has a public access point. Even in a private context, some malicious code that infected internal computers may launch an automated DDoS attack or sniffing data for example.

Instead of making a service unavailable, the consequence of a DoS attack against a Cloud system can be economical. Supposing that the Cloud provider has associated to the considered service an auto-scaling functionality which allows an unrestricted scalability, by adding nodes (VMs) when the amount of service requests grows up. During the attack, many nodes should be added, with the corresponding additional costs [192].

A permanent monitoring of the network can help in mitigating the risks of DoS and DDoS attacks [193]. Spoofing and sniffing attacks can be reduced by using data-in-transit encryption and user authentication techniques, performing filtering for incoming and outgoing network packets, or even implementing techniques like virtual private network (VPN) that aims to encapsulate and encrypt all the packets over the network.

5. Data Storage security

At data storage level, securing the Cloud system first means ensuring the availability of stored data (data-at-rest) in case of storage infrastructure failure (technical fault, disaster at some Cloud location…). Data backup and redundant data storage allow data to be (almost) always accessible. Data backup also minimizes the consequences of a successful attack.

Securing the Cloud system also means ensuring the confidentiality of stored data. To achieve this, data has to be encrypted. Many tools exist to encrypt individual file content before pushing it on the storage support and decrypt after getting it in memory. The case of data stored in databases is different: database engine directly reads and writes data on the support and should have decrypted data in memory in order to treat them. Thus, most database engines include internal encryption mechanisms. Using storage devices or file systems with built-in encryption mechanism are also solutions. All these solutions are based on key management systems. Keys should not be lost; otherwise data cannot be accessed anymore.

A particular security threat against stored data, which is minimal in a private Cloud but could be severe in a public Cloud, is Data Remanence. This refers to the data left out in the storage support when data is transferred or removed [194]. Other security threats against Cloud data storage have been discussed in [189].

8.1.2. Reference Deployment Models for Cloud Computing Security

Five reference deployment models for Cloud computing that progressively address the main user security concerns have been presented in the recent publications: the separation model, availability model, migration model, tunnel model, and encryption model [195], [196].

1. Separation Model. The main idea of the separation model is to implement separation of duty, which is one of the key approaches for preventing fraud, errors, and abuse of privileges. Separation of duty assumes that at least two or more principals are involved in any single transaction. Thereby only a part of the transaction can be in responsibility of each principal, so that any of the principals can have control over critical processes.

Adapting this approach to Cloud computing should address two basic application cases: data needs to be processed and stored. That means that two independent services are responsible for data processing and data storage.

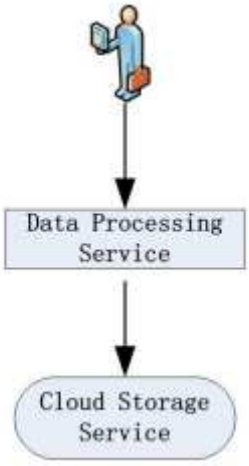

Figure 36 Separation Model [196]

Figure 36 demonstrates a possible implementation of the separation model. The data processing service processes the data and presents it to the user. The Cloud storage service is responsible for making the data persistent and accessible any time for the user.

2. Availability Model. The main idea of the availability model is to provide for the availability of the users' data by means of replication. Implementing the model for Cloud computing supposes that there are at least two independent services for data processing and storage as well as a mechanism for synchronization and replication of the data.

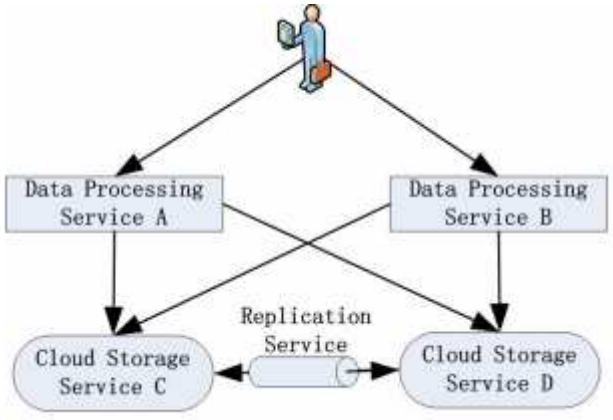

Figure 37 Availability Model [196]

Figure 37 illustrates a possible implementation of the availability model. Two independent providers provide equivalent data processing services and equivalent data storage services imposing redundancy on both data processing and storage. Replication of data between the two Cloud storage services is bi-directional and transparent to users. The replication service is responsible for the data replication and synchronization.

3. Migration Model. The migration model concerns the capability of migrating data from one Cloud to another. So that the users' concern about the excessive control of their data by the Cloud provider is minimized by knowing that they can easily switch to another service provider by moving their data from the current Cloud storage to another.

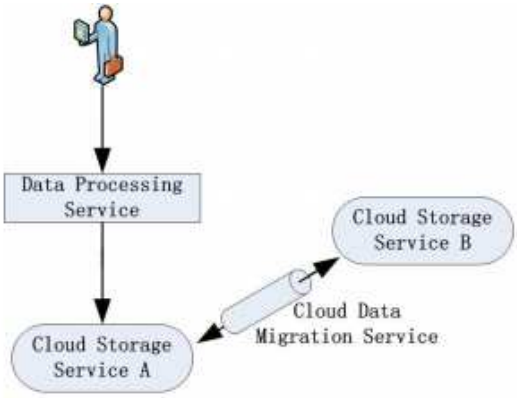

Figure 38 Migration Model [195]

Figure 38 illustrates a possible implemantation of the migration model with two Cloud storages provided by two independent Cloud providers. They support an opportunity for data import and export. A Cloud data migration service interacts between the storage services and guarantees easy data migration between the storage services.

4. Tunnel Model. The separation model separates the processing from the storing in order to prevent frauds and errors. The tunnel model goes forward and isolates the two service providers by cutting all the direct communication between them. In doing so, it makes sure that neither of the service providers will be able to identify each other, as well as additional filters can be imposed on their communication.

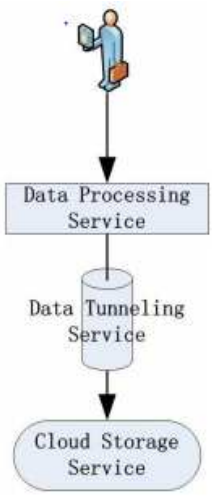

Figure 39 Tunnel Model [196]

A possible implementation of the tunnel model is shown in Figure 39. A tunnel service is located between the data processing and the data storage services. It encapsulates the communication between them and is responsible for providing the corresponding interfaces. Thus, the data processing service does not need to care about details of the Cloud storage service (e.g. location, identity, and interface). The Cloud storage service, in its turn, will not be able to relate the stored data with a specific data processing service. Thus, not only data processing and storage are completely isolated, but also the both service providers.

5. Encryption Model. The encryption model focuses on preventing unauthorized data disclosure or unauthorized modification on the data stored in the Cloud. In doing so, it relies on cryptography support.

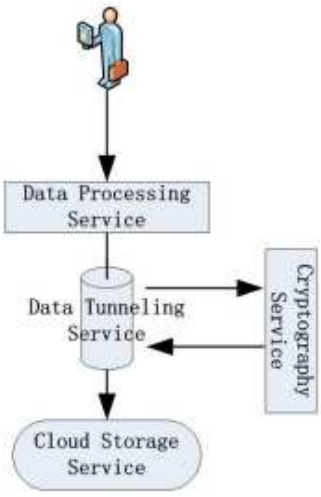

Figure 40 Encryption Model [196]

A possible implementation of the encryption model is presented in Figure 40. The tunnel model is augmented with a cryptography service providing for cryptographic operations on data. The data tunnelling service invokes the cryptography service before giving over the data to the Cloud storage

service. Thus, the data stored in the Cloud is processed cryptographically depending on the security requirements (e.g., encrypted, digitally signed). The data access implies that the data tunnelling service fetches the stored data and invokes the cryptography service (e.g. decryption, verification of signature) again before sending it back to the data processing service.

The above discussed deployment models for the cloud security are to be taken into consideration while working on the Medolution architecture.

8.2. Data security standards

As it was discussed in the previous sections in the context of Big Data, in particular in Healthcare domain, maintaining privacy and security at all stages of working with data, starting from data gathering from heterogeneous sources, is of paramount importance. Extensible Markup Language (XML) is still widely used for electronic data exchange across the world for health data exchange, although JSON uptake is gaining momentum lately. XML is enhanced with a sophisticated access control mechanism that allows not only to securely browse healthcare XML documents but also to securely update each document element [197], which allows for its effective application for the exchange of electronic health data.

Thus, the following sections provide an overview of some widely used XML mechanisms as they apply in health care followed by a discussion on the key integration profiles that facilitate centralized user authentication management.

8.2.1. OASIS Security Assertion Markup Language

The Security Assertion Markup Language (SAML) defines the syntax and processing semantics of assertions made about a subject by a system entity. SAML version 2.0 was approved as an OASIS Standard in March 2005 [198]. SAML is an XML-based framework for communicating user authentication, entitlement, and attribute information. It allows business entities to make assertions regarding the identity, attributes, and entitlements of a subject (an entity that is often a human user) to other entities, such as a partner company or another enterprise application. The main drivers behind the adoption of the SAML standard are single sign-on, federated identity and being applicable in Web services and other industry standards.

SAML consists of building-block components that together allow a number of use cases to be supported. The core SAML specification defines the structure and content of both *assertions* and *protocol* messages used to transfer authentication, attribute, and entitlement information. SAML assertions carry statements about a principal that an asserting party claims to be true. The valid structure and contents of an assertion are defined by the SAML assertion XML schema. SAML protocol messages are used to make the SAML-defined requests and return appropriate responses. Similarly, the structure and contents of these messages are defined by the SAML-defined protocol XML schema.

The means by which lower-level communication or messaging protocols (such as HTTP or SOAP) are used to transport SAML protocol messages between participants is defined by the SAML *bindings*. Finally, SAML *profiles* are defined to satisfy a particular business use case, for example the Web Browser single sign-on (SSO) profile. Profiles typically define constraints on the contents of SAML assertions, protocols, and bindings in order to solve the business use case in an interoperable fashion. The relationship between these basic SAML concepts is provided in the following figure. (See Figure 41 below).

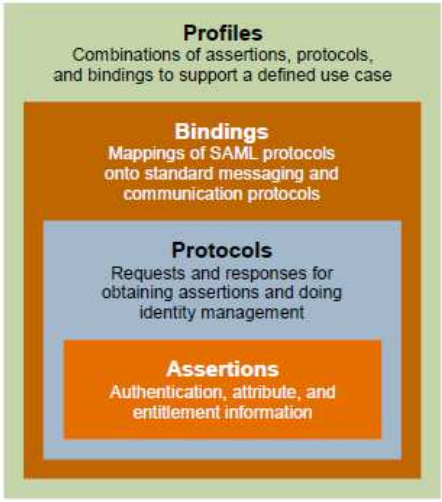

Figure 41 Basic SAML concepts [199]

SAML defines three kinds of statements that can be carried within an *assertion*:

- **Authentication statements**: These are created by the party that successfully authenticated a user. At a minimum, they describe the particular means used to authenticate the user and the specific time at which the authentication took place.
- **Attribute statements**: These contain specific identifying attributes about the subject (for example, that user "John Doe" has "Gold" card status).
- **Authorization decision statements**: These define something that the subject is entitled to do (for example, whether "John Doe" is permitted to buy a specified item).

SAML defines a number of generalized request/response *protocols*, some of which are presented below:

- **Authentication Request Protocol**: Defines a means by which a principal can request assertions containing authentication statements and, optionally, attribute statements.
- **Single Logout Protocol**: Defines a mechanism to allow near-simultaneous logout of active sessions associated with a principal.
- **Assertion Query and Request Protocol:** Defines a set of queries by which SAML assertions may be obtained.

SAML *bindings* detail exactly how the various SAML protocol messages can be carried over underlying transport protocols. Finally, SAML *profiles* define how the SAML assertions, protocols, and bindings are combined and constrained to provide greater interoperability in particular usage scenarios.

8.2.2. OASIS eXtensible Access Control Markup Language

eXtensible Access Control Markup Language (XACML) [200] is an XML-based language for access control that has been standardized in OASIS. XACML describes both an access control policy language and a request/response language. The policy language is used to express access control policies (**who** can access **what**, under what **conditions**, and for what **purpose**). The request/response language expresses queries about whether a particular access should be allowed (requests) and describes answers to those queries (responses). The latest approved version of XACML is 2.0; work is in progress for version 3.0.

XACML defines some major roles as presented in the following basic data-flow diagram. It should be noted that some of the data-flows represented in the diagram may be facilitated by a repository, and XACML does not prescribe a particular communication protocol for any of the data-flows. (See Figure 42 below).

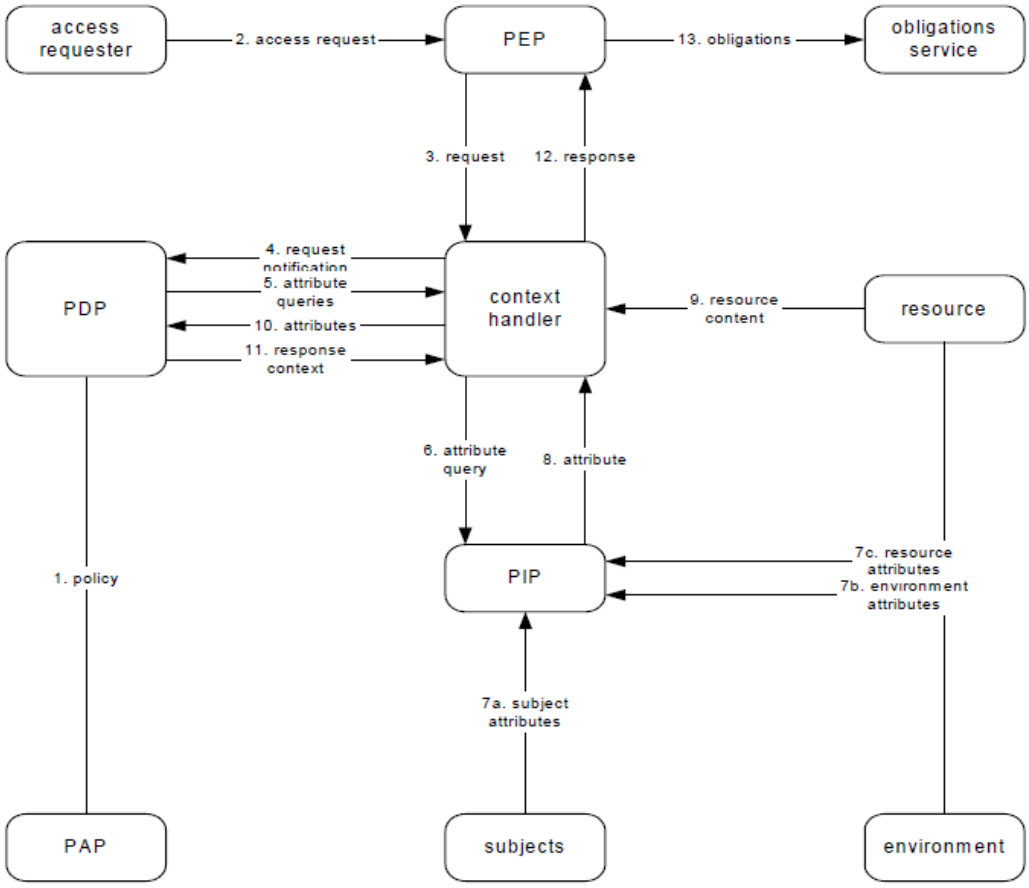

Figure 42 XACML Data-flow Diagram [201]

Policy Enforcement Point (PEP) is responsible for protecting access to one or more resources. When a resource access is attempted, the PEP sends a description of the attempted access to a Policy Decision Point (PDP) in the form of an authorization decision request. PEP may obtain attributes from on-line Attribute Authorities (AA) or from Attribute Repositories into which AAs have stored attributes. The PDP evaluates this request against its available policies and attributes and produces an authorization decision that is returned to the PEP. The PEP is responsible for enforcing the decision. The Policy Administration Point (PAP) basically administers and maintains the policies. The Policy Information Point (PIP) facilitates the PDP in acquiring any additional security attributes of resources and subjects in order to determine whether an access request is to be granted or denied.

XACML Policy language model

XACML defines three top-level policy elements: <Rule>, <Policy> and <PolicySet>. The <Rule> element contains a Boolean expression that can be evaluated in isolation, but that is not intended to be accessed in isolation by a PDP. So, it is not intended to form the basis of an authorization decision by itself. It is intended to exist in isolation only within an XACML PAP, where it may form the basic unit of management, and be re-used in multiple policies. The <Policy> element contains a set of <Rule> elements and a specified procedure for combining the results of their evaluation. It is the basic unit of policy used by the PDP, and so it is intended to form the basis of an authorization decision. The <PolicySet> element contains a set of <Policy> or other <PolicySet> elements and a

specified procedure for combining the results of their evaluation. It is the standard means for combining separate policies into a single combined policy. (See Figure 43 below).

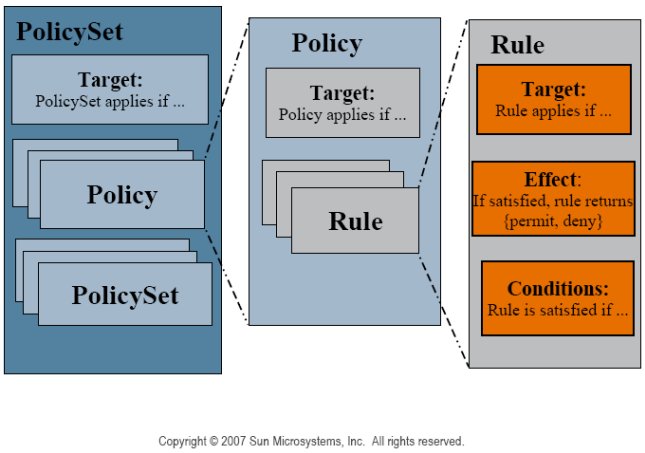

Figure 43 XACML Basic Policy Structure [201]

8.2.3. XACML Security Assertion Markup Language Profile

XACML itself defines the content of some of the messages necessary to implement this model, but deliberately confines its scope to the language elements used directly by the PDP and does not define protocols or transport mechanisms. Full implementation of the usage model depends on use of other standards to specify assertions, protocols, and transport mechanisms. XACML also does not specify how to implement a Policy Enforcement Point, Policy Administration Point, Attribute Authority, Context Handler, or repository, but XACML can serve as a standard format for exchanging information with these entities when combined with other standards.

One standard suitable for providing the assertion and protocol mechanisms needed by XACML is the OASIS Security Assertion Markup Language (SAML), Version 2.0. Hence, XACML SAML Profile [202] defines how to use SAML 2.0 to protect, transport, and request XACML schema instances and other information needed by an XACML implementation. There are also other XACML profiles such as Core and hierarchical role based access control (RBAC) profile and Privacy policy profile of XACML v2.0; however, these are not presented in this document.

There are 6 types of queries and statements used in the SAML 2.0 profile of XACML v2.0:

- AttributeQuery
- AttributeStatement
- XACMLPolicyQuery
- XACMLPolicyStatement
- XACMLAuthzDecisionQuery.
- XACMLAuthzDecisionStatement

Figure 44 below illustrates the XACML use model and the messages that are used to communicate between the various components. Not all components are necessary to be used in every implementation.

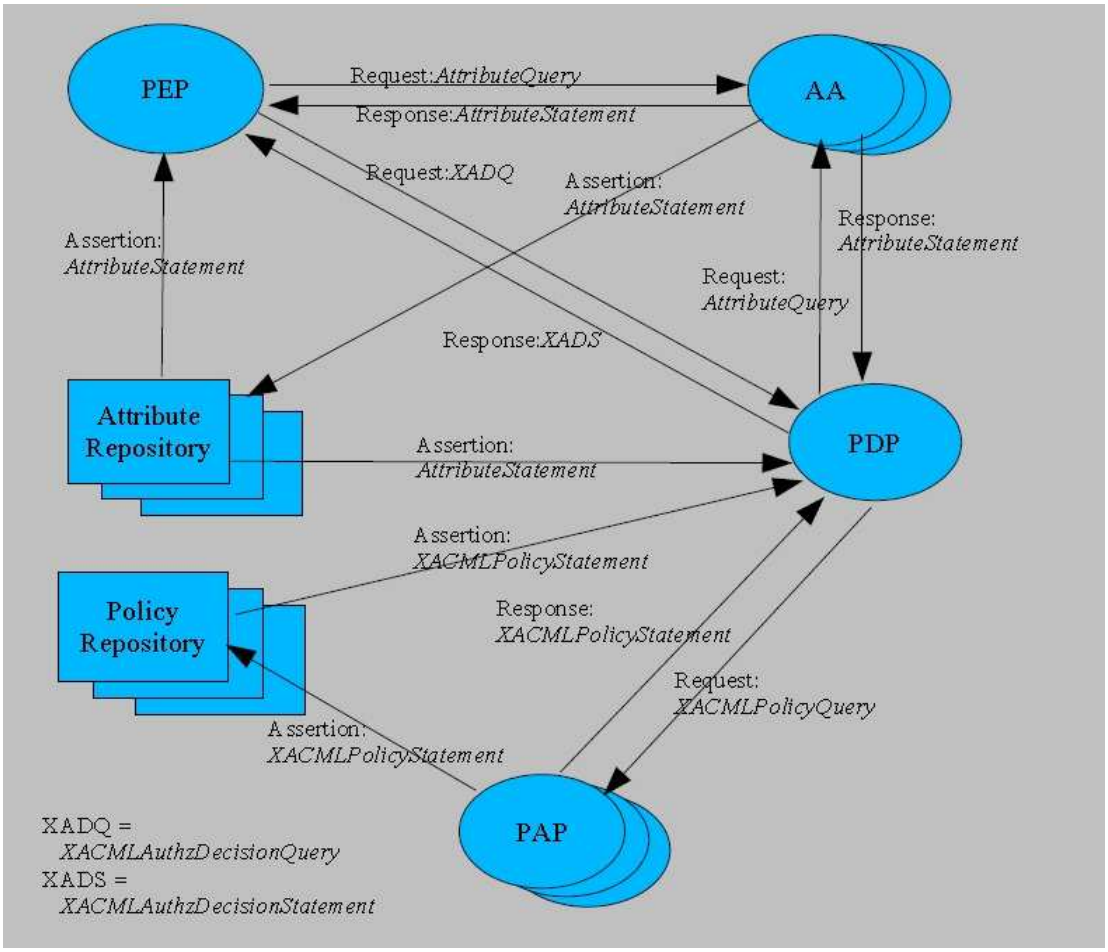

Figure 44 Use of XACML and SAML together [202]

8.2.4. OASIS Cross-Enterprise Security and Privacy Authorization Profile of SAML

The XSPA profile of SAML describes the minimum vocabulary necessary to provide access control over resources and functionality within and between healthcare information technology (IT) systems [203].

Figure 45 below displays an overview of interactions between parties in the exchange of healthcare information. The XSPA profile of SAML supports sending all requests through an Access Control Service (ACS). The Access Control Service on the Service User side receives the Service User request and responds with a SAML assertion containing user authorizations and attributes. To perform its function, the ACS collects all the attributes (e.g. organization-id, structural role, functional role, purpose of use, requested resource, and actions) necessary to create the Service User requested assertion. The Service Provider ACS is responsible for the parsing of assertions, evaluating the assertions against the security and privacy policy, and making and enforcing a decision on behalf of the Service Provider. The XSPA profile of SAML actually defines the semantics of the Service Request, Identity Assertion and Authorization Attributes that are seen in the figure below.

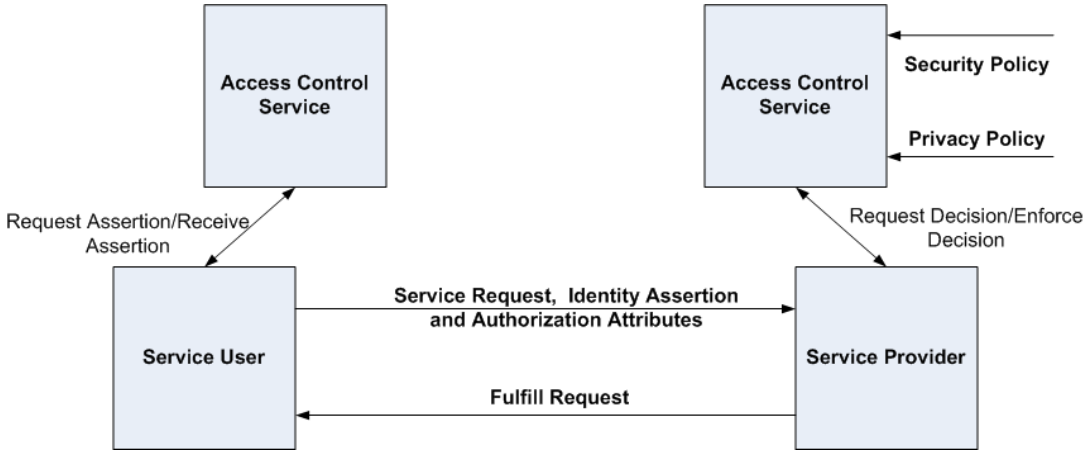

Figure 45 Interaction between parties in healthcare information exchange [203]

8.2.5. OASIS Cross-Enterprise Security and Privacy Authorization Profile of XACML

The Cross-Enterprise Security and Privacy Authorization (XSPA) profile of XACML [204] describes several mechanisms to authenticate, administer, and enforce authorization policies controlling access to protected information residing within or across enterprise boundaries. The policies being administered and enforced relate to security, privacy, and consent directives. This profile may be used in coordination with additional standards including Web Services Trust Language (WS-Trust) and Security Assertion Markup Language (SAML).

This profile specifies the use of XACML 2.0 to promote interoperability within the healthcare community by providing common semantics and vocabularies for interoperable policy request/response, policy lifecycle, and policy enforcement.

Similarly, with XSPA profile of SAML, the following figure provides an overview of interactions between parties in the exchange of healthcare information. (See Figure 46 below).

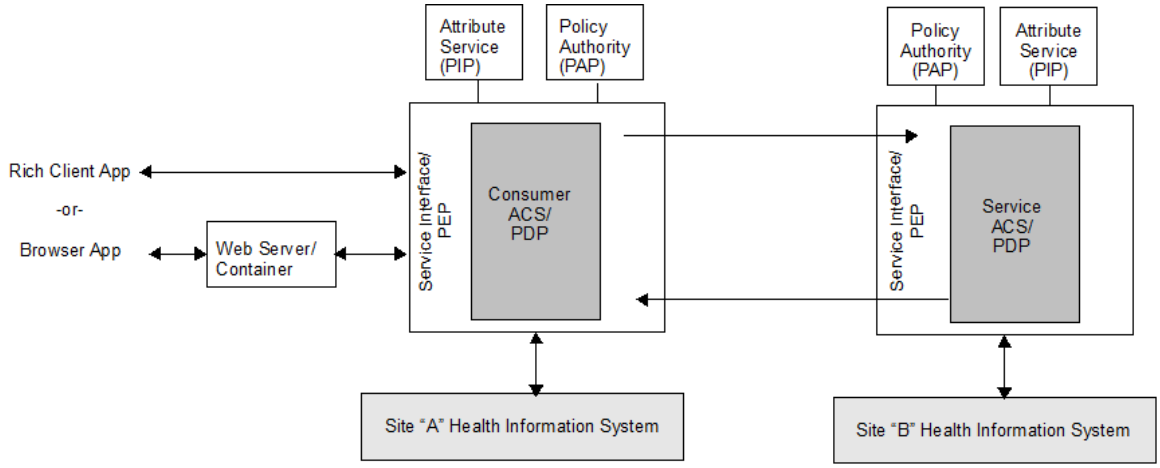

Figure 46 Interaction between parties in healthcare information exchange [204]

With the help of XSPA profile for XACML, all XACML request and response attributes are identified by a Uniform Resource Name (URN) from its vocabulary. This enables seamless mapping of data values between the client interface and policy services.

8.2.6. IHE Enterprise User Authentication Integration Profile

IHE Enterprise User Authentication Profile (EUA) defines means to establish one name per user that can then be used on all of the devices and software that participate in this integration profile, within an enterprise [205]. EUA facilitates centralized user authentication management and provides users with the convenience and speed of a single sign-on. This profile leverages Kerberos (RFC 1510) and the HL7 CCOW standard, specifically the user subject. In brief, CCOW or Clinical Context Object Workgroup is an HL7 standard protocol designed to enable disparate applications to synchronize in real-time, and at the user-interface level. CCOW is the primary standard protocol in healthcare to facilitate "Context Management", which is the process of using particular "subjects" of interest (e.g., user, patient, clinical encounter, charge item, etc.) to 'virtually' link disparate applications so that the end-user sees them operate in a unified, cohesive way.

User authentication is a necessary step for most application and data access operations and it is a workflow improvement for the users. The IHE EUA Profile adds value to the CCOW specification for the user subject by specifying the user subject and CCOW user subject suffix. EUA profile does not address security features such as audit trails, access control, authorization management and PKI.

The most important property of EUA is that, the environment is assumed to be a single enterprise, governed by a single security policy and having a common network domain. On the other hand, health care information exchange necessitates cross-enterprise transactions in many use cases; hence the Cross-Enterprise User Assertion Profile (XUA) is proposed by the IHE as explained in the next section.

8.2.7. IHE Cross-Enterprise User Assertion Integration Profile

In order to provide accountability in cross-enterprise transactions, there is a need to identify the requesting user in a way that enables the receiver to make access decisions and proper audit entries. The Cross-Enterprise User Assertion Profile (XUA) provides a means to communicate claims about an authenticated principal (user, application, system etc.) in transactions that cross enterprise boundaries [205]. The previous IHE profiles for an authenticated user identity (IHE Enterprise User Authentication Profile [EUA]) are not intended to function in cross-enterprise transactions. In a cross-enterprise environment, it is more likely that the transactions appear between two enterprises that maintain their own independent user directories. Hence, these type of transactions need the focus of Identity Federation standards.

The XUA Profile leverages Web-Services Security, SAML 2.0 Token Profile and the various profiles from W3C, and OASIS to support identity federation. XUA Profile is focused on Web service transactions, and specifies that when a Cross-Enterprise User Assertion is needed, these Web service transactions will additionally use the Web Services Security header with a SAML 2.0 Token containing the identity Assertion.

A very clear need on all Medolution use-cases is the recording of the user identity in security audit logs. The XUA profile does not define these auditable events; these are driven by other IHE transactions such as the Retrieve Document Set transaction. The method of authenticating the principal (user) and the method that the X-Service User Actor (e.g., XDS.b Document Consumer) uses to get the Identity Assertion are outside the scope of this profile. There are principal (user) attributes that can be needed in the use-cases: Doctor, Patient, Guardian, Emergency-Access. The Identity Assertion can contain attributes about the principal (user). However, yet XUA does not identify what standards to use to represent these attributes and their values, so this is left to specific implementations that have defined a local vocabulary or vocabulary translation.

The actors and transactions involved in XUA Integration Profile are shown in the following figure. Actually, XUA defines only two actors: X-Service User and X-Service Provider, and one transaction:

Provide X-User Assertion [ITI-40]. The actors and transactions in dashed lines are the ancillary ones, whose specifications are not defined by this profile. (See Figure 47 below).

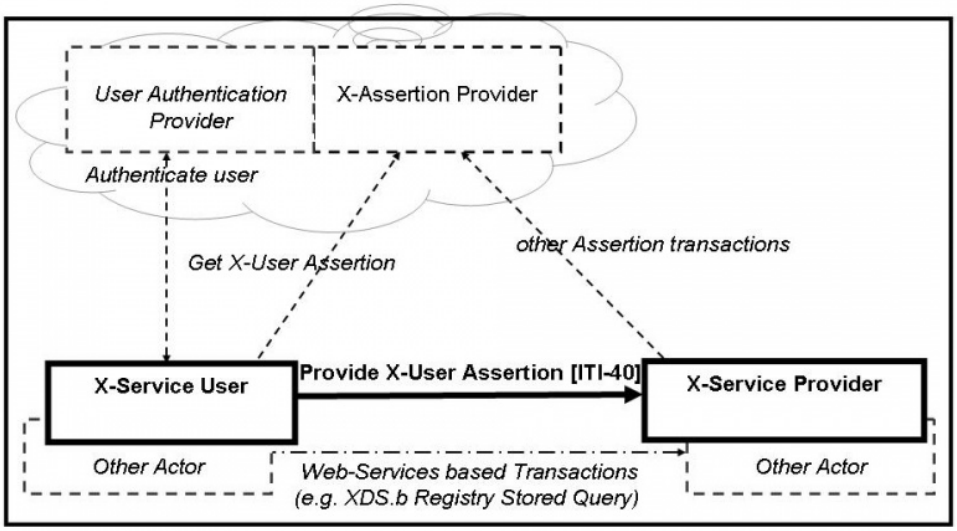

Figure 47 Cross-Enterprise User Assertion Actor Diagram [205]

8.2.8. IHE Audit Trail and Node Authentication Integration Profile

The Audit Trail and Node Authentication (ATNA) Integration Profile establishes security measures which, together with the security policy and procedures, provide patient information confidentiality, data integrity and user accountability [205]. This environment is considered the Security Domain and can scale from a department, to enterprise or cross-border Affinity Domain. The ATNA model considers that within the secure domain the following is true:

- All machines are host authenticated. This authentication identifies the machine as being one that is known to the security system of the organization, with known security characteristics.
- The host identification is used to determine what (if any) access should be granted to automated processes on that host, and/or persons under the direction of that host's access controls.
- The secure node is responsible for providing reasonable access controls.
- The secure node is also responsible for providing security audit logging to track security events.

Basically, ATNA Integration Profile defines the Secure Node actor, which is to be grouped with any IHE Actor according to the user requirements, and 2 transactions: Authenticate Node [ITI-19] and Record Audit Event [ITI-20]. It also benefits from the Maintain Time [ITI-1] transaction for consistent time handling. The relationship among these actors and transactions is presented in the following figure. (See Figure 48 below).

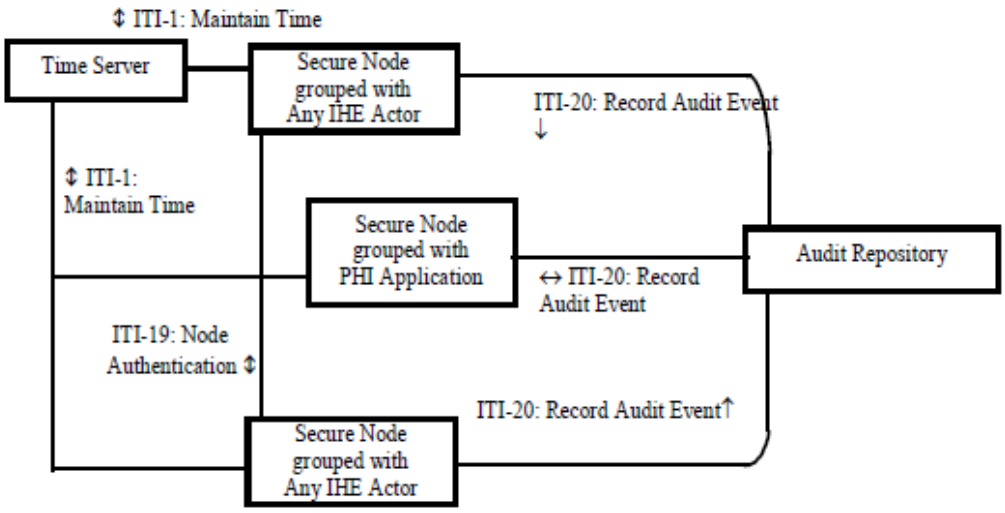

Figure 48 ATNA Actors and Transactions [205]

The Secure Node Actor shall include:

- The Authenticate Node [ITI-19] transaction for all network connections that may expose private information.
- All local user activity (login, logout, etc.) protected to ensure only authorized users.
- The Record Audit Event [ITI-20] transaction.

8.2.9. IHE Basic Patient Privacy Consents (BPPC) Integration Profile

Basic Patient Privacy Consents (BPPC) Integration Profile provides a mechanism to record the patient privacy consent(s) and a method for Content Consumers to use to enforce the privacy consent appropriate to the use [205]. This profile complements Cross-Enterprise Document Sharing (XDS) Integration Profile by describing a mechanism whereby an XDS Affinity Domain can develop and implement multiple privacy policies, and describes how that mechanism can be integrated with the access control mechanisms supported by the XDS Actors (e.g. EHR systems). BPPC profile provide mechanisms to:

- Record the patient privacy consent(s),
- Enforce the privacy consent appropriate to the use.

There are two actors in the BPPC profile, the Content Creator and the Content Consumer. Content is created by a Content Creator and is to be consumed by a Content Consumer. The sharing or transmission of content or updates from one actor to the other is addressed by the use of appropriate IHE profiles described in the Section 8.2. on Content Bindings with XDS, XDM and XDR.
In the BPPC profile, the Affinity Domain (i.e. healthcare information network in XDS terms) organizers create a set of policies (i.e. patient consents). Each of these policies are given an object identifier (OID). Each OID can clearly identify one of the policies defined by the healthcare information network. The Affinity Domain organizers can define their own policies in as clear of language as is necessary for the patients, providers, and systems to understand.
The BPPC profile shows how to capture a patient's acknowledgment and/or signature of one or more of these previously generated policies. This is captured using a CDA document with optionally a scanned copy or optionally a digital signature. Preferably, the scanned copy is with the patient's wet signature on paper acknowledgment. Patients need to know what they are consenting to, and they can understand human text; not many can understand technological aspects nor fully assess the long-term implications of their decision.

When a document is used, the document consumer actors are obligated to enforce the acceptable use. The document consumer actor is required to block access to documents that are not authorized. Any OIDs that are not understood by the document consumer actor must not be used to enable access. The BPPC profile was developed for the first time in 2006, and the profile is called "basic" because there are still many gaps that need to be addressed. For example, the profile does not address directly computer processable and executable privacy consent document formats, such as the ones that can be defined with OASIS XACML 2.0. The patients have to choose among the previously defined set of policies, they cannot define their access control settings dynamically. Also, the profile does not present how access control is applied; this is left to implementers.

8.2.10. oAuth

OAuth is an open standard to allow secure authorization in a simple and standard method from web, mobile and desktop applications. OAuth provides to clients a "secure delegated access" to server resources on behalf of a resource owner. It specifies a process for resource owners to authorize third-party access to their server resources without sharing their credentials. Designed specifically to work with Hypertext Transfer Protocol (HTTP), OAuth essentially allows access tokens to be issued to third-party clients by an authorization server, with the approval of the resource owner. The third party then uses the access token to access the protected resources hosted by the resource server [206]. The OAuth 1.0 protocol was published as RFC 5849 in December 2007. The OAuth 2.0 framework was published as RFC 6749 in October 2012 which is not backwards compatible with OAuth 1.0.

8.2.11. Relevance to Medolution

When patient data needs to be accessed and processed by Medolution architecture, depending on the security and privacy measures already been used by local care sites, the Health Data Ingestion stack will need to implement these security and privacy standards, in particular OASIS XACML and SAML for managing authorization and access control, IHE EUA and ATNA for authentication and audit logging and IHE BPPC for consent management.

8.3. Privacy strategies

8.3.1. Big Data encryption

Cryptography can be defined as a study of communication over an “insecure” channel. The two most basic goals of cryptography are privacy and authenticity. Depending on the structure of the keys, encryption schemes could be of two types: symmetric (or private key) and asymmetric (or public key) [202].

Symmetric encryption takes readable data, scramble it to make it unreadable, and then unscramble it again when it’s needed. Various symmetric encryption algorithms include Triple DES, Blowfish and Twofish [208]. **Asymmetric encryption** takes readable data, scrambles it, and unscrambles it again at the other end, but a different key is used for each end. This method is easier since only the party that needs to decrypt needs access to the private key [209]. **RSA** is considered an asymmetric algorithm [208]. Examples of other methods are **Hashing** [210], Advanced Encryption Standard (**AES**), Honey Encryption and Quantum key distribution [208].

Underlying these encryption methods, there are a few challenges posed by Big Data which are summarized below along with their possible solutions:

- **Hidden Metadata:** Network encryption hides the content of data but does not hide the amount of plain text or the identities of the communicating parties. For high latency applications such as email, cryptographic remailers can be used. For low latency applications, a secure IPsec connection can be established to a proxy server. In both cases, compromise of the remailer or server (or a warrant) would undermine the security [211].

- **Data Storage:** Data stored on a device should be protected with authenticated encryption, and the key (of at least 128 bits) should be stored in a secure place. A common solution is to store the decryption key on the device. This is clearly not an effective solution against a motivated opponent; a better option is to store the key in an external secure device (e.g., a smart card) or to use biometrics to authenticate the user and to derive a key that can be used to access the decryption key. Key management for the encryption of stored data is tricky as it requires a backup solution for most applications, in case something happens to the user or the user leaves the organization that owns the data [206].
- **Performing Operations over the data:** Data cannot be sent encrypted by the users if the Cloud needs to perform operations over the data. A solution for this is to use "**Fully Homomorphic Encryption**" (FHE), which allows data stored in the Cloud to perform operations over the encrypted data so that new encrypted data will be created. When the data is decrypted, the results will be the same as if the operations were carried out over plain text data. Therefore, the Cloud will be able to perform operations over encrypted data without knowledge of the underlying plain text data [212]. In a breakthrough result in 2009, Gentry constructed the first fully homomorphic encryption scheme which allows to compute the encryption of arbitrary functions of the underlying plaintext [213].
- **Trusting a centralized entity:** Another challenge is when multiple data owners have sensitive Big Data sets, as they may not trust each other with their valuable data. A trusted third party (TTP) can be used but might be neither realistic nor feasible due to legal barriers. Multi-Party Computation (MPC) is a cryptographic technique that allows for secure computation without having to trust a centralized entity.
- **Possible attacks on cryptosystems:** There are various other attacks possible on cryptosystems such as mathematical attacks, attacks using quantum computers and side-channel and fault attacks [211]. In order to prevent side-channel attacks, the opponent must be prevented from accessing the physical properties of the devices, e.g., by shielding the implementation and providing an internal power source. However, perfect shielding is not possible, and timing information can be obtained remotely. A second approach is to add countermeasures to the implementation [211].
- **Improper key generation:** Another common problem is that crypto implementations do not generate keys properly. A typical mistake is that not enough randomness is used in the key generation process, which makes the keys easy to predict to an attacker. A solution to this could be to use a key generation algorithm that only generates a small subset of the keys or the deliberate insertion of a side channel [211].
- **Cost consideration:** A cost barrier for cryptographic deployment is the cost of key management. This includes not only key establishment but also the full life cycle management, which includes generation, revocation, archival and destruction. Additionally, if high security level against physical attacks is required, the cost and complexity of the implementation is very high [211].
- **Access control policies:** Another major challenge in order to ensure that most sensitive private data is end-to-end secure, data must be encrypted based on access control policies i.e. decryption is allowed only if the entity trying to access the information is authorized by an access control policy. Attribute-Based Encryption can help in providing fine-grained access control of encrypted data [212]. Specific research in this area is still to be made more efficient and scalable [214].

All in all, proper use of current encryption methods will ensure that the data remains protected. However, proper due diligence is required in order to overcome the above mentioned challenges in the context of Big Data, which should be considered for Medolution applications.

8.3.2. Anonymization and pseudo anonymization approaches

Data anonymization is a type of information sanitization whose intent is personal privacy protection. It is the process of either encrypting or removing personally identifiable information from data sets, so that the individuals whom the data describe remain anonymous. Data anonymization enables the transfer of information while reducing the risk of unintended disclosure and thus violation of privacy laws as it was discussed in Appendix A of this document.

While the precise legal terminology was discussed earlier, it can be reminded that in general terms anonymized data refers to data from which the patient cannot be identified by the recipient of the information. Attributes within the data set which immediately identify an individual (e.g. name, address, postal code, etc.) must be suppressed or pseudo-anonymized (i.e. substituted with random data). Other attributes which, in conjunction with other data held by or disclosed to the recipient, could identify the patient must be generalized (e.g. birthdate changed from yyyymmdd to yyyymm or yyyy). These attributes are called quasi-identifiers. De-anonymization is the reverse process in which anonymous data is cross-referenced with other data sources to re-identify the anonymous data source.

The need for data anonymization is a clear requirement in the Medolution project. The general use case for anonymization can be stated as following:

- *We, as data custodians, want to share a data set with an authorized third-party, while being assured that the risk of re-identifying any individual passes a minimum risk threshold.*

There are many ways to address this use case but two techniques will be discussed below, which are more relevant for Medolution into two high level methods of data anonymization: K-Anonymity and Differential Privacy. For each of the high level methods two different implementations will be presented, the pros and cons for each will be discussed to reach a conclusion about suggested approach for Medolution.

8.3.2.1. K-Anonymity

K-Anonymity is the technique “to release person-specific data such that the ability to link to other information using the quasi-identifier is limited” [215]. K-Anonymity achieves this through suppression of identifiers and output perturbation. A release of data is said to have the k-anonymity property if the information for each person contained in the release cannot be distinguished from at least k-1 individuals whose information also appear in the release. More specifically, the data set has k-anonymity if all records within the data have at least k-1 records (also in the data set) with the same combination of quasi-identifier values (an equivalence class) [216]. A data holder can often identify attributes in their data that also appear in outside sources, and these attributes are candidates for linking. They are called quasi-identifiers, and it is essentially the combinations of these quasi-identifiers that must be protected [217]. For example, if k = 5 and the quasi-identifiers are age and gender, then a k-anonymized dataset has at least 5 records for each value combination of age and gender [218].

There are essentially two methods for generating anonymized data, which provides k-anonymity: generalization and suppression. Generalization is a process of replacing values of a given attribute with more general values. Suppression is an additional technique, which removes some outlier records from a dataset to avoid having to generalize the rest of the data set too much. Examples of the generalization hierarchies for the de-identification algorithms are discussed in [219]. The key challenge when dealing with dataset with many quasi-attributes and generalization methods is the

substantial computational complexity in selecting the combination, which quickly produces a k-anonymized dataset with minimal data loss. This is often referred to as the **k-optimal solution**. Also, a k-optimal solution might still be susceptible to attacks such as the "Unsorted matching attack", "Complementary release attack" and "Temporal attack" [215]. Solving this computational challenge is the study of much research [220] and for Medolution, the following review is limited only to the most prominent methods which are used to generate a k-optimal data set.

Sweeney Algorithm
Latanya Sweeney patented a k-anonymity algorithm for calculating a k-optimal data set from a given set of source data [221]. The Sweeney algorithm was novel but since it exhaustively computes the generalized data sets for all the possible equivalence classes in the lattice graph, it is not practical for large datasets with many quasi-identifiers. We reference it because of its place in the history of state of the art for k-anonymity calculations but we do not comment further on it for Medolution.

Samarati Algorithm
In an attempt to find a more effective way to achieve k-anonymity of generalized data sets, Pierangela Samarati presented an approach and an algorithm [222] that is based on the concept of a k-minimal generalization with suppression. A generalization of a data set is k-minimal if data are not generalized more than necessary to provide k-anonymity.

For any given table of data, there are different possible generalizations, but not all generalizations are equally satisfactory from a perspective of k-anonymity. For instance, the trivial generalization bringing each attribute to the highest possible level of generalization, thus collapsing all entries in table to the same list of values, provides k-anonymity at the price of a strong generalization of the data. Such extreme generalization is not needed if a more specific table (i.e., containing more specific values) exists which satisfies k-anonymity. This concept is captured by the definition of k-minimal generalization [222],[220].

The definition of k-minimal generalization with suppression introduced by Samarati is based on the concept of a **distance vector**. Like the name implies, a distance vector describes how far apart (or close) two values are. We can talk about distance between tables, or entries in a table, or entries between generalized tables. Each value in the distance vector describes "distance" between the generalization used are for corresponding attributes domains. Remember that each attribute has a sequence of generalization schemes that are monotonically increasing in generality. In the case of a distance vector between entries in a table (or entries between related generalized tables), the values in the vector describe how related the data is. That relationship is affected by the generalization applied to the entries. If distinct entries did not experience generalization at all, the distance value between the distinct entries attribute would be high.

The problem of finding minimal k-anonymous tables, with attribute generalization and tuple suppression (suppressing specific entries in a table), is challenging because of computation time, which grows exponentially with the number of data attributes that compose quasi-identifier. The key idea exploited by Samarati to cut down the computation time required to find k-minimal generalized tables is the observation that going up in the generalization lattice, the number of entries that must be removed to guarantee k-anonymity decreases.

The above observation allows concluding an absence of a solution that guarantees k-anonymity suppressing less than a set maximum amount at specific height in the generalization schemes. In other words, there cannot exist a solution, with a lower height of the generalization lattice that guarantees it. This property is exploited by using a binary search approach on the lattice of distance vectors corresponding to the domain generalization hierarchy of the domains of the quasi-identifier [222]. The algorithm goes through the lattice with a binary search, always cutting the search space in half, going down if a solution is found at a level, or up if not. Eventually, the algorithm finds the

solution with the lowest height, thus with the least generalizations. Then the best solution on that level (i.e. with the least information loss) with respect to a given preference (i.e. information loss metric) is chosen [223].

By looking at the distance vectors between the entries in a table it could be determined whether a generalization at a given vector satisfies k-anonymity by suppressing less than the maximum entries suppressed without computing the generalization. More precisely, for each distance vector, the minimum required suppression for the k-anonymity constraint to be satisfied by the generalization corresponding to the distance vector could be determined.

A table may have more than one minimal generalization satisfying a k-anonymity constraint for a suppression threshold. However, multiple solutions may exist which satisfy this condition. Samarati's algorithm returns a k-minimal generalization with the lowest height among all those existing. Although this may be considered a generally acceptable preference criterion, other preference criteria may exist depending on subjective measures and preferences of the data recipient. While this algorithm does use binary search and monotonicity property on the generalization lattice, the number of k-minimal generalizations itself remains exponential and can easily become too large to enumerate efficiently [224]. In short, this solution returned is not guaranteed to be globally optimal, although at the benefit of better run-time performance.

Optimal Lattice Algorithm
Very similarly to Samarati's algorithm, the goal of the Optimal Lattice Anonymization algorithm (which will be referred to as OLA, for short) is to find a node that enforces the k-anonymity property but also minimizes information loss. While a binary search to find k-anonymous nodes in the generalization lattice is still utilized and a k-minimal node is looked for. OLA is focused on selecting a globally optimal solution. To achieve that, the notion of a generalization strategy is used, which is understood as a series of connected paths from the bottom node to the top node. The optimal lattice anonymization algorithm seeks the optimal node in the generalization lattice with three broad steps, which can be summarized as follows [218]:

Step 1. For each generalization strategy, conduct a binary search to find all the k-anonymous nodes. To help with this process, predictive tagging is used to skip computations on particular nodes since these computations can be time consuming.

We consider a child node to be a node that is more generalized with respect to a single quasi-identifier by a single level in that quasi-identifier's generalization scheme. Likewise if a node not k-anonymous, then all parent's nodes are not as well. When evaluating if a particular node is k-anonymous or not, we may run into a case where if we suppress the equivalence classes with less than k entries, then the data is k-anonymous. Suppression is preferable to generalization because the former affects single records whereas generalization affects all the records in the dataset.

Therefore, when searching for a solution, a solution that imposes more suppression would be selected instead of one that imposes more generalization. However, because of the negative impact of suppressed data on the ability to perform meaningful data analysis, the end-users will want to impose limits on the amount of suppression that is allowed (MaxSup, which is referred to in the description for Samarati's algorithm). It is assumed that the data analyst will specify MaxSup such that complete case analysis can be performed or imputation techniques can be used to compensate for the missing data.

Step 2. For each generalization strategy with k-anonymous nodes, only the k-anonymous node with the lowest height within the strategy is retained. Effectively, k-anonymous nodes with no parents are also k-anonymous. All such nodes will be classified as k-minimal nodes. Note the distinction

between Samarati's and OLA approaches, where in Samarati's we only cared to find one k-minimal node, and that was sufficient. In OLA, we find all such nodes.

Step 3. Now that we have a set of k-minimal nodes, these are compared in terms of their information loss and the node with the smallest information loss is selected as the globally optimal solution. If there are still more competing nodes, then we calculate which node has less risk of being identified [216]. Information loss is calculated with an information loss metric. There are three commonly used metrics: the Precision Metric (also known as the Prec Metric, Introduced by Sweeney), the Discernability Metric, and the Modified Discernability Metric.

A required property that the selected information loss metric should have is the monotonicity property (which the above stated metrics all have). An information loss metric having the monotonicity property means that information loss of any particular node in the generalization lattice is equal to or greater than of its parents [218]. Because of the monotonicity property, the k-minimal node with the smallest information loss must also have the smallest information loss among all k-anonymous nodes in the lattice.

To reduce the time required for the calculations and comparisons OLA maintains a list of k-minimal nodes that are potential solutions, which are k-anonymous nodes that have the lowest height within their generalization strategies. Whenever a node "N", is tagged as k-minimal, OLA checks if there are other k-minimal nodes above it on the generalization strategies that pass through "N". If there are, then these higher nodes are removed from the k-minimal solutions list and node N is added to the list [218].

A limitation of OLA is that it does enforce that information loss metrics are monotonic with respect to generalization strategies in the lattice, where other algorithms do not enforce this. However, the case has been made that even if an information loss metric is non-monotonic, it rarely exhibits this non-monotonic behavior in practice. To the extent that this empirical observation can be generalized broadly, other non-monotonic metrics, such as basic entropy or the original discernability metric, may still produce optimal results with OLA [218].

Aside from the clear trade-off from the computation time of a given query and the restrictions of how we model the information loss on the data, the benefit of a globally k-minimal data set will have the most usefulness and flexibility. So the data beyond the creation of it with OLA will be at its richest regardless of use case.

8.3.2.2. Differential Privacy

Differential privacy promises to protect individuals from any additional harm that they might face due to their data being in the private database. In particular, the risk of harm is not significantly greater when compared to not being in the private database. Although individuals may indeed face harm once the results of a differentially private mechanism have been released, differential privacy promises that the probability of harm was not significantly increased by their choice to participate. To satisfy the differential privacy constraint, a query-releasing mechanism needs to send a randomized query output to the analyst in a way such that the probability distribution of the query output does not differ too much, whether or not any individual record is in the database. In application, it attempts to do “two important things at once. First, it defines a measure of privacy, or rather, a measure of disclosure—the opposite of privacy. And second, it allows data producers to set the bounds of how much disclosure they will allow” in a given set of database queries [225], [226]. While differential privacy is an extremely strong guarantee, it does not promise unconditional freedom from harm [227]. At the same time, the algorithms used must satisfy both the goals of privacy and usefulness. Hence the most important question is how these two attributes must be traded off against each other [228].

A quick summary of two mechanisms used in differential privacy is provided below. These mechanisms serve as the building blocks and can also be combined to design more robust algorithms. The only thing to be kept in mind is that the combination of two differentially private algorithms must be differentially private itself.

Laplace Mechanism

The Laplace mechanism is a symmetric exponential distribution that offers a differentially private interface through which the data can be accessed, but still maintaining its privacy. It is useful for privately answering numerical valued queries independently. The query output is perturbed by adding random noise that conforms to the Laplace statistical distribution [229]. If the noise is sufficiently large, it will assist in preserving the differential privacy while the utility of the output will deteriorate. On the flip side, if the noise is too small, utility is increased but the privacy constraints are compromised [230]. Therefore the noise must be an optimal level to hide the contribution of any single participant, irrespective of the underlying database [231]. To consider also is that the trade-off between utility and privacy of any anonymization technique would depend largely on the attacker's background knowledge [229], [232].

The Laplace distribution depends on only one attribute called the "scale", which is directly proportional to its standard deviation or noisiness. It also depends on the privacy parameter ε reflecting the level of anonymization desired [233]. There's evidence available that by adding a random Laplace variable to a query, ε differential privacy is guaranteed. This ε is more of a privacy budget rather than purely statistical upper bound of the query [229]. As we query multiple times, we yield different ε, but we are concerned about the total ε, which tells about the maximum privacy release allowable. Once this privacy budget is exceeded, the user cannot query further. An informative simulation has also been provided by Anthony Tockar, which signifies that how uninformative the information could be after adding random noise [234]. So ultimately how much this scale should be set to? It will depend on the nature of the query itself. Smaller sensitivities of the query would mean less distortion [235]. Larger the sensitivities, more noise is needed to mask the data [233].

There are a set of steps (See Figure 49 below) needed for implementing differential privacy which are: 1) run query on database; 2) calculate the most influential observation; 3) calculate the Laplace noise distribution; 4) add Laplace noise distribution to the query results, and 5) publish perturbed query results [232].

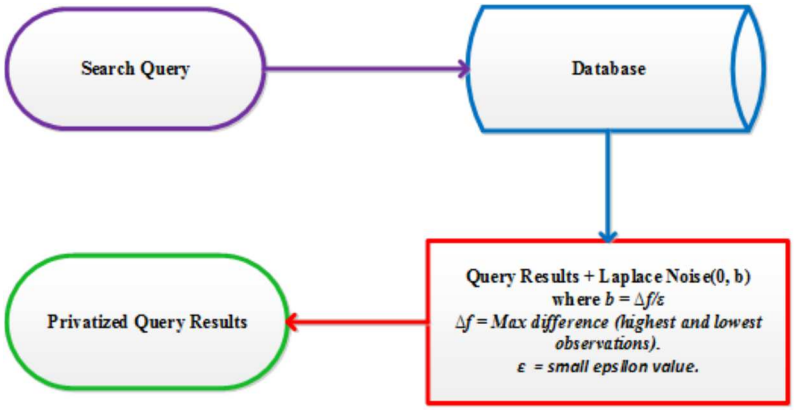

Figure 49 An overview of an interactive differential privacy technique [232].

Among advantages of the Laplace mechanism are the following: 1) queries from databases with low sensitivities can be answered with very little noise [231]. 2) this mechanism can be useful when it is not very easy to determine the sensitivity of the underlying query; 3) it can also be used to run an

iterative algorithm within the given privacy budget [236]. 4) Laplace noise can be scaled to obtain variants of DP [233]. As the Laplace mechanism is exponentially concentrated, it provides an excellent approximation to the true sum [231], [237].
On the flip side, this method is not suitable for categorical data and non-numeric valued queries. Also, it is specifically dependent on the perceived risk to the most different individual, which is referred to as the "sensitivity of the query". This mechanism cannot be used to answer correlated queries [229].

Exponential Mechanism
Exponential mechanism is the most general approach used to output non-numeric queries and is the natural building block for answering queries with arbitrary utilities (and arbitrary non-numeric range), while preserving differential privacy. The exponential mechanism was designed for situations in which we wish to choose the "best" response, but adding noise directly to the computed quantity can completely destroy its value [238]. This mechanism is said to capture all differential privacy mechanisms and could be most efficiently used for sampling from arbitrary sets [233], [238].

It is based on constructing a scoring function such as w: Inputs x Outputs → R and the utility of this mechanism will depend on the choice of scoring function. Every differentially private algorithm is captured by exponential mechanism by choosing the appropriate scoring function [239]. Given a query f with range P (the range P could consist of nominal, categorical or integer values), this mechanism assigns probabilities to the different elements in P based on their scores, where a higher score means a more desired output and hence a higher probability. If such a score function exists, then a differentially private output could be produced based on the sensitivity of the score function. It was argued similarly to the Laplace mechanism that if the sensitivity of the score function is low, then high quality output can be obtained [239], [240].

The exponential mechanism can often give strong utility guarantees, because it discounts outcomes exponentially quickly. It can also be an effective option for the non-interactive, or offline, case. The accuracy of this method is also linked to the rate at which the probability that the empirical distribution concentrates in a small ball around the true distribution [241]. It should be considered however, that the exponential mechanism can define a complex distribution over a large arbitrary domain, and so it may not be possible to implement the exponential mechanism efficiently when the range of the scoring function is super-polynomially large [238]. Also, in this mechanism all queries must be given up-front, while in contrast the Laplace mechanism answers queries independently. Additionally, the exponential mechanism is inefficient and sometimes it can destroy the values where systematically noise is added [241].

This mechanism has also been used in conjunction with the Multiplicative Weights approach. This combination (MWEM) is viewed as combining expert learning techniques (multiplicative weights) with an active learning component (via the exponential mechanism) and shows promising potential, including matching significantly higher theoretical accuracy guarantees for differentially private data analysis with linear queries, even for challenging case of restrictions on privacy for complex data and query sets, as well as improving on experimental error and overall simplification [237].

8.3.3. Limitations of existing solution and relevancy for Medolution

The analysis of the main anonymization techniques allows to conclude that one of the key current challenges is the computational complexity for large data-sets with large numbers of pseudo-identifiers, which is an often the case in healthcare domain. It has been illustrated that the creation of an efficient algorithm that guarantees a certain level of data anonymization and that can handle a generic data set with many quasi-identifiers is an on-going research problem. Two important points need to be considered in the context of Medolution however. First, differential privacy is an active research area, however to date has only been applied to a few operational systems. One of

them is the Census Bureau's "OnTheMap" website, which uses differential privacy to create reasonably accurate block-level synthetic census data, and Google's "Chrome" web browser, which uses randomized responses to collect aggregate statistics about the Windows process names running on the user's computer and the user's home page. In the last case, although the statistics are accurate in aggregate, the use of randomization makes it impossible to reliably determine a users' processes or home page. Second, the differential privacy, at least to date, comes at the cost decreased result accuracy. Thus, a recent research conducted to determine the impact of using differential privacy to create a statistical model for correlating genomic information and warfarin dosage based on clinical trial data has found that the models constructed using differential privacy would result in worse clinical outcomes for a significant number of patients compared to those models created without differential privacy (although this finding was only tested in simulation and not on actual clinical trial) [242]. Of the techniques reviewed, the Optimal Lattice Algorithm (OLA) shows the most promise for providing a globally optimal solution with a strong assurance of privacy while ensuring substantial result accuracy. However, due to the number and variety of the data sets and sources that Medolution intends to address, a standard OLA implementation will still face problems with computational complexity.

8.4. User-centric data privacy

8.4.1. Monitoring and traceability

- In order to ensure medical content traceability, within the MEDUSA project framework (See Chapter 10), two innovative technologies have proven their effectiveness (Figure 50):*watermarking*: active tracking technique, identifying the owner and the information leaking source thanks to some additional information inserted into the content to be tracked [243];
- *fingerprinting*: passive tracking technique, achieving the automatic tracking of unauthorized distribution thanks to some salient information extracted from the content to be tracked itself [244].

Note that rather than ensuring monitoring by themselves, watermarking and fingerprinting should be considered as generic tools providing the information to be processed by an external monitoring system. For instance, the fingerprinting detection can technically trigger any type of action (from content delivery denial to patient database updating). However, for medical data purposes, the monitoring system itself shall be design not only on a technical ground but shall also strictly observe to the current laws, rules and regulations. Moreover, as the legislative framework itself is still subject to ethical, societal and/or deontological controversial discussions, the patient itself should be able to express his/her own constraints related to his/her personal data processing. (See Figure 50 below).

Figure 50 Medical information tracking

Medical information tracking is achieved through two types of methods: (1) watermarking (left) consisting in inserting an invisible mark in the original content and by subsequently extracting it so as

to identify the owner and the leaking source and (2) fingerprinting (right) consisting in extracting some salient signature from the content and in matching it to a pre-processed database.

8.4.2. Watermarking

Digital watermarking can be defined as the process of embedding a pattern of information into a cover digital content (image, audio, video, etc.). The insertion of the mark is always controlled by some secret information referred to as a key. The subsequent watermark detection can serve to a large variety of applications, from property and/or integrity proof to augmented reality. Once watermarked, the host data can be transmitted and/or stored in a hostile environment, i.e. in an environment where changes attempting to remove the watermark are likely to occur. While the key should be kept secret (i.e. known only by the owner), the embedded information and even the embedding method can be public.

There are no universal requirements to be satisfied by all watermarking applications. Nevertheless, some general directions can be given for most of the applications. In order to be effective, the watermark should be perceptually invisible for a human observer (transparency) and its detection should be successful even when the watermarked content is attacked (robustness). Moreover, it should allow the insertion of a sufficient amount of information (data payload) required by the targeted application (e.g. a serial number identifying a user, a time stamp, etc.). The definitions for these general properties, as well for some additional practical features, are detailed below.

Performance criteria

Transparency

The notion of transparency is related to the perception (visual, auditory, etc.) of artefacts resulted from the insertion process. Watermarking should be imperceptible and invisible to a human observer (the embedded watermark should not affect the quality of the host data).

Robustness

Robustness is the ability of the mark to survive changes undergone by the host media. These changes (be they intentional or unintentional) define the set of attacks. The various possible attacks against watermarked video can be structured into four classes, according to the way they act: removal attacks, geometric attacks, cryptographic attacks, and protocol attacks.

The removal attacks try to make the watermark unreadable. This class includes attacks by noise addition, de-noising, transcoding quantization, etc. The geometric attacks aim to destroy the synchronization of the watermark. After such an attack, the watermark is still present in the video, but its location is unknown at the decoder. Rotations, curvatures, jitter of pixels individually considered or combined, fall into this category.

Protocol attacks aim to make watermark unusable by creating some ambiguities concerning the mark usage. Attacks by inversion and copy belong to this class. The former creates a false key so that by applying the detection procedure, the watermark indicates a different owner for the video.

The cryptographic attacks try to manage the watermark (detect/copy/insert a new one) without knowledge of the secret key. One example is represented by the brute-force search. Another example, known as the oracle attack, consists in creating an unmarked version of the signal by exploiting the response of a detector (assuming it is available). In any case, this type of attack is very restrictive in practice because of its complexity.

Fragility and semi-fragility

A watermark system is fragile to an attack when the watermark cannot be detected after slightest modifications generated by this attack.

A watermarking system is semi-fragile when both particular robustness and fragility properties are imposed to the system. Once the classes of allowed and non-allowed attacks have been defined based on the targeted application, the watermark must survive all manipulation belonging in the former class (the robustness), but it should be destroyed by the manipulations belonging to the latter (the fragility).

The required degree of each requirement presented above depends on the watermarking application. A watermarking application is effective when it ensures the functional balance of the three requirement degrees of transparency, robustness, data payload. However, some applications can require additional features, like cost minimization, constant bit-rate, format compliance, etc.

Data payload
This is the total amount of information (in bits) inserted into original content. According to the targeted applications, the specifications on this factor may be very different, from 64 bits per sequence for the identification of ownership up to hundreds of kilobits per frame for application of hyper-video.

Cost
The technical cost of the algorithm is also significant feature of any watermarking method. From this point of view, the complexity of the algorithm is the main criterion of practical acceptance.

Constant bit-rate
Watermarking method should not increase the size of the compressed data and the bit-rate, at least for constant bit-rate applications where the transmission channel has to be obeyed.

Watermarking for medical imaging
Digital watermarking proposed for medical imaging is a special subset of image watermarking. That particularity is relied on the critical use of medical imaging in patient diagnosis. Consequently, watermarked medical images should not differ perceptually form their originals, in the sense that the watermarking technique should not bias the diagnosis in any way.

Generally, three main classes of watermarking method were identified for medical images.
The first class includes methods that embed the mark within the region of non-interest (RONI) in order not to bias the diagnosis interpretation. Various works suggest that RONI refer in general to black background of the image; however, RONI can include grey level portion of little interest, hence leaves some more room for watermarking. Since there is no interference with interest medical image content, transparency is less strict; thus increasing the method data payload. Despite no interferences occur between the RONI and the data potentially used for the diagnosis, it has been shown that modifying black background by salt and pepper noisy pattern may bother medical interpretations. Therefore, the watermark information amplitude should be correctly set.

The second approach corresponds to reversible watermarking method. Once the embedded information is detected, the watermark is removed, allowing the reconstruction of the original image. Reversible methods are generally fragile and deployed for integrity verification. Methods which tried to achieve high robustness level introduced in the image visible salt-and-pepper noise.

While reversible watermarking facilitates the watermark information updating, the resulting watermarked images remain unprotected and may be moved and replaced by other marks. In addition, the mark must be removed before any interpretation, which may cause additional time delay for the physician.
The third approach consists in using jointly classical watermarking method and distortion minimization. In that case, the watermark replaces some image details by watermark information such as the least significant bit.

8.4.3. Fingerprinting

Image fingerprints can be best defined in relation with human fingerprints, as illustrated in Figure 51 below. While the human fingerprint can be seen as a human summary (a signature) that is unique for every person, the image fingerprint can be seen as some short image feature (e.g. a string of bits, colour histograms), which can uniquely identify that image. In practice, image fingerprints are

used just as human fingerprints: they are first computed and then searched for in a database, according to a given similarity measure.

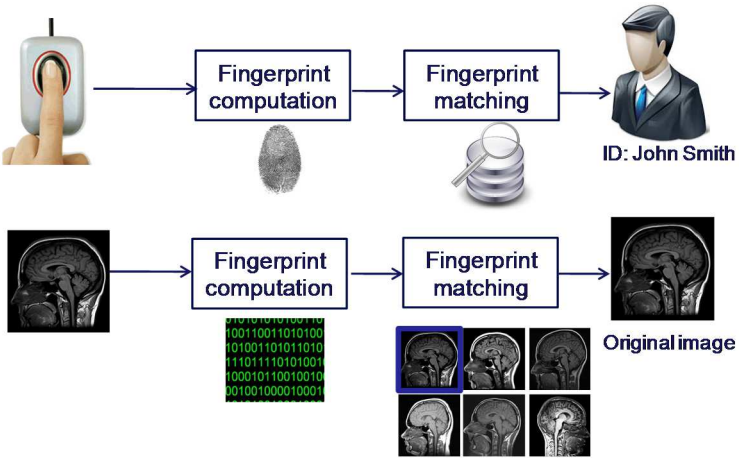

Figure 51 Human fingerprinting and medical image fingerprinting [244]

Performance criteria

Assume the case in which an image has its fingerprints computed and is searched for in the database. A correct answer in such a matching procedure is obtained when the same visual content is detected not only in its original image, but also in all its replica images; let *tp* be the number of such correct answers. A correct answer is also obtained when two images with different content are detected as different; let *tn* be the number of such situations. Practical fingerprinting methods may also come across with two types of matching errors. First, some image content existing in the database might not be retrieved; let *fn* be the number of such wrong decisions. Secondly, the detection procedure can also yield a false positive i.e. take some visual content for another one. Let *fp* be the number of such situations.

Image fingerprinting has two main properties:

- *Uniqueness:* fingerprints extracted from different content images should be different. This property is assessed by the probability of false alarm (P_{fa}) defined by the following formula:
  - $$P_{fa} = \frac{fp}{tp + fn + fp + tn}$$
- *Robustness to distortions*: fingerprints extracted from an original image and its replicas should be similar in the sense of the considered similarity metric. The robustness property is also quantified by the probability of missed detection (P_{md}), as defined below:
  - $$P_{md} = \frac{fn}{tp + fn + fp + tn}$$

On the one hand, an efficient fingerprinting method should ensure a low probability of false alarm (i.e. low probability of retrieving image which are neither the query nor its replicas) and low probability of missed detection (i.e. a low probability of not retrieving replica images of the query). According to the targeted application, additional functional properties, such as the database search efficiency can be set.

Fingerprinting for medical imaging

Under the medical imaging applicative framework, fingerprinting may serve three types of applications.

Image identification and retrieval: Given a very large database of images and a query image, the identification of such a query can pose complex challenges (e.g. time requirements, human observes). An image fingerprinting system enables the identification of a particular image by

computing its fingerprint and by efficiently querying it among the reference fingerprints without using human observers.

Authentication of multimedia content: Due to powerful software (e.g. Photoshop, Windows Movie Maker, Pinnacle) for multimedia manipulation, content became very easy to manipulate and therefore in many cases the originality of the content might need to be checked. An authentication system based on fingerprinting verifies the originality of the content and aims at detecting the malicious transformation. This is achieved by designing a fingerprint and a similarity metric able to detect any minor transformation in the query compared to the original version.

Copyright infringement prevention: In order to achieve copyright infringement-free image database by means of image fingerprinting, content owners would have to provide reference fingerprints to content sites, which would allow the identification of the images through the matching procedure. According to this identification and to the business or copyright rules established for each image, action could be taken, e.g. allow, filter or notify.

8.4.4. Limitations of existing solutions and relevancy for Medolution

Two classes user-centric medical data tracking techniques, watermarking and fingerprinting, can be of interest for Medolution as they provide:

- *complementarity in their approaches*: while the watermarking techniques ensure content personalisation without any noticeable artefacts, the fingerprinting can relay solely on the original content. In this way, a potential wide range of content personalisation can be targeted;
- *independence with respect to the way the medical data is acquired /processed/ transferred/ stored/ etc.*: note that these two techniques relate to the content itself and are robust (invariant) with respect to current day image processing transformations. This allows for the tracking to be performed as an added-value service, at any level inside the MEDOLUTION platform, without imposing any constraint for the rest of the workflow (either in processing or in formatting aspects);
- They preserve, at any point in the workflow, the value of the tracked content.

8.5. Limitations of existing solution and relevancy for Medolution

In respect to privacy and security solutions for IoT and Big Data systems in Healthcare the challenges are manifold:

- With regards to the application of *Cloud services/computing for healthcare solutions* such major challenges as integrity, confidentiality and availability need to be addressed. Approaching cloud security on different layers of a cloud system should allow for mitigation of related risks effectively. Various deployment models for cloud computing security should to be taken into consideration while working on the Medolution architecture layers to progressively address the main user security concerns.
- When *patient data needs to be accessed and processed* by Medolution architecture, depending on the security and privacy measures already been used by local care sites, appropriate security and privacy standards, in particular OASIS XACML and SAML for managing authorization and access control, IHE EUA and ATNA for authentication and audit logging and IHE BPPC for consent management can be implemented.
- With regards to the *privacy strategies relevant to the health data analytics*, in particular anonymization, one of the key current challenges is the computational complexity for large data-sets with large numbers of pseudo-identifiers. While there is a number of techniques currently implemented in commercial products and actively researched, their further

adaptation will be required to provide a globally optimal solution with a strong assurance of privacy required for Medolution while ensuring substantial result accuracy.

- With regards to the *user-centric medical data privacy techniques*, watermarking and fingerprinting can be utilized, as they are complementary and ensure independence with respect to the way the medical data is acquired /processed/ transferred/ stored/ etc. whilst preserving the value of the tracked content.

9. Medolution Innovations

Medolution's objective is to research and develop "*Smart Patient Environments"* to be jointly used by healthcare professionals and patients themselves.
A "Smart Patient Environment" is an environment consisting of multiple automated devices (such as vital sign sensors or artificial heart pumps) and systems (such as protocol management tools, EHR systems, diagnostic tools and data analytics applications), locally and remotely connected over smart networks, continuously monitored and automatically interpreted on the basis of personalised medical protocols for medical "alerting" and clinical decision support.
Current solutions target mainly a one-to-one data flow, where data input from a single sensor is being utilised as information towards a single specialised application, mostly for a single (or limited number of) patient(s). Medolution allows scaling to millions of patients in parallel, supporting information flows from a multitude of sensor devices to many specialised medical applications.

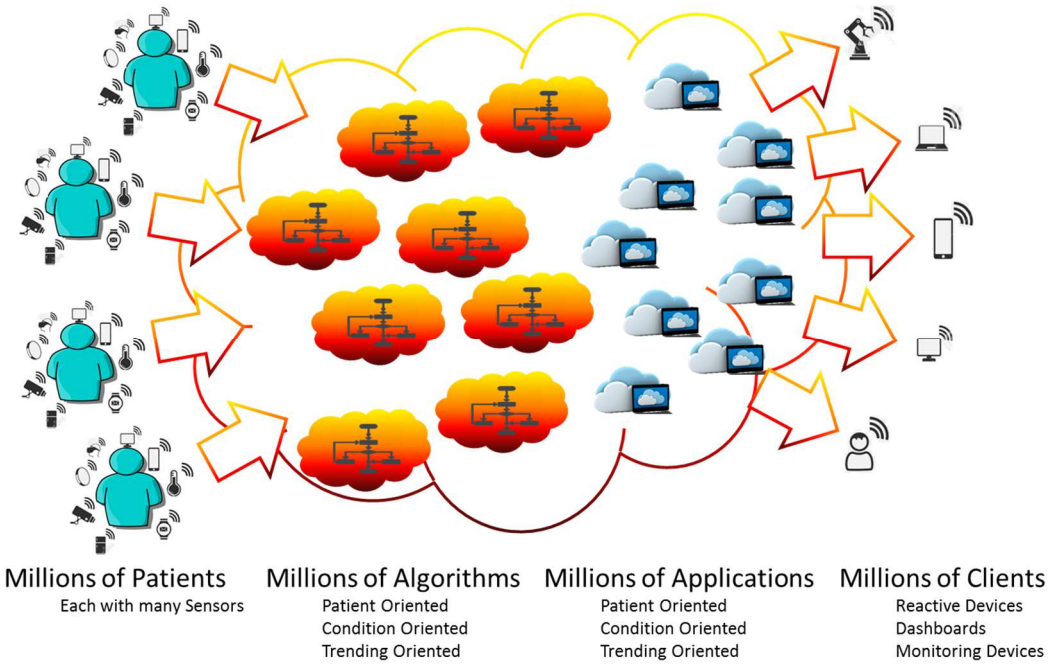

Figure 52 Medolution Innovation - Scaled to the Extend

Within Figure 52 above, the scalability needs for the Medolution platform is schematically expressed. Given millions of patients, each wearing many sensors, it is clear that the Medolution platform needs to provide a huge scaling ability in order to support the many individual, simultaneous data-streams from each of these sensors. This stresses for innovations on the Internet of Things with respect to Big Data Solutions for Healthcare (as described in paragraph 9.1 hereafter), on Dependability (paragraph 9.2 hereafter) and the Devices and IoT solutions (paragraph 9.3 hereafter). The innovations on the processing of these many simultaneous data-streams by millions of independent algorithms is described in paragraph 9.5 hereafter.

All these millions time millions of individual and simultaneous data-streams and processing algorithms demand innovation on the Automated Technical Management, which is described in detail in paragraph 9.4 hereafter.
The medical healthcare domain in which Medolution resides, demands for strict Privacy and Security Solutions. The Medolution innovations in this area are described in paragraph 9.6 hereafter.

9.1. IoT and Big Data Solutions for Healthcare

The goal of the Medolution Platform in terms of Big Data Platform will be to benefit from the Integrated Distributions advantages, e.g. reduced configuration steps and ready to use behaviour, while eliminating their drawbacks, e.g. allowing components choice, business culture evolution, and Cloud deployment. This last point will also permit to make such distribution usage available as online services. In addition, the focus will also be set on integrating advanced real-time Big Data processing technologies. This is addressed in Work Package 4 of the project by providing a platform that is:

- Shortening the Big Data processing latency and extracting insights on-the-fly to deliver them at the right time for the healthcare staff. To achieve this goal, Medolution will leverage the latest advances for stream processing and will enable experimentation with both the emerging in-memory processing and lambda architectures approaches for deploying new reactive applications.
- Allowing applications to be built by easily composing, configuring, deploying on the Cloud and managing several software pieces dealing with data collection, data mediation, data processing and analysis, etc. Such data access and processing services will be virtualized, “Cloudified” and made available in a catalogue.

9.2. Dependability

The major envisioned innovations of Medolution in the fields of dependability, architectural elements and modelling languages are:

- High dependability functions will be supported even by common customer devices like smartphones or home computers.
- Architectural elements will be developed that, on the one hand, are suited to the restrictions of small devices and, on the other hand, are capable to support the required grade of dependability.
- The quality of the systems will be ensured by an appropriate specific modelling which is based on enhancements of architecture modelling languages and supports rigorous analyses of the systems’ dependability properties.
- Particularly, the systems comprise high numbers of networked small devices providing sensor data to dependable functions and the big data processing in the cloud. The developed methods for dependability support, modelling and analysis will be able to cope with that IoT integration.

9.3. Devices and IoT Solutions for Healthcare System

The major planned Medolution innovations regarding to sensors and IoT are:

- Approach for the controlled development of dependable device systems based on requirements modelling and analysis and a service-oriented architecture model comprising special architectural design patterns.
- Development Guideline for interoperability between high risk medical devices and less reliable sensors such as Smartwatches or Smartphones with respect to Medical Device Certification and FDA approval.
- Development of an API to support fast and easy interoperability between medical devices, Big Dependable System, service suppliers, and other Sensors as a part of Health Data Ingestion stack. The API needs to respect the development guideline and therefore the development process. Additionally, the API will support the deployment of risk management, privacy and security as core features.

- The BDS itself interconnecting medical devices and sensors, analytics, Cloud storage, and less reliable sensors.
- Integration of cohesive sets of permanent sensors on the body as well as in the environment of a patient for continuous real-time decision support in a secure and privacy respecting way.

9.4. Automated Technical Management for Medical Systems

The Medolution technological innovation is an elaborated approach to technical management of medical systems. In contrast to existing policy-based management solutions, it concentrates on the tasks of clearly management functional areas (fault, configuration, accounting, performance, and security) and does not impair the application logic. Concentrating on these functional areas should provide for supporting and facilitating **dependable behaviour** of medical systems (BDMS), discussed in Chapter 2. In particular, the fault tolerance patterns identified in Chapter 5 as a main mechanism for achieving the desired behaviour can be used to support the management of the Medolution system.

A sophisticated automated runtime management is supposed to manage medical systems in accordance with technical requirements. The technical requirements are to be derived from the previously defined medical domain specific requirements. At this point, the model support comes to the fore. Model-based management not only assists by automation of the refinement process, but also facilitates further verification and formal correctness proof of the system. The system developers are provided with basic development paradigms, architectural patterns, elements and structures that allow **for medical product certification** to be supported already during the development process. This is of a special interest taking into consideration the variety of regulations, standards and regulatory cconstraints applied to the medical and healthcare devices and systems presented in Chapters 3 and 8. In particular, it will be addressed in the context of the high risk medical devices (for example, LVAD).

9.5. Healthcare Data Exploitation

9.5.1. Healthcare Data Integration

Medolution aims to provide a core platform, which will be able to ingest EHR data using well-established international standards. For Europe, HL7 CDA (CCD) and CEN/ISO 13606 are the two widely adopted EHR protocol standards, therefore; Medolution will implement the necessary mechanisms to ingest EHR data from these formats into the Big Data platform for further data analysis. This will lead to whole data analytics combining EHR data with data collected from devices, wearables and other sources within the defined use cases.

For large datasets, continuous data, or data residing in raw format, additional data sharing mechanisms are needed to achieve automated linking. In order to achieve interoperability, for each candidate technology, interfaces should be defined based on the specific requirements of public health data. The Medolution project will analyse these requirements within the project scope, specify the appropriate ways and define the interfaces for each of them based on the latest open technologies in Big Data domain. This way, health data can be shared in the most appropriate and secure way with the applications that need to process the data.

9.5.2. Healthcare Data Analytics

Data analytics
Developing efficient Medolution data analytics algorithms that can do classification, regression or clustering depends on how we can take into account the dimensionality of the problems, data over-

fitting, repeated measures, missing values and missing variables, data redundancy or incidental endogeneity.

- *Dimensionality Reduction:* To solve the problem of over-fitting data in the Medolution data lake, traditional approaches can be used and optimized by using heuristics that go beyond the basic technique of dividing the data sets into two parts (one set to train the models while the remaining is used for final evaluation of the model). In general, eliminating non-corresponding data can be managed only by data scientists helped by clinicians.
- *Handling missing values and missing variables*: These are two issues that must be solved in Medolution by using an optimized approach. Machine learning algorithms must cope with these missing elements and appropriate approaches must be proposed by involving clinical experts as well as patients to identify and capture missing variables.
- *Removing Endogeneity*: Exogeneity is an aspect of small dataset and enhances the regression performance because the explanatory variables, or predictors, are independent of the residual term. But in big data sets the regression performance can be diminished due to the incidental endogeneity, which refers to a genuine relationship between variables and the error term in contrast to spurious correlation. Therefore, the validity of most of the statistical methods used in regression depends on how endogeneity can be removed [174]. One of the approaches that will be planned to remove endogeneity in Medolution is mathematical techniques such as structural equations.

Medical image analytics

- Integrate clinical (image) data with extra-mural sensor data enabling advanced patient specific monitoring and clinical decision making.
- *Radiomics:* Improvements in medical imaging technology have resulted in a higher accuracy and more versatile imaging. Capitalising on this, advanced methods should be applied to extract more information from these images, such as functional information (like perfusion rather than solely morphological data). As a result, more personalised care is made possible. Medolution will enable high-throughput extraction of quantitative descriptions and combine these with additional sensors. This unique approach will be implemented and validated for selected patient populations, and contributes to the medical innovation objectives.
- *Flexible integration of new knowledge:* this involves innovation to include existing information, such as medical evidence and best practice benchmarks, in data and image analysis as well as in decision support and medical image visualisation, assuming new technologies for data and text mining combined with predictive logic.

9.5.3. Healthcare Decision Support Systems

The decision support system in Medolution will build on capabilities developed in previous ITEA research projects (Edafmis, Medusa – ITEA projects). These capabilities contain basically two components: a so called Rule Editor and a Real Time Rules Engine. Among the innovative aspects that bring our proposed DSS beyond the state of the art is its ability to offer following functionalities and features in one unified framework:

- Declarative and human comprehensive way of describing expert knowledge and designing decision rules via a Rule Editor: The Rule Editor allows users that are not programmers, to define algorithms in human language that are automatically transformed into computer code that can be executed by the Real Time Rules Engine.
- Scalable cloud-based real time rule engine: The Real Time Rules Engine runs in the cloud and is connected to the data sources of a user (vital signs, contextual parameters, EHR data, etc.); this is multiplied for each user, because each user has basically his or her own

personal infrastructure in the cloud infrastructure. In this way we can handle millions of users in parallel, in real-time.

- Integration of the rule editing capabilities in the doctor's workstation.
- Flexible reasoning framework uniting different paradigms: deterministic, fuzzy and probabilistic: The Rule Editor allows end users to define the logic that we need in the project: you can define rules with all kinds of parameters. Values can be absolute numeric, can be trends, can be fuzzy, and can have a time factor. You can define single rules, but also sets of rules that are connected with the AND factor. Rules or sets of rules include an Alert or another Trigger as a THEN clause. Per definition Single Rules or Sets of Rules are seen as parallel rules because the Real Time Rules Engine executes them in parallel. So in fact they are OR connected. The context parameters can be described, defined and executed real time in this system.

9.6. Privacy and Security Solutions for IoT and Big Data Systems

The major Medolution advancements and innovations regarding the anonymization will be:

1. A pragmatic approach towards the computational complexity based on the correlation between the datasets attributes and determined acceptable de-identification risk threat holds to enable dynamic and improved application of the OLA anonymization algorithm.
2. An improved method for traversing the OLA lattice to identify the globally optimal solution.
3. Development of a framework for a data custodian and a data analyst (consumers of the Medolution data) to publish and utilize the meta-data of the available data sets to perform the analysis.
4. Development of an API to support fast and easy dynamic application of the improved OLA anonymization algorithm for large datasets with large numbers of pseudo-identifiers.
5. A framework for international, cross-legal privacy and security law experimentation.

10. Conclusions

This document describes in depth the results of a survey of components that impact the areas of innovation by Medolution. The investigations have focused on the following areas, and can be found in the corresponding chapters:

- Internet of Things and Big Data Solutions for Healthcare
- Dependability
- Devices and the IoT solutions for Healthcare
- Automated Technical Management for Medical Systems
- Healthcare Data Exploitation
- Privacy and Security Solutions for IoT and Big Data Systems

In addition, the results of a survey of two other related areas are presented in the Appendixes to this document in order to complement the state of the technical art analysis by and overview of relevant to Medolution:

- Regulatory privacy and security constraints
- European research projects in healthcare data processing

The aim of the project is to build a proof of concept Medolution core platform. This core platform will support the use cases as defined in the Medolution Full Project Proposal, integrating:

- Control of heterogeneous devices
- Decision support and visualisation of real time and long term image and data analytics
- Cloud management upon the cloud infrastructure
- Virtual workspaces accessing the hospital infrastructure.

This proof of concept to the Medoluton core platform is intended to target the objectives as stated in the Full Project Proposals, involving:

- Dependable integration of heterogeneous devices (Objective T1 & T2)
- Big data analysis for real-time and long term healthcare purposes (Objective T3, T5 & P1)
- Deployment on the cloud (Objective T4 & P2)
- Case studies and demonstrators showing feasibility in healthcare settings (all Objectives)

Chapter 9 offers a concise, retrospective view on these aspects.

The investigations on the impacted areas as described within this document give the participating partners a solid base ground to start the innovations intended within the Medolution project.

11. Glossary

AA	Attribute Authorities
AADL	Architecture Analysis and Design Language
ACS	Access Control Service
AD	Activity Diagram
ADL	Archetype Description Language
ADL	Architecture Description Language
AES	Advanced Encryption Standard
AMPLab	Algorithms, Machines and People Lab (of Berkeley University)
ASTM	American Society for Testing and Materials
ATNA	Audit Trail and Node Authentication (Integration Profile)
AUI	Abstract User Interface
AWS	Amazon Web Service
BDHS	Big Dependable Healthcare System
BDMS	Dependable Behaviour of Medical Systems
BDPaaS	Big Data Platform as a Service
BDD	Block Definition Diagram
BPPC	Basic Patient Privacy Consents (Integration Profile)
BRIDG	Biomedical Research Integrated Domain Group
CaaS	Compute-as-a-Service Cloud
CEDD	Clinical Element Data Dictionary
CEM	Clinical Element Models
CEN/TC	Health Informatics committee of the European Committee for standardization
CCD	Continuity of Care Document
CCR	Continuity of Care Record
CCOW	Clinical Context Object Workgroup
CDA	Clinical Document Architecture
CDM	Common Data Model
CDS	Clinical Decision Support
CDSS	Clinical Decision Support Systems
CIGs	Computer Interpretable Guidelines
CIM	Common Information Model
CDASH	Clinical Data Acquisition Standards Harmonization
CDE	Common Data Element
CDISC	Clinical Data Interchange Standards Consortium
CPMSes	Cyber-Physical Medical System
CTMC	Continuous-time Markov chains
CWM	Common Warehouse Metamodel™
DAM	Domain Analysis Model
DDOS	Distributed Denial of Service
DICOM	Digital Imaging and Communications in Medicine
DMTF	Distributed Management Task Force
DPWS/WS4D	Device Profile for WebServices / WebService for Devices

DoS	Denial of Service
EC2	Elastic Compute Cloud
EHRs	Electronic Health Records
eMix	Electronic Medical Information Exchange
EMF	Eclipse Modelling Framework
EMR	Elastic Map Reduce
EMV2	Error-Model-Annex
ETL	Extract, Transform, and Load (data)
EUA	Enterprise User Authentication (Integration Profile)
IHE	Integrating the Healthcare Enterprise
IoT	Internet of Things
ISO/TC	International Organization for Standardization /Technical Committee
FHE	Fully Homomorphic Encryption
FHIM	Federal Health Information Model
FHIR	Fast Healthcare Interoperability Resources
FUI	Final User Interface
GERA	Generalized Enterprise Reference Architecture
GSPN	Generalized stochastic Petri nets
HCI	Human Computer Interaction
HDP	Health Device Profile
HIS	Hospital Information System
HITSP	Health Information Technology Standards Panel
HL7	Health Level Seven
IaaS	Infrastructure as a Service
IBD	Internal Block Diagram
IETF	Internet Engineering Task Force
IMC	Interactive Markov chains
JMEDS	Java Multi Edition DPWS Stack
LE	Bluetooth Low Energy
LDAP	Lightweight Directory Access Protocol
LPWAN	Low Power Wide Area Network
VLAD	Left Ventricular Assist Device
MBE	Model-based system engineering
MCAP	Multi-Channel Adaptation Protocol
MOF	Meta Object Facility
MPC	Multi-Party Computation
MTTF	Mean Time To Failure
MTTR	Mean Time To Repair
MWEM	Multiplicative Weights Exponential Mechanism
NBDRA	The NIST Big Data Reference Architecture
OCL	Open Constraint Language
OID	Object Identifier
OLA	Optimal Lattice Algorithm
OWL	Web Ontology Language
PaaS	Platform as a Service layer
PDP	Policy Decision Point

PAP	Policy Administration Point
pbdR	Programming with Big Data in R
PEP	Policy Enforcement Point
PCIM	Policy Core Information Model
PRA	Patient Record Architecture
RBAC	Role Based Access Control
RD	Requirement Diagram
RDS	Relational Database Service
RIM	Reference Information Model
RM-ODP	Reference Model for Open Distributed Processing
RONI	Region of Non-Interest
OLA	Optimal Lattice Anonymization algorithm
OSGi Alliance	Open Service Gateway initiative
SaaS	Software as a Service
SAML	Security Assertion Markup Language
SAN	Stochastic activity networks
SAVE	System AVailability Estimator
SDLC	System Development Life Cycle
SDO	Standard Development Organization
SDTM	Study Data Tabulation Model
SFM	Structure-function model
SPA	Stochastic process algebra
SPEM	Software & Systems Process Engineering Metamodel™
SSH	Secure Shell
SODA	Service-Oriented Device Architecture
SysML	Systems Modelling Language
TOGAF	The Open Group's Architecture Framework
TMR	Ttriple Modular Redundancy
ToC	Transitions of Care Initiative
TSDB	Time Series Database
TTP	Trusted Third Party
UML	Unified Modelling Language
URN	Uniform Resource Name
VM	Virtual Machine
VPN	Virtual Private Network
WAF	Web Application Firewalls
WBEM	Web-Based Enterprise Management
WS-Trust	Web Services Trust Language
XACML	eXtensible Access Control Markup Language
XML	Extensible Markup Language
XSPA	Cross-Enterprise Security and Privacy Authorization
XSS	Cross Site Scripting
XUA	Cross-Enterprise User Assertion (Integration Profile)

12. References

[1] NIST Big Data Public Working Group, “DRAFT NIST Big Data Interoperability Framework: Volume 6, Reference Architecture.” NIST.
[2] NIST Big Data Public Working Group and Technology Roadmap Subgroup, “NIST Big Data Interoperability Framework: Volume 7, Standards Roadmap. NIST Special Publication 1500-7,” National Institute of Standards and Technology (NIST), Sep. 2015.
[3] IoT-A Consortium, “Architectural Reference Model for IoT.” IoT-A Consortium.
[4] “Time series database,” *Wikipedia, the free encyclopedia*. 28-Aug-2016.
[5] “Elasticsearch | Elastic.” [Online]. Available: https://www.elastic.co/products/elasticsearch. [Accessed: 05-Aug-2016].
[6] “Logstash | Elastic.” [Online]. Available: https://www.elastic.co/products/logstash. [Accessed: 05-Aug-2016].
[7] “Kibana: Explore, Visualize, Discover Data | Elastic.” [Online]. Available: https://www.elastic.co/products/kibana. [Accessed: 05-Aug-2016].
[8] “Welcome to Apache™ Hadoop®!” [Online]. Available: http://hadoop.apache.org/. [Accessed: 05-Aug-2016].
[9] “Apache Mahout: Scalable machine learning and data mining.” [Online]. Available: http://mahout.apache.org/. [Accessed: 05-Aug-2016].
[10] “R: The R Project for Statistical Computing.” [Online]. Available: https://www.r-project.org/. [Accessed: 05-Aug-2016].
[11] “Apache Spark™ - Lightning-Fast Cluster Computing.” .
[12] “Python Data Analysis Library — pandas: Python Data Analysis Library.” [Online]. Available: http://pandas.pydata.org/. [Accessed: 01-Sep-2016].
[13] “Spark Streaming | Apache Spark.” [Online]. Available: http://spark.apache.org/streaming/. [Accessed: 05-Aug-2016].
[14] “Apache Storm.” [Online]. Available: http://storm.apache.org/. [Accessed: 05-Aug-2016].
[15] “Samza.” [Online]. Available: http://samza.apache.org/. [Accessed: 05-Aug-2016].
[16] “Apache Apex.” [Online]. Available: http://apex.apache.org/. [Accessed: 05-Aug-2016].
[17] “Introducing Kafka Streams: Stream Processing Made Simple.” [Online]. Available: http://www.confluent.io/blog/introducing-kafka-streams-stream-processing-made-simple. [Accessed: 05-Aug-2016].
[18] “Apache Flink: Scalable Batch and Stream Data Processing.” [Online]. Available: https://flink.apache.org/. [Accessed: 05-Aug-2016].
[19] “Beam Incubation Status - Apache Incubator.” [Online]. Available: http://incubator.apache.org/projects/beam.html. [Accessed: 05-Aug-2016].
[20] N. Marz, “How to beat the CAP theorem - thoughts from the red planet - thoughts from the red planet,” 13-Oct-2011. [Online]. Available: http://nathanmarz.com/blog/how-to-beat-the-cap-theorem.html. [Accessed: 23-Nov-2016].
[21] N. Marz and J. Warren, *Big Data: Principles and Best Practices of Scalable Realtime Data Systems*, 1st ed. Greenwich, CT, USA: Manning Publications Co., 2015.
[22] D. Jebaraj, “Lambda Architecture: Design Simpler, Resilient, Maintainable and Scalable Big Data Solutions,” *InfoQ*, 12-Mar-2014. [Online]. Available: https://www.infoq.com/articles/lambda-architecture-scalable-big-data-solutions. [Accessed: 23-Nov-2016].
[23] Hortonworks community, “Defining Enterprise Hadoop - Hortonworks.” [Online]. Available: http://hortonworks.com/blog/defining-enterprise-hadoop/.
[24] “The modern platform for data management and analytics - Cloudera.” [Online]. Available: http://www.cloudera.com/. [Accessed: 05-Aug-2016].
[25] “The modern platform for data management and analytics - Cloudera.” .
[26] “MapR Converged Data Platform | MapR.” [Online]. Available: https://www.mapr.com/products/mapr-converged-data-platform. [Accessed: 05-Aug-2016].
[27] “Pivotal Big Data Suite | Big Data | Pivotal.” [Online]. Available: https://pivotal.io/big-data/pivotal-big-data-suite. [Accessed: 05-Aug-2016].
[28] “HP HAVEn for Big Data.” .
[29] B. Gourley, “Top Five Nominees for the 2012 Government Big Data Solutions Award,” *Cloudera Engineering Blog*, 13-Nov-2012. [Online]. Available:

http://blog.cloudera.com/blog/2012/11/top-five-nominees-for-the-2012-government-big-data-solutions-award/. [Accessed: 01-Sep-2016].
[30] "Google Developing 'Brillo' Software for Internet of Things," *The Information*. [Online]. Available: https://www.theinformation.com/Google-Developing-Brillo-Software-for-Internet-of-Things. [Accessed: 01-Sep-2016].
[31] E. Dubrova, *Fault-Tolerant Design*, 2013th ed. New York: Springer, 2013.
[32] A. A. Ucla, A. Avizienis, J. Laprie, and B. Randell, *Fundamental Concepts of Dependability*. 2001.
[33] R. Hanmer, *Patterns for Fault Tolerant Software*, 1st ed. Wiley, 2013.
[34] I. Sommerville, *Software Engineering*, 10 edition. Boston: Pearson, 2015.
[35] "Dependability Modelling and Evaluation." [Online]. Available: https://depend.cs.uni-sb.de/index.php?483. [Accessed: 11-Aug-2016].
[36] "ISO/IEC/IEEE Systems and software engineering – Architecture description," *ISOIECIEEE 420102011E Revis. ISOIEC 420102007 IEEE Std 1471-2000*, pp. 1–46, Dec. 2011.
[37] K. Echtle, W. Brauer, and G. Goos, *Fehlertoleranzverfahren*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1990.
[38] Object Management Group (OMG), "OMG Systems Modeling Language (OMG SysML™)." Sep-2015.
[39] S. Friedenthal, A. Moore, and R. Steiner, *A Practical Guide to SysML: The Systems Modeling Language*. Morgan Kaufmann, 2014.
[40] "OMG SysML." [Online]. Available: http://www.omgsysml.org/. [Accessed: 26-Jul-2016].
[41] F. Kordon, J. Hugues, A. Canals, and A. Dohet, Eds., *Embedded Systems: Analysis and Modeling with SysML, UML and AADL*, 1 edition. London : Hoboken, NJ: Wiley-ISTE, 2013.
[42] S. A. E. Aerospace, "SAE AS5506B: Architecture Analysis & Design Language (AADL) standard document." 2012.
[43] P. H. Feiler, B. A. Lewis, and S. Vestal, "The SAE Architecture Analysis #x00026; Design Language (AADL) a standard for engineering performance critical systems," in *2006 IEEE Conference on Computer Aided Control System Design, 2006 IEEE International Conference on Control Applications, 2006 IEEE International Symposium on Intelligent Control*, 2006, pp. 1206–1211.
[44] P. H. Feiler and D. P. Gluch, *Model-Based Engineering with AADL: An Introduction to the SAE Architecture Analysis & Design Language*, 1st ed. Addison-Wesley Professional, 2012.
[45] P. H. Feiler, D. P. Gluch, and J. J. Hudak, "The architecture analysis & design language (AADL): An introduction," DTIC Document, 2006.
[46] J. Delange and P. Feiler, "Architecture fault modeling with the aadl error-model annex," in *2014 40th EUROMICRO Conference on Software Engineering and Advanced Applications*, 2014, pp. 361–368.
[47] A. SAE, "SAE Architecture Analysis and Design Language (AADL) Annex Volume 3: Annex E: Error Model Annex." 23-Jan-2013.
[48] B. Larson, J. Hatcliff, K. Fowler, and J. Delange, "Illustrating the AADL Error Modeling Annex (V.2) Using a Simple Safety-critical Medical Device," in *Proceedings of the 2013 ACM SIGAda Annual Conference on High Integrity Language Technology*, New York, NY, USA, 2013, pp. 65–84.
[49] I. Y. Jung, G.-J. Jang, J.-M. Yang, and J. Yoo, "Design of a Situation Aware Service for Internet of Things," *Int. J. Distrib. Sens. Netw.*, vol. 11, no. 9, p. 641312, Jan. 2015.
[50] L. CHEN and A. Avizienis, "N-VERSION PROGRAMMING : A FAULT-TOLERANCE APPROACH TO RELIABILITY OF SOFTWARE OPERATION," IEEE, Reprinted from FTCSB 1978, pp. 3–9, 1995.
[51] "DIGITAL HEALTH: Q3 2014 STATE OF THE INDUSTRY." Personal Connected Health Alliance, Oct-2014.
[52] M. Freeman and J. Goldstein, "BIOMEDICAL DEVICE INTEGRATION Getting Vitals Where and When Its Needed Most Industrial Case Study on Sentara Healthcare." DIVURGENT, 2012.
[53] "Design Guidelines." Continua Alliance, 2011.
[54] "ISO/IEEE 11073-00103:2015 - Health informatics -- Personal health device communication -- Part 00103: Overview," *ISO*. [Online]. Available: http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?ics1=35&ics2=240&ics3=80&csnumber=64941. [Accessed: 10-Jun-2016].

[55] "Merge eMix™." [Online]. Available: http://www.merge.com/Solutions/Interoperability/Merge-eMix%E2%84%A2.aspx. [Accessed: 14-Sep-2016].
[56] "Medical-Grade Remote Care." [Online]. Available: http://www.qualcommlife.com/mobile-medical-solutions. [Accessed: 13-Sep-2016].
[57] D. M. West, "Improving Health Care through Mobile Medical Devices and Sensors," 2013.
[58] N. Stylianides, M. D. Dikaiakos, H. Gjermundrød, G. Panayi, and T. Kyprianou, "Intensive care window: real-time monitoring and analysis in the intensive care environment," *IEEE Trans. Inf. Technol. Biomed. Publ. IEEE Eng. Med. Biol. Soc.*, vol. 15, no. 1, pp. 26–32, Jan. 2011.
[59] M. Blount *et al.*, "Real-time analysis for intensive care: development and deployment of the artemis analytic system," *IEEE Eng. Med. Biol. Mag. Q. Mag. Eng. Med. Biol. Soc.*, vol. 29, no. 2, pp. 110–118, Apr. 2010.
[60] Technical Marketing Workgroup 1.0, "A technical overview of LoRa® and LoRaWAN™." Nov-2015.
[61] "ISO/IEEE 11073-10407:2010 - Health informatics -- Personal health device communication -- Part 10407: Device specialization -- Blood pressure monitor," *ISO*. [Online]. Available: http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?ics1=35&ics2=240&ics3=80&csnumber=54573. [Accessed: 10-Jun-2016].
[62] "ISO/IEEE 11073-10408:2010 - Health informatics -- Personal health device communication -- Part 10408: Device specialization -- Thermometer," *ISO*. [Online]. Available: http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?ics1=35&ics2=240&ics3=80&csnumber=54309. [Accessed: 10-Jun-2016].
[63] "ISO/IEEE 11073-10415:2010 - Health informatics -- Personal health device communication -- Part 10415: Device specialization -- Weighing scale," *ISO*. [Online]. Available: http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?ics1=35&ics2=240&ics3=80&csnumber=54310. [Accessed: 10-Jun-2016].
[64] "ISO/IEEE 11073-10417:2014 - Health informatics -- Personal health device communication -- Part 10417: Device specialization -- Glucose meter," *ISO*. [Online]. Available: http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?ics1=35&ics2=240&ics3=80&csnumber=61896. [Accessed: 10-Jun-2016].
[65] "What Is AWS IoT?" [Online]. Available: http://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html. [Accessed: 13-Sep-2016].
[66] "AWS IoT Gateway Review - Pros & Cons - Internet of Things," *Internet of Things at Home*, 29-Feb-2016. [Online]. Available: http://iotathome.org/2016/02/amazon-web-services-iot-device-gateway/. [Accessed: 26-Aug-2016].
[67] "Watson Internet of Things." .
[68] "Kura," 18-Oct-2016. [Online]. Available: http://www.eclipse.org/kura/.
[69] "Om2m." 18-Oct-2016.
[70] "PlatformIO." [Online]. Available: http://platformio.org/. [Accessed: 18-Oct-2016].
[71] "Kaa IoT Platform," 18-Oct-2016. [Online]. Available: http://www.kaaproject.org/.
[72] "Compose. Collaborative Open Market to Place Objects at your Service," 18-Oct-2016. [Online]. Available: http://www.compose-project.eu/.
[73] Ricker et al, "Service-Oriented Device Architecture (SODA)." eclipsecon, 2007.
[74] "Principles, Architecture and Horizontal Services." The OSAmI Consortium, 19-Aug-2010.
[75] H.-G. Hegering, S. Abeck, and B. Neumair, *Integrated Management of Networked Systems: Concepts, Architectures, and Their Operational Application*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998.
[76] "Recommendation X.700 (09/92): Management Framework for Open Systems Interconnection (OSI) for CCITT Applications," International Telecommunication Union, 1992.
[77] K. McCloghrie, D. Perkins, and J. Schoenwaelder, *Structure of Management Information Version 2 (SMIv2)*. IETF, 1999.
[78] D. C. Robinson and M. S. Sloman, "Domains: a New Approach to Distributed System Management," in *Workshop on the Future Trends of Distributed Computing Systems in the 1990s*, 1988, pp. 154–163.
[79] M. Sloman and J. Moffett, "Domain Model of Autonomy," in *Proceedings of the 3rd Workshop on ACM SIGOPS European Workshop: Autonomy or Interdependence in Distributed Systems*, Cambridge, United Kingdom, 1988, pp. 1–4.
[80] B. Moore, *Policy Core Information Model (PCIM) Extensions*. IETF, 2003.

[81] B. Moore, E. Ellesson, J. Strassner, and A. Westerinen, *Policy Core Information Model – Version 1 Specification*. IETF, 2001.
[82] A. Westerinen *et al.*, *Terminology for Policy-Based Management*. IETF, 2001.
[83] R. Kamal, M. S. Siddiqui, H. Rim, and C. S. Hong, "A Policy based Management Framework for Machine to Machine Networks and Services," in *13th Asia-Pacific Network Operations and Management Symposium (APNOMS)*, 2011, pp. 1–4.
[84] Y. Snir, Y. Ramberg, J. Strassner, R. Cohen, and B. Moore, *Policy Quality of Service (QoS) Information Model*. IETF, 2003.
[85] N. Matthys and W. Joosen, "Towards Policy-based Management of Sensor Networks," in *Proceedings of the 3rd International Workshop on Middleware for Sensor Networks*, Leuven, Belgium, 2008, pp. 13–18.
[86] J. E. L. D. Vergara, V. A. Villagrá, and J. Berrocal, "Semantic Management: Advantages of Using an Ontology-based Management Information Meta-model," in *Proceedings of the HP Openview University Association Ninth Plenary Workshop (HP-OVUA'2002)*, 2002, pp. 11–13.
[87] B. C. Williams, P. P. Nayak, and U. Nayak, "A Model-based Approach to Reactive Self-Configuring Systems," in *AAAI-96*, 1996, pp. 971–978.
[88] M. Garschhammer, R. Hauck, B. Kempter, I. Radisic, H. Roelle, and H. Schmidt, "The MNM Service Model – Refined Views on Generic Service Management," *J. Commun. Netw.*, vol. 3, no. 4, pp. 297–306, Dec. 2001.
[89] H. Foster, S. Uchitel, J. Magee, and J. Kramer, "Model-based verification of Web service compositions," in *18th IEEE International Conference on Automated Software Engineering*, 2003, pp. 152–161.
[90] J. Bezivin, "On the unification power of models," *Softw. Syst. Model.*, vol. 4, no. 2, pp. 171–188, 2005.
[91] Object Management Group, *Model Driven Architecture (MDA). MDA Guide Rev. 2.0.* Object Management Group, 2006.
[92] DMTF, *Common Information Model (CIM) Metamodel. Specification, DMTF Standard, Version: 3.0.0*. DMTF, 2012.
[93] Object Management Group, *OMG Unified Modeling Language (OMG UML), Superstructure. Version 2.4.1*. 2011.
[94] "CIM | DMTF." [Online]. Available: https://www.dmtf.org/standards/cim. [Accessed: 03-Aug-2016].
[95] DMTF, *Common Information Model (CIM) Schema. Specification, DMTF Standard, Version 2.45.0*. 2016.
[96] DMTF, *Common Information Model (CIM) Infrastructure. Specification, DMTF Standard, Version 2.7.0*. 2012.
[97] DMTF, *Managed Object Format (MOF). Specification, DMTF Standard, Version 3.0.0*. 2012.
[98] "WBEM | DMTF." [Online]. Available: https://www.dmtf.org/standards/wbem. [Accessed: 03-Aug-2016].
[99] R. Lenz and M. Reichert, "IT Support for Healthcare Processes - Premises, Challenges, Perspectives," *Data Knowl Eng*, vol. 61, no. 1, pp. 39–58, Apr. 2007.
[100] M. Burwitz, H. Schlieter, and W. Esswein, "Agility in medical treatment processes – A model-based approach," in *Modellierung*, 2012, vol. 201, pp. 267–279.
[101] S. Andreassen *et al.*, "Model-Based Medical Decision Support – A Road to Improved Diagnosis and Treatment?," in *15th Nordic-Baltic Conference on Biomedical Engineering and Medical Physics*, 2011, vol. 34, pp. 257–260.
[102] *Model-Based Engineering of Embedded Systems: The SPES 2020 Methodology*. Klaus Pohl and Harald Hönninger and Reinhold Achatz and Manfred Broy, 2012.
[103] A. Banerjee, S. K. S. Gupta, G. Fainekos, and G. Varsamopoulos, "Towards Modeling and Analysis of Cyber-physical Medical Systems," in *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies*, Barcelona, Spain, 2011, p. 154:1–154:5.
[104] A. Ray, R. Jetley, P. Jones, and Y. Zhang, "Model-based Engineering for Medical-Device Software," *Biomed Instrum Technol.*, vol. 44, no. 6, pp. 507–518, Nov. 2010.
[105] "SPES 2020 - Software Plattform Embedded Systems 2020." [Online]. Available: http://spes2020.informatik.tu-muenchen.de/spes-home.html. [Accessed: 15-Nov-2016].
[106] IEEE Architecture Working Group, "IEEE Std 1471-2000, Recommended Practice for Architectural Description of Software-intensive Systems," IEEE, 2000.

[107] H. Hungar and E. Reyzl, “Software-Entwicklung und Zertifizierung im Umfeld sicherheitskritischer und hochverfügbarer Systeme: Bedeutung modellbasierter und formaler Ansätze für effiziente Entwicklung und Zertifizierung,” in *Software Engineering (Workshops)*, 2008, vol. 122, pp. 299–302.
[108] N. Damianou, N. Dulay, E. Lupu, and M. Sloman, “Ponder: A Language for Specifying Security and Management Policies for Distributed Systems The Language Specification Version 2.3,” Imperial College of Science, Technology and Medicine, Department of Computing, 180 Queen’s Gate, London SW7 2BZ, U.K., Imperial College Research Report DoC 2000/1, Oct. 2000.
[109] K. Twidle, N. Dulay, E. Lupu, and M. Sloman, “Ponder2: A Policy System for Autonomous Pervasive Environments,” in *Fifth International Conference on Autonomic and Autonomous Systems, 2009. ICAS ’09*, 2009, pp. 330–335.
[110] K. J. Turner, S. Reiff-Marganiec, L. Blair, G. A. Campbell, and F. Wang, “APPEL: An Adaptable and Programmable Policy Environment and Language,” Computing Science and Mathematics, University of Stirling, Technical report CSM-161, Apr. 2009.
[111] Foundation for Intelligent Physical Agents, *FIPA Device Ontology Specification*. 2002.
[112] G. Klyne *et al.*, “Composite Capability/Preference Profiles (CC/PP): Structure and Vocabularies 1.0,” W3C, Jan. 2004.
[113] “openEHR.” [Online]. Available: http://www.openehr.org/.
[114] “Health Level Seven (HL7).” [Online]. Available: http://www.hl7.org/.
[115] “IHE - Integrating the Healthcare Enterprise.” [Online]. Available: http://ihe.net/.
[116] “ISO - Technical committees - ISO/TC 215 - Health informatics,” *ISO*. [Online]. Available: http://www.iso.org/iso/iso_technical_committee?commid=54960. [Accessed: 13-Jun-2016].
[117] “CEN/TC 251 - Health informatics.” [Online]. Available: https://standards.cen.eu/dyn/www/f?p=204:7:0::::FSP_ORG_ID:6232&cs=18CA078392807EDD402B798AAEF1644E1.
[118] “American Society for Testing and Materials (ASTM) - Committee E31 on Healthcare Informatics.” [Online]. Available: https://www.astm.org/COMMITTEE/E31.htm.
[119] “The CEN/ISO EN13606 standard.” [Online]. Available: http://www.en13606.org/the-ceniso-en13606-standard.
[120] M. Eichelberg, T. Aden, J. Riesmeier, A. Dogac, and G. B. Laleci, “A Survey and Analysis of Electronic Healthcare Record Standards,” *ACM Comput Surv*, vol. 37, no. 4, pp. 277–315, Dec. 2005.
[121] SALUS Consortium, “SALUS - D1.1.8 Final Report.” 26-Apr-2016.
[122] “HL7/ASTM Implementation Guide for CDA R2 -Continuity of Care Document (CCD) Release 1.” Health Level Seven International.
[123] W. Ceusters and B. Smith, “Semantic Interoperability in Healthcare State of the Art in the US.” New York State Center of Excellence in Bioinformatics and Life Sciences, 03-Mar-2010.
[124] “Home - SemanticHealthNet.” [Online]. Available: http://www.semantichealthnet.eu/. [Accessed: 13-Jun-2016].
[125] “C 154 - HITSP Data Dictionary.” Healthcare Information Technology Standards Panel (HITSP), 25-Jan-2010.
[126] “C 32 - HITSP Summary Documents Using HL7 Continuity of Care Document (CCD) Component.” Healthcare Information Technology Standards Panel (HITSP), 08-Jul-2009.
[127] “Federal Health Information Model Home Page FHIMS.org.” [Online]. Available: http://www.fhims.org/. [Accessed: 14-Jun-2016].
[128] “Standards & Interoperability (S&I) Framework - Transitions of Care (ToC) Initiative.” [Online]. Available: http://wiki.siframework.org/Transitions+of+Care+(ToC)+Initiative. [Accessed: 14-Jun-2016].
[129] “Standards & Interoperability (S&I) Framework - S&I Framework CEDD Overview.” [Online]. Available: http://wiki.siframework.org/S%26I+Framework+CEDD+Overview. [Accessed: 14-Jun-2016].
[130] “Standards & Interoperability (S&I) Framework - Query Health.” [Online]. Available: http://wiki.siframework.org/Query+Health. [Accessed: 14-Jun-2016].
[131] “CDISC | Strength Through Collaboration.” [Online]. Available: http://www.cdisc.org/. [Accessed: 14-Jun-2016].
[132] “Study Data Tabulation Model (SDTM) | CDISC.” [Online]. Available: http://www.cdisc.org/sdtm. [Accessed: 14-Jun-2016].

[133] "Clinical Data Acquisition Standards Harmonization (CDASH) | CDISC." [Online]. Available: http://www.cdisc.org/cdash. [Accessed: 14-Jun-2016].
[134] "BRIDG." [Online]. Available: http://www.bridgmodel.org/. [Accessed: 14-Jun-2016].
[135] "Reference Information Model (RIM) Downloads." [Online]. Available: http://www.hl7.org/implement/standards/rim.cfm. [Accessed: 14-Jun-2016].
[136] "Mini-Sentinel." [Online]. Available: http://www.mini-sentinel.org/. [Accessed: 14-Jun-2016].
[137] "Observational Medical Outcomes Partnership." [Online]. Available: http://omop.org/. [Accessed: 14-Jun-2016].
[138] "i2b2: Informatics for Integrating Biology & the Bedside." [Online]. Available: https://www.i2b2.org/. [Accessed: 14-Jun-2016].
[139] International Health Terminology Standards Development Organisation, "SNOMED CT." .
[140] D. Fensel, "Ontologies," in *Ontologies*, Springer Berlin Heidelberg, 2001, pp. 11–18.
[141] "OWL 2 Web Ontology Language Document Overview (Second Edition)." [Online]. Available: https://www.w3.org/TR/owl2-overview/. [Accessed: 07-Nov-2016].
[142] M. del C. Legaz-García, C. Martínez-Costa, M. Menárguez-Tortosa, and J. T. Fernández-Breis, "A semantic web based framework for the interoperability and exploitation of clinical models and EHR data," *Knowl.-Based Syst.*, vol. 105, pp. 175–189, Aug. 2016.
[143] "Late-Binding Data Warehouse," *Health Catalyst*. .
[144] "DIVING IN:NAVIGATING A DATA LAKE FOR PREDICTIVE CAR." Aug-2015.
[145] "The Future of Personalized Healthcare: Predictive Analytics," *Rock Health*. [Online]. Available: https://rockhealth.com/reports/predictive-analytics/. [Accessed: 14-Jun-2016].
[146] A. A. Sinaci and G. B. Laleci Erturkmen, "A federated semantic metadata registry framework for enabling interoperability across clinical research and care domains," *J. Biomed. Inform.*, vol. 46, no. 5, pp. 784–794, Oct. 2013.
[147] M. Marrs, "The Difference Between Data, Analytics, and Insights," *Business 2 Community*. [Online]. Available: http://www.business2community.com/business-intelligence/difference-data-analytics-insights-01540318. [Accessed: 21-Sep-2016].
[148] C. Hill, "Analysis vs. Analytics: What's the Difference?," *1to1 Media*, 21-Jun-2011. [Online]. Available: http://www.1to1media.com/data-analytics/analysis-vs-analytics-whats-difference. [Accessed: 21-Sep-2016].
[149] A. T. Janke, D. L. Overbeek, K. E. Kocher, and P. D. Levy, "Exploring the Potential of Predictive Analytics and Big Data in Emergency Care," *Ann. Emerg. Med.*, vol. 67, no. 2, pp. 227–236, Feb. 2016.
[150] A. Holzinger and I. Jurisica, "Knowledge Discovery and Data Mining in Biomedical Informatics: The Future Is in Integrative, Interactive Machine Learning Solutions," in *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics*, A. Holzinger and I. Jurisica, Eds. Springer Berlin Heidelberg, 2014, pp. 1–18.
[151] F. Wang, L. S. Docherty, K. J. Turner, M. Kolberg, and E. H. Magill, "Services and Policies for Care at Home," *1st Int. Conf. Pervasive Comput. Technol. Healthc.*, p. 7.1-7.10, 2006.
[152] N. T. Issa, S. W. Byers, and S. Dakshanamurthy, "Big data: the next frontier for innovation in therapeutics and healthcare," *Expert Rev. Clin. Pharmacol.*, vol. 7, no. 3, pp. 293–298, May 2014.
[153] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," *Health Inf. Sci. Syst.*, vol. 2, p. 3, 2014.
[154] K. Yu, X. Wu, W. Ding, and J. Pei, "Towards Scalable and Accurate Online Feature Selection for Big Data," in *2014 IEEE International Conference on Data Mining*, 2014, pp. 660–669.
[155] R. Fang, S. Pouyanfar, Y. Yang, S-C. Chen, and S.S. Iyengar, "Computational Health Informatics in the Big Data Age: A Survey," *ACM Comput. Surv.*, vol. 49, no. 1, Jul. 2016.
[156] G. E. Hinton, "A Practical Guide to Training Restricted Boltzmann Machines (Version 1)," *ResearchGate*, vol. 9, no. 1, Aug. 2010.
[157] D. Sanders, "Healthcare 2.0: The Age of Analytics," Aug-2013.
[158] "Video and Image Analytics - PARC, a Xerox company." [Online]. Available: https://www.parc.com/services/focus-area/video_image_analytics/. [Accessed: 21-Sep-2016].
[159] F. Venter and A. Stein, "Images & videos: really big data," *Analytics Magazine*, 05-Nov-2012. [Online]. Available: http://analytics-magazine.org/images-a-videos-really-big-data/. [Accessed: 21-Sep-2016].

[160] J. Kobielus, “Using advanced image analytics to spot hidden cancer patterns | IBM Big Data & Analytics Hub,” 09-May-2014. [Online]. Available: http://www.ibmbigdatahub.com/blog/using-advanced-image-analytics-spot-hidden-cancer-patterns. [Accessed: 21-Sep-2016].
[161] J. S. Lewis, S. Ali, J. Luo, W. L. Thorstad, and A. Madabhushi, “A quantitative histomorphometric classifier (QuHbIC) identifies aggressive versus indolent p16-positive oropharyngeal squamous cell carcinoma,” *Am. J. Surg. Pathol.*, vol. 38, no. 1, pp. 128–137, Jan. 2014.
[162] IBM, “Medical Image Analytics - IBM,” 22-Mar-2013. [Online]. Available: http://researcher.watson.ibm.com/researcher/view_group.php?id=4829. [Accessed: 21-Sep-2016].
[163] IBM, “Medical Sieve - IBM,” 22-Mar-2013. [Online]. Available: http://researcher.watson.ibm.com/researcher/view_group.php?id=4384. [Accessed: 21-Sep-2016].
[164] IBM, “IBM Watson.” [Online]. Available: http://www.ibm.com/watson/. [Accessed: 21-Sep-2016].
[165] IBM Research, *IBM Research Accelerating Discovery: Medical Image Analytics*. 2013.
[166] “VTK - The Visualization Toolkit.” .
[167] “ITK - Segmentation & Registration Toolkit.” [Online]. Available: https://itk.org/. [Accessed: 10-Nov-2016].
[168] “MATLAB - MathWorks.” [Online]. Available: https://www.mathworks.com/products/matlab/?requestedDomain=nl.mathworks.com. [Accessed: 10-Nov-2016].
[169] “MeVisLab: MeVisLab.” [Online]. Available: http://www.mevislab.de/. [Accessed: 10-Nov-2016].
[170] “syngo.via Frontier.” [Online]. Available: https://www.healthcare.siemens.com/medical-imaging-it/clinical-imaging-applications/syngo-via-frontier. [Accessed: 10-Nov-2016].
[171] “MEDUSA Exploitable Results.pdf.” .
[172] “Data visualization,” *Wikipedia*. 19-Nov-2016.
[173] “Data Visualization: What it is and why it matters.” [Online]. Available: http://www.sas.com/en_sg/insights/big-data/data-visualization.html. [Accessed: 12-Aug-2016].
[174] L. Goodwin, M. VanDyne, S. Lin, and S. Talbert, “Data mining issues and opportunities for building nursing knowledge,” *J. Biomed. Inform.*, vol. 36, no. 4–5, pp. 379–388, Oct. 2003.
[175] “Clinical Decision Support Systems: State of the Art.” Agency for Healthcare Research and Quality U.S. Department of Health and Human Services, 2009.
[176] “Clinical Decision Support Systems.” [Online]. Available: http://www.openclinical.org/dss.html. [Accessed: 10-Nov-2016].
[177] M. J. Hardin and D. C. Chhieng, “Data Mining and Clinical Decision Support Systems.” .
[178] S. L. Harris, J. H. May, and L. G. Vargas, “Predictive analytics model for healthcare planning and scheduling,” *Eur. J. Oper. Res.*, vol. 253, no. 1, pp. 121–131, Aug. 2016.
[179] N. M. S. kumar, T. Eswari, P. Sampath, and S. Lavanya, “Predictive Methodology for Diabetic Data Analysis in Big Data,” *Procedia Comput. Sci.*, vol. 50, pp. 203–208, Jan. 2015.
[180] “Clinical Decision Support System (CDSS) Market Segment Forecasts up to 2023 Research Reports- TransparencyMarketResearch.” [Online]. Available: http://www.transparencymarketresearch.com/clinical-decision-support-system-market.html. [Accessed: 26-Aug-2016].
[181] “iWander app for iPhone: reviews, screenshots, forum, users community.” [Online]. Available: http://iwander.iapps4you.com/. [Accessed: 07-Sep-2016].
[182] “Medusa Medical Technologies | Home |Medusa Medical | Siren ePCR Suite | Patient Records Software.” [Online]. Available: http://www.medusamedical.com/. [Accessed: 07-Sep-2016].
[183] “Swing (Java),” *Wikipedia*. 15-Oct-2016.
[184] “Standard Widget Toolkit,” *Wikipedia*. 19-Jul-2016.
[185] Krish Ramachandran, IT Specialist, IBM, “Adaptive user interfaces for health care applications.” 20-Jan-2009.
[186] “Introduction to Model-Based User Interfaces.” [Online]. Available: https://www.w3.org/TR/mbui-intro/. [Accessed: 07-Sep-2016].
[187] “EclipseCon 2008: Fundamentals of the Eclipse Modeling Framework,” 17:56:58 UTC.

[188] "NIST Cloud Computing Reference Architecture (500-292) PDF." [Online]. Available: http://www.nist.gov/customcf/get_pdf.cfm?pub_id=909505. [Accessed: 13-Jun-2016].
[189] "Survey on Security Issues in Cloud Computing - International Journal of Computer Applications (0975–888) PDF." [Online]. Available: http://research.ijcaonline.org/volume47/number18/pxc3880578.pdf. [Accessed: 13-Jun-2016].
[190] J. Clarke-Salt, *SQL Injection Attacks and Defence*. Syngress, 2009.
[191] "Security Issues in Cloud Deployment Models - TechNet Articles - United States (English) - TechNet Wiki." [Online]. Available: http://social.technet.microsoft.com/wiki/contents/articles/4509.security-issues-in-cloud-deployment-models.aspx. [Accessed: 14-Jun-2016].
[192] "Mitigating economic denial of sustainability (edos) in cloud computing using in-cloud scrubber service," presented at the Computational Intelligence and Communication Networks (CICN), 2012 Fourth International Conference, pp. 535–539.
[193] "Survey of Network Based Defense Mechanisms Count ering the DoS and DDoS Problems - ACM Comput. Surv. 39 , 1, Article 3 (April 2007)." [Online]. Available: https://www.cs.ucf.edu/~dcm/Teaching/COT4810-Spring2011/Literature/DenialOfServiceAttacks.pdf. [Accessed: 14-Jun-2016].
[194] F. Sabahi, "Secure Virtualization for Cloud Environment Using Hypervisor-based Technology," *Int J. Mach. Learn. Comput.*, vol. 2, no. 1, pp. 39–45, Feb. 2012.
[195] G. Zhao, C. Rong, M. G. Jaatun, and F. E. Sandnes, "Deployment Models: Towards Eliminating Security Concerns from Cloud Computing," in *International Conference on High Performance Computing and Simulation (HPCS)*, 2010, pp. 189–195.
[196] G. Zhao, C. Rong, M. G. Jaatun, and F. E. Sandnes, "Reference Deployment Models for Eliminating User Concerns on Cloud Security," *J. Supercomput.*, vol. 61, no. 2, pp. 337–352, 2012.
[197] R. Asija, "Enhancing Security and Privacy of Healthcare Data using XML Schema," *Int. J. Comput. Appl. 0975 8887*, vol. 116, p. 6, Apr. 2015.
[198] "Security Assertion Markup Language (SAML) v2.0." OASIS Security Services TC, Mar-2005.
[199] "OASIS SAML v2.0 Technical Overview." OASIS Security Services TC, 25-Mar-2008.
[200] "OASIS XACML 2.0 Specification Set." OASIS eXtensible Access Control Markup Language (XACML) TC, Feb-2005.
[201] "eXtensible Access Control Markup Language (XACML) Version 2.0, Core Specification." OASIS eXtensible Access Control Markup Language (XACML) TC, 01-Feb-2005.
[202] "SAML 2.0 profile of XACML v2.0." OASIS eXtensible Access Control Markup Language (XACML) TC, 19-Aug-2014.
[203] "OASIS Cross-Enterprise Security and Privacy Authorization (XSPA) Profile of Security Assertion Markup Language (SAML) for Healthcare Version 1.0, Committee Specification." OASIS Cross-Enterprise Security and Privacy Authorization (XSPA) TC, 25-Aug-2009.
[204] "OASIS Cross-Enterprise Security and Privacy Authorization (XSPA) Profile of XACML v2.0 for Healthcare Version 1.0, Committee Specification." OASIS eXtensible Access Control Markup Language (XACML) TC, 27-Aug-2009.
[205] "IHE IT Infrastructure (ITI) Technical Framework, Volume 1 (ITI TF-1) Integration Profiles." 22-Apr-2016.
[206] "RFC 6749 - The OAuth 2.0 Authorization Framework." .
[207] J. H. Lee, "Authenticated encryption in the symmetric and asymmetric settings," Ph.D., University of California, San Diego, United States -- California, 2001.
[208] C. Bradford, "5 Common Encryption Algorithms and the Unbreakables of the Future - StorageCraft," *StorageCraft Technology Corporation*, 31-Jul-2014. [Online]. Available: http://www.storagecraft.com/blog/5-common-encryption-algorithms/. [Accessed: 18-Aug-2016].
[209] M. Behrens, "Understanding Encryption – Symmetric, Asymmetric, & Hashing," *Atomic Spin*, 20-Nov-2014. .
[210] Webopedia, "What is Hashing? Webopedia Definition." [Online]. Available: http://www.webopedia.com/TERM/H/hashing.html. [Accessed: 30-Sep-2016].
[211] B. van der Sloot, D. Broeders, and E. Schrijvers, *Exploring the Boundaries of Big Data Bart van der Sloot, Dennis Broeders & Erik Schrijvers*. Amsterdam University Press, 2016.

[212] G. Lafuente, “Big Data Security - Challenges & Solutions,” *MWR InfoSecurity*. [Online]. Available: https://www.mwrinfosecurity.com/our-thinking/big-data-security-challenges-and-solutions/. [Accessed: 19-Aug-2016].
[213] Big Data Working Group, “Top Ten Big Data Security and Privacy Challenges,” Apr. 2013.
[214] Big Data Working Group, “Top Ten Big Data Security and Privacy Challenges,” Nov. 2012.
[215] L. Sweeney, “k-ANONYMITY: A MODEL FOR PROTECTING PRIVACY,” 2002.
[216] K. Babu, N. Ranabothu, and N. Kumar, “Achieving k-anonymity Using Improved Greedy Heuristics for Very Large Relational Databases,” *Trans. Data Priv.*, no. 6(1), pp. 1–17, 2013.
[217] P. Samarati and L. Sweeney, “Generalizing data to provide anonymity when disclosing information (abstract),” 1998.
[218] K. El Emam, F. Dankar, R. Issa, E. Jonker, and D. Amyot, “A Globally Optimal k-Anonymity Method for the De-Identification of Health Data,” *J. Am. Med. Inform. Assoc. JAMIA*, vol. 16(5), pp. 670–682, 2009.
[219] K. El Emam, *Guide to the De-Identification of Personal Health Information*. Publisher Taylor & Francis Group, 2013.
[220] V. Ciriani, S. De Capitani di Vimercati, S. Foresti, and P. Samarati, “k-Anonymity,” *Springer US Adv. Inf. Secur.*, pp. 1–36, 2007.
[221] L. Sweeney, “Systems and methods for deidentifying entries in a data source,” 7,269,578 B2, 09-Jan-2007.
[222] P. Samarati, “Protecting Respondents’ Identities in Microdata Release,” *IEEE Trans. Knowl. Data Eng.*, vol. 13, no. 6, pp. 1010–1027, 2001.
[223] R. Issa, “Satisfying k-anonymity: New algorithm and empirical evaluation,” M.C.S., Carleton University (Canada), Canada, 2009.
[224] R. Agrawal and R. J. Bayardo, “Data Privacy Through Optimal k-Anonymization,” 2005.
[225] J. A. Sherer, J. Le, and A. Taal, “Big Data Discovery, Privacy, and the Application of Differential Privacy Mechanisms,” *Comput. Internet Lawyer*, vol. 32, no. 7, pp. 10–16, Jul. 2015.
[226] Q. Geng, “The optimal mechanism in differential privacy,” Ph.D., University of Illinois at Urbana-Champaign, United States -- Illinois, 2013.
[227] D. C. Roth A., *The Algorithmic Foundations of Differential Privacy*, Roth, A. vols. Now Publishers, 2014.
[228] A. Roth, “New algorithms for preserving differential privacy,” Ph.D., Carnegie Mellon University, United States -- Pennsylvania, 2010.
[229] A. Tockar, “Laplace Mechanism,” 08-Sep-2014. .
[230] Q. Geng, “The optimal mechanism in differential privacy,” Ph.D., University of Illinois at Urbana-Champaign, United States -- Illinois, 2013.
[231] A. Roth, “Privacy for Non-Numeric Queries.”
[232] K. Mivule, “An investigation of data privacy and utility using machine learning as a gauge,” D.Sc., Bowie State University, United States -- Maryland, 2014.
[233] G. Cormode, “Distributed Streams.”
[234] A. Tockar, “Differential Privacy: The Basics – Neustar Research Blog,” 2014. [Online]. Available: http://content.research.neustar.biz/blog/differential-privacy/WhiteQuery.html. [Accessed: 29-Jul-2016].
[235] W. Yuxiang, “Differential Privacy: An Introduction,” 19-Sep-2012.
[236] C. Dwork, “A Firm Foundation for Private Data Analysis,” *Commun. ACM*, Jan. 2011.
[237] M. Hardt, K. Ligett, and F. McSherry, “A simple and practical algorithm for differentially private data release,” *ArXiv10124763 Cs*, p. 2, Dec. 2010.
[238] C. Dwork and A. Roth, “The Algorithmic Foundations of Differential Privacy,” *Found Trends Theor Comput Sci*, vol. 9, no. 3–4, pp. 211–407, Aug. 2014.
[239] A. Machanavajjhala, “Privacy in a Mobile-Social World.”
[240] F. K. Dankar and K. El Emam, “Practicing Differential Privacy in Health Care: A Review,” *Trans Data Priv.*, vol. 6, no. 1, pp. 35–67, Apr. 2013.
[241] L. Wasserman and S. Zhou, “A Statistical Framework for Differential Privacy,” *J. Am. Stat. Assoc.*, vol. 105, no. 489, pp. 375–389, 2010.
[242] S. L. Garfinkel, “De-Identification of Personal Information. NISTIR 8053.” NIST, 2015.
[243] M. Mitrea and F. Prêteux, “Multimedia Content Watermarking,” in *The Wireless and Mobile Network Security H. Chaouchi and M. Laurent-Maknavicius Editors*, Willey, 1999, pp. 149–201.

[244] A. Garboan, M. Mitrea, and F. Prêteux, “Cinematography sequences tracking by means of fingeprinting technique,” *Ann. Telecommun.*, vol. 68, no. 3–4, pp. 187–199, Apr. 2013.
[245] “BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES.” Executive Office of the President, 2014.
[246] C. J. Bennett and C. Raab, *The Governance of Privacy: Policy Instruments in Global Perspective*, 1st ed. Cambridge: MIT Press, 2006.
[247] “Safeharbor.” [Online]. Available: export.gov/safeharbor. [Accessed: 17-Jul-2016].
[248] “The EU-U.S. Privacy Shield.” .
[249] “Privacy Shield Framework,” 08-Aug-2016. [Online]. Available: https://www.privacyshield.gov/Program-Overview.
[250] C. J. Bennett and R. M. Bayley, “Privacy Protection in the era of ‘Big Data’: Regulatory challenges and social assessment,” in *Exploring the Boundaries of Big Data*, Amsterdam University Press, 2016, pp. 205–227.
[251] J. Hoboken van, “From collection to use in the privacy regulation? A forward-looking comparison of European and US framework for personal data processing,” in *Exploring the Boundaries of Big Data*, Amsterdam University Press, 2016, pp. 231–258.
[252] B. Sloot van der, “The individual in the Big Data era: moving towards an agent-based privacy paradigm,” in *Exploring the Boundaries of Big Data*, Amsterdam University Press, 2016, pp. 177–203.
[253] O. Tene and J. Polonetsky, “Big Data for All: Privacy and User Control in the Age of Analytics,” *Northwest. J. Technol. Intellect. Prop.*, vol. 11, no. 5, pp. 239–273, 2013.
[254] “Big Data Ethics Initiative.” [Online]. Available: http://informationaccountability.org/big-data-ethics-initiative/.
[255] E. Douilhet and A. P. Karanasiou, “Legal Responses to the Commodification of Personal Data in the Era of Big Data: The Paradigm Shift from Data Protection towards Data Ownership,” in *Effective Big Data Management and Opportunities for Implementation*, IGI Global, 2016, pp. 130–139.
[256] Big Data Working Group, “Expanded Top Ten Big Data Security and Privacy Challenges.” Cloud Security Alliance, 2013.
[257] T. Van Overstraeten, “Legal Guidelines on the Use of Electronic Patient Data,” presented at the eHeatlh 2010 Conference, Barcelona, Spain, 2010.
[258] “ARTICLE 29 DATA PROTECTION WORKING PARTY. Opinion 4/2007 on the concept of personal data.” European advisory body on data protection and privacy, 20 June.
[259] B. Hawkes, “Data Protection Guidelines on research in the Health Sector.” Data Protection Commissioner, Ireland, Nov-2007.
[260] “Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).” European Parliament, Council of the European Union, Apr-2016.
[261] B. W.S., “GDPR: Getting Ready for the New EU General Data Protection Regulation.” nformation Law Group. InfoLawGroup LLP, 05-May-2016.
[262] “European Commission - Fact Sheet. Questions and Answers - Data protection reform"." European Commission, 21-Dec-2015.
[263] “Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.” European Parliament, Council of the European Union, 24-Oct-1995.
[264] “Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications).” European Parliament, Council of the European Union, 2002.
[265] “Opinion 02 /2013 on apps on smart devices.” THE WORKING PARTY ON THE PROTECTION OF INDIVIDUALS WITH REGARD TO THE PROCESSING OF PERSONAL DATA, 27-Feb-2013.
[266] “Work stream 1.2: Enable me to make the right health and care choices: providing citizens with access to an accredited set of NHS and social care ‘apps.’” National Information Board, UK.
[267] “PAS 277:2015 on Health and wellness apps – Quality criteria across the life cycle – Code of practice, as a set of quality criteria for the development, testing and releasing of health and wellness apps.” BSI, UK, 2015.

[268] "SUMMARY OF THE MEETING STAKEHOLDER MEETING ON QUALITY AND RELIABILITY OF MOBILE HEALTH APPLICATIONS 06TH JULY 2015, BRUSSELS." European Commission, Directorate-General for Communications Networks, Content and Technology, Jul-2015.
[269] "mHealth Master Plan." El Departament de Salut, Generalitat de Catalunya, Jan-2015.
[270] "COMMISSION STAFF WORKING DOCUMENT on the existing EU legal framework applicable to lifestyle and wellbeing apps Accompanying the document GREEN PAPER on mobile Health ('mHealth')." European Parliament, Council of the European Union, 2014.
[271] "Draft Code of Conduct on privacy for mobile health applications." DG Connect, European Commission, 2015.
[272] "The Health Insurance Portability and Accountability Act of 1996 (HIPAA) P.L. No. 104-191, 110 Stat. 1938 (1996)." .
[273] "An Introductory Resource Guide for Implementing the Health Insurance Portability and Accountability Act (HIPPA) Security Rule;" NIST Special Publication 800-66 Revision1, 06-Oct-2016.
[274] "Standards for Security Categorization of Federal Information and Information Systems. Federal Information Processing Standards Publication." 2004.
[275] "Minimum Security Requirements for Federal Information and Information Systems. Federal Information Processing Publication." FIPS Publication 200, 2006.
[276] "Minimum Necessary Requirement: 45 CFR 164.502(b), 164.514(d)." U.S. Department of Health & Human Services, Health Information Privacy, 04-Apr-2003.
[277] "Guidance regarding methods for De-identification of protected health information in accordance with the Health Insurance Portabilityand Accountability Act (HIPAA) Privacy Rule. Washington, DC.: HHS." 2012.
[278] Institute of Medicine (IOM), "Sharing clinical trial data: Maximizing benefits, minimizing risk." The National Academies Press, Washington, D.C., 2015.
[279] "Standards for Privacy of Individually Identifiable Health Information; Proposed Rule. 45 CFR Parts 160 and 164 Office of the Secretary." 67 Fed.Reg.14776,14799, 27-Mar-2001.
[280] El Emam K., *Guide to the deidentification of personal health information.* Boca Raton, FL: CRC Press (Auerbach Publications), 2013.
[281] "Statistical Policy Working Paper 22 (Second Vertion, 2005). Report on Statistical Disclosure Limitation Methodology." Federal Committee on Statistical Methodology, 2005.
[282] C. Dwork, "Differential Privacy," *33rd Int. Colloq. Autom. Lang. Program. Part II ICALP 2006*, vol. 4052, pp. 1–12, 2006.
[283] R. Gellman, "The Deidentification Dilemma: A Legislative and Contractual Proposal," *Fordham Intellect. Prop. Media Entertain. Law J.*, vol. 21, no. 1, pp. 33–55, 2011.
[284] C. M. O'Keefe and J. O. Chipperfield, "A Summary of Attack Methods and Confidentiality Protection Measures for Fully Automated Remote Analysis Systems," *Int. Stat. Rev.*, vol. 81, no. 3, pp. 426–455, 2013.
[285] BIG Consortium, "D.4.2.2. Final version of IPR, Standardization recommendations." 2014.
[286] K. Terry, "Patient records: The struggle for ownership," *Medical Economics*, 10-Dec-2015. [Online]. Available: http://medicaleconomics.modernmedicine.com/medical-economics/news/patient-records-struggle-ownership. [Accessed: 12-Aug-2016].
[287] "Medical Records (Ownership and Storage) (Hansard, 30 November 1976)." [Online]. Available: http://hansard.millbanksystems.com/written_answers/1976/nov/30/medical-records-ownership-and-storage. [Accessed: 12-Aug-2016].
[288] "Policy and Procedure For Records: Retention & Disposal." Mersey Care NHS Trust, Dec-2003.
[289] "§ 630f BGB Dokumentation der Behandlung," *dejure.org*. [Online]. Available: http://dejure.org/gesetze/BGB/630f.html. [Accessed: 12-Aug-2016].
[290] "National Patient Rights Legislation: The Netherlands." [Online]. Available: http://europatientrights.eu/countries/signed/netherlands/netherlands.html. [Accessed: 19-Aug-2016].
[291] A. Novotny and S. Spiekermann, "Personal Information Markets AND Privacy: A New Model to Solve the Controversy.," *11th Int. Conf. Wirtsch. Leipz. Ger.*, 2013.
[292] M. Mun, S. Hao, and N. Mishra, "Personal Data Vaults: A Locus of Control for Personal Data Streams." ACM CoNext, 2010.

[293] Y.-A. Monjoye de, E. Shmueli, and S. Wang, “Open PDS: Protecting the Privacy of Metadata through SafeAnswers.” 2014.
[294] V. Markl, H. Krcmar, and T. Hoeren, “Big Data Management: Innovation potential analysis for the new technologies for managing and analyzing large amounts of data.” Mar-2014.
[295] J. Harer, *Anforderungen an Medizinprodukte: Praxisleitfaden für Hersteller und Zulieferer*, 2., Überarbeitete Auflage. München: Hanser, 2014.
[296] Council of the European Union, “Council Directive 93/42/EEC of 14 June 1993 concerning medical devices.” European Parliament, Council of the European Union, 1993.
[297] CE Marking Consulting Service, “CE Marking Logo,” *CE-Marking*. [Online]. Available: http://www.ce-marking.com/CE-marking-logo.html. [Accessed: 08-Jun-2016].
[298] European Standards Organisation: CEN, CENELEC, or ETSI, “Summary list of titles and references harmonised standards under Directive 93/42/EEC for Medical devices,” *Medical devices*. [Online]. Available: http://ec.europa.eu/growth/single-market/european-standards/harmonised-standards/medical-devices/index_en.htm. [Accessed: 06-Jun-2016].
[299] J. Lincoln, “The Medical Device Design History File, Technical File / Design Dossier,” *MasterControl*. [Online]. Available: http://de.mastercontrol.com/newsletter/medical_device/medical-device-design-history-file-0710.html. [Accessed: 02-Jun-2016].
[300] EMERGO GROUP, “European Authorized Representative for Medical Device and IVD Companies,” *EMERGO*. [Online]. Available: http://www.emergogroup.com/services/europe/european-authorized-representative. [Accessed: 02-Jun-2016].
[301] EMERGO GROUP, “Europe CE Marking Regulatory Process for Medical Devices,” *EMERGO*. [Online]. Available: http://www.emergogroup.com/resources/europe-process-chart. [Accessed: 02-Jun-2016].
[302] “Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on in vitro diagnostic medical devices.” European Parliament, Council of the European Union, 1998.
[303] Council of the European Union, “Council Directive 90/385/EEC of 20 June 1990 on the approximation of the laws of the Member States relating to active implantable medical devices.” European Parliament, Council of the European Union, 1990.
[304] International Organization for Standardization, “ISO 13485:2016 Medical devices -- Quality management systems -- Requirements for regulatory purposes.” 2016.
[305] “EN 60601-1,” *LinkFang*. [Online]. Available: http://www.linkfang.de/wiki/EN_60601-1. [Accessed: 08-Jun-2016].
[306] International Organization for Standardization, “IEC 62304:2006(en) Medical device software — Software life cycle processes.” .
[307] EUROPEAN COMMISSION ENTERPRISE AND INDUSTRY DIRECTORATE GENERAL, “CLINICAL EVALUATION: A GUIDE FOR MANUFACTURERS AND NOTIFIED BODIES.” EUROPEAN COMMISSION ENTERPRISE AND INDUSTRY DIRECTORATE GENERAL, 2009.
[308] mdc medical device certification GmbH, “Basic Information about the European Directive 93/42/EEC on Medical Devices.” mdc medical device certification GmbH, 2009.
[309] E. French-Mowat and J. Burnett, “How are medical devices regulated in the European Union?,” *SAGE Publ.*, vol. 105, pp. p22-28, 2012.
[310] DKE, “GERMAN STANDARDIZATION ROADMAP Mobile Diagnostic Systems.” VDE ASSOCIATION FOR ELECTRICAL, ELECTRONIC & INFORMATION TECHNOLOGIES, May-2015.
[311] S. Ramakrishna, L. Tian, C. Wang, S. Liao, and T. Wee Eong, *Medical Devices: Regulations, Standards and Practices*. Woodhead Publishing, 2015.
[312] Emergo Group Inc., “An Overview of the US Regulatory Process for Medical Devices,” *SlideShare*. [Online]. Available: http://de.slideshare.net/emergogroup/us-fda-medical-device-regulatory-approval-process. [Accessed: 09-Jun-2016].
[313] FDA U.S. Food and Drug Administration, “How to Study and Market Your Device,” *FDA U.S. Food and Drug Administration*. [Online]. Available: http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/#step3. [Accessed: 10-Jun-2016].
[314] M. Adib Bamiah, S. Nawaz Brohi, and S. Chuprat, “TRUSTED CLOUD COMPUTING FRAMEWORK FOR HEALTHCARE SECTOR,” *J. Comput. Sci.*, vol. 10, no. 2, pp. 240–250.

[315] S. Hallett, G. Parr, and S. McClean, “Cloud-based Healthcare: Towards a SLA Compliant Network Aware Solution for Medical Image Processing,” *CLOUD Comput. 2012 Third Int. Conf. Cloud Comput. GRIDs Virtualization*, pp. 219–223, 2012.
[316] “Service level agreement| ZH Healthcare,” *EHR & Practice Management Solution | ZH Healthcare*. [Online]. Available: http://zhhealthcare.com/legal-documents/service-level-agreement/. [Accessed: 02-Sep-2016].
[317] WCHN, “Women’s and Children’s Health Network, Service Level Agreement,” Jul. 2016.
[318] CALHN, “Central Adelaide Local Health Network, Service Level Agreement 1 July 2015 – 30 June 2016,” Agreement, Jul. 2015.
[319] M. Gibbons, R. Wilson, and L. Samal, “Impact of Consumer Health Informatics Applications,” Evidence Report Technology Assessment (Full Report), 2009.
[320] A. Logan *et al.*, “Mobile Phone–Based Remote Patient Monitoring System for Management of Hypertension in Diabetic Patients,” *Am. J. Hypertens.*, vol. 20, no. 9, pp. 942–948, Sep. 2007.
[321] S. Aggarwal and L. McCabe, “The Compelling TCO Case for Cloud Computing in SMB and Mid-Market Enterprises 1st ed,” 2009.
[322] H. D. Herr Dr. Diesing, DIN e.V, and Beuth Verlag, *Zulassung von Medizinprodukten Ein Leitfaden*. 2016.
[323] MEDUSA, “MEDUSA poster presented at the ITEA co-summit, March 2015, Berlin – Germany.” 2015.
[324] MEDUSA Consortium, “MEDUSA final brochure.” .
[325] “SALUS Project,” 26-Apr-2016. [Online]. Available: http://www.salusproject.eu/. [Accessed: 26-Apr-2016].
[326] G. B. Laleci, M. Yuksel, and A. Dogac, “Providing Semantic Interoperability Between Clinical Care and Clinical Research Domains,” *IEEE J. Biomed. Health Inform.*, vol. 17, no. 2, pp. 356–369, Mar. 2013.
[327] M. Yuksel *et al.*, “An Interoperability Platform Enabling Reuse of Electronic Health Records for Signal Verification Studies,” *BioMed Res. Int.*, vol. 2016, pp. 1–18, 2016.
[328] A. A. Sinaci *et al.*, “Postmarketing Safety Study Tool: A Web Based, Dynamic, and Interoperable System for Postmarketing Drug Surveillance Studies,” *BioMed Res. Int.*, vol. 2015, pp. 1–10, 2015.
[329] “Common Data Model | Observational Medical Outcomes Partnership,” 26-Apr-2016. [Online]. Available: http://omop.org/CDM. [Accessed: 26-Apr-2016].
[330] T. Krahn, M. Eichelberg, S. Gudenkauf, G. B. L. Erturkmen, and H.-J. Appelrath, “Adverse Drug Event Notification System: Reusing Clinical Patient Data for Semi-automatic ADE Detection,” 2014, pp. 251–256.
[331] “Query for Existing Data Profile - IHE Wiki,” 26-Apr-2016. [Online]. Available: http://wiki.ihe.net/index.php/Query_for_Existing_Data_Profile. [Accessed: 26-Apr-2016].
[332] “Care Management Profile - IHE Wiki,” *Care Management Profile - IHE Wiki*, 26-Apr-2016. [Online]. Available: http://wiki.ihe.net/index.php/Care_Management_Profile. [Accessed: 26-Apr-2016].
[333] “HL7 Standards Product Brief - HL7 Version 3 Standard: Representation of the Health Quality Measure Format (eMeasure) DSTU, Release 2,” 26-Apr-2016. [Online]. Available: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=97. [Accessed: 26-Apr-2016].
[334] IHE ITI Commitee, “Audit Trail and Node Authentication - IHE Wiki,” *Audit Trail and Node Authentication - IHE Wiki*, 26-Apr-2016. [Online]. Available: http://wiki.ihe.net/index.php/Audit_Trail_and_Node_Authentication. [Accessed: 26-Apr-2016].
[335] “OPEN ATNA Project,” 26-Apr-2016. [Online]. Available: https://www.projects.openhealthtools.org/sf/projects/openatna/. [Accessed: 26-Apr-2016].
[336] “iCARDEA Project,” 26-Apr-2016. [Online]. Available: http://www.srdc.com.tr/projects/icardea/. [Accessed: 26-Apr-2016].
[337] G. Laleci *et al.*, “Personalized Remote Monitoring of the Atrial Fibrillation Patients with Electronic Implant Devices,” *J. Healthc. Eng.*, vol. 2, no. 2, pp. 183–196, Jun. 2011.
[338] “PCD Implantable Device Cardiac Observation - IHE Wiki,” 26-Apr-2016. [Online]. Available: http://wiki.ihe.net/index.php/PCD_Implantable_Device_Cardiac_Observation. [Accessed: 26-Apr-2016].
[339] “HL7 Standards Product Brief - HL7 Version 3 Standard: Common Terminology Services (CTS), Release 1,” 26-Apr-2016. [Online]. Available:

http://www.hl7.org/implement/standards/product_brief.cfm?product_id=10. [Accessed: 26-Apr-2016].
[340] "10014-EASI-CLOUDS-WP-5-D51A-Projectflyer." .
[341] "EASI-CLOUDS Project Leaflet." .
[342] "EASI-CLOUDS Posters Co-Summit 2015." .
[343] "OSAmI D4.3." OSAmI Consortium, Nov-2011.
[344] "CloudPort Project." [Online]. Available: http://www.prologue.fr/en/pages/cloudport/. [Accessed: 25-May-2016].
[345] "CloudPort_Phase3_3-2-Rapport_Technique_Général" - CloudPort Consortium." May-2014.
[346] "CompatibleOne - Open source Cloud broker research project." [Online]. Available: https://projects.ow2.org/bin/view/compatibleone/. [Accessed: 25-May-2016].
[347] "OCCI - Open Cloud Computing Interface." .
[348] "OpenIoT - Open Source cloud solution for the Internet of Things." .
[349] "Global Sensor Networks - a middleware for processing sensor data in the Internet." [Online]. Available: http://lsir.epfl.ch/research/current/gsn/. [Accessed: 24-May-2016].
[350] "Linked Sensor Middleware - brings together the live real world sensed data and the Semantic Web." [Online]. Available: https://code.google.com/archive/p/deri-lsm/. [Accessed: 24-May-2016].
[351] "OpenIotOrg/openiot," *GitHub*. [Online]. Available: https://github.com/OpenIotOrg/openiot. [Accessed: 15-Nov-2016].
[352] "eXtended Global Sensor Network." [Online]. Available: http://ceur-ws.org/Vol-1401/paper-04.pdf. [Accessed: 24-May-2016].
[353] "Cloud4Health Web Page," *Cloud-Computing für Big-Data-Analysen in der Medizin*. [Online]. Available: http://www.cloud4health.de/. [Accessed: 23-Jun-2016].
[354] S. Claus, H. Schwichtenberg, and J. Laufer, "Cloud4health - On Effective Ways to Deal with Sensitive Patient Data in a Secure Cloud Environment," in *43. Jahrestagung der Gesellschaft für Informatik e.V. (GI), Informatik angepasst an Mensch, Organisation und Umwelt*, 2013, pp. 376–384.
[355] BSI, "Technische Richtlinie. Kryptographische Verfahren: Empfehlungen und Schlüssellängen," Bundesamt für Sicherheit und Informationstechnik, Bonn, Germany, BSI TR-02102-1, Feb. 2016.
[356] BSI, "Technische Richtlinie. Kryptographische Verfahren: Empfehlungen und Schlüssellängen. Teil 2 - Verwendung von Transport Layer Security (TLS)," Bundesamt für Sicherheit und Informationstechnik, Bonn, Germany, BSI TR-02102-2, Feb. 2016.
[357] "TRESOR Web Page," *Trusted Ecosystem for Standardized and Open cloud-based Resources*. [Online]. Available: http://www.cloud-tresor.com/. [Accessed: 23-Jun-2016].
[358] D. Thatmann, M. Slawig, S. Zickau, and A. Küpper, "Deriving a Distributed Cloud Proxy Architecture for Managed Cloud Service Consumption," in *IEEE Sixth International Conference on Cloud Computing*, 2013, pp. 614–620.
[359] Zickau, S. and Küpper, A., "Towards Location-based Services in a Cloud Computing Ecosystem," in *Ortsbezogene Anwendungen und Dienste*, Universitätsverlag Chemnitz, 2012, pp. 187–190.
[360] D. Thatmann, M. Slawik, S. Zickau, and A. Küpper, "Towards a Federated Cloud Ecosystem: Enabling Managed Cloud Service Consumption," in *GECON*, 2012, vol. 7714, pp. 223–233.
[361] S. Zickau, D. Thatmann, T. Ermakova, Jonas Repschläger, R. Zarnekow, and A. Küpper, "Enabling Location-based Policies in a Healthcare Cloud Computing Environment," in *3rd IEEE International Conference on Cloud Networking, CloudNet 2014*, Luxembourg, 2014, pp. 333–338.
[362] J. van der H. johan.van.der.heide[at]itea3.org, "12011 BaaS," *itea3.org*. [Online]. Available: https://itea3.org/project/baas.html. [Accessed: 26-Jul-2016].
[363] "www.i-treasures.eu | Capturing the intangible." [Online]. Available: http://i-treasures.eu/. [Accessed: 02-Sep-2016].
[364] "Human–computer interaction," *Wikipedia, the free encyclopedia*. 16-Aug-2016.
[365] "Ofelia - Related Projects and Initiatives." [Online]. Available: http://www.fp7-ofelia.eu/about-ofelia/related-projects-and-initiatives/. [Accessed: 02-Sep-2016].
[366] E. Lupu *et al.*, "AMUSE: Autonomic Management of Ubiquitous e-Health Systems," *Concurr. Comput. Pract. Exp.*, vol. Volume 20, Issue 3, p. pp.: 277-295, 2008.

[367] S. Keoh *et al.*, “Policy-based Management for Body-Sensor Networks,” 2007, pp. 92–98.
[368] G. Russello, C. Dong, and N. Dulay, “A Policy-Based Framework for e-Health Applications,” in *Proceedings of the UK e-Science All Hands Meeting 2007*, 2007.
[369] J. Singh, J. Bacon, and K. Moody, “Dynamic Trust Domains for Secure, Private, Technology-assisted Living,” *2nd Int. Conf. Availab. Reliab. Secur. ARES 2007*, 2007.
[370] G. Russello, C. Dong, and N. Dulay, “Authorization and Conflict Resolution for Hierarchical Domains,” in *Policies for Distributed Systems and Networks*, 2007.
[371] F. Wang and K. J. Turner, “Towards Personalised Home Care Systems,” 2008.
[372] F. Wang and K. J. Turner, “Policy Conflicts in Home Care Systems,” *9th Int Conf Feature Interact. Softw. Commun. Syst.*, 2007.
[373] S. Ali and S. Kiefer, “Semantic Medical Devices Space: An Infrastructure for the Interoperability of Ambient Intelligent Medical Devices,” *IEEEEMBS Conf. Inf. Technol. Biomed.*, 2006.
[374] J. J. Carroll *et al.*, “Jena: Implementing the Semantic Web Recommendations,” 2003, pp. 74–83.

1. Appendix A: International and national data privacy constraints

There are some important differences in the privacy frameworks in those countries following the EU model and in the United States. The European approach, which is based on a view that privacy is a fundamental human right, generally involves top-down regulation and the imposition of across-the-board rules restricting the use of data or requiring explicit consent for that use. The United States, in contrast, employs a sectorial approach that focuses on regulating specific risks of privacy harm in particular contexts, such as health care and finance. This places fewer broad rules on the use of data, allowing industry to be more persuasive its products and services, while also sometimes leaving unregulated potential uses of information that fall between sectors [245].

Despite these important differences, both privacy frameworks are based on the "Fair Information Practice Principles" or "FIPPs," which since 1970s form the bedrock of modern data protection regimes. While the principles are later redefined in law and international agreements in different ways, at their core, the FIPPs articulate basic protections for handling personal data. They provide that an individual has a right to know what data is collected about him or her and how it is being used. The individual should further have a right to object to some uses and to correct inaccurate information. The organization that collects information has an obligation to ensure that the data is reliable and kept secure. The FIPPs form a common thread through national statutes of many countries and a variety of international agreements. Whether national policy is framed in terms of "privacy" or as "data protection", there has been, and will continue to be, a remarkable consensus on the basic legal principles [245][246] . They also formed the basis for the original U.S.-E.U. and U.S.-Switzerland Safe Harbor Frameworks [247][248], as well as they lay down the basis for a new EU-U.S. Privacy Shield [249], which harness the global consensus around the FIPPs as a means to build bridges between U.S. and European law.

There are three general and overlapping aspects of the FIPPs that critics argue are fundamentally challenged by Big Data and its implications [250][251]:

1. *The definition of personally identifiable information*, the collection and processing of which normally triggers regulation in this area. The line between personal and non-personal data is increasingly difficult to draw due to the fact that personal data can be more easily re-identified from the contributions of data elements which, on their own, say little or nothing about any one particular person. Additionally, in the context of IoT (Internet of Things), interferences about one's personal life and behavior can be more easily drawn by capturing data from the technical identifiers linked to various objects and based on interferences that are drawn about the categorical group to which one is presumed to belong. The use of related metadata, group data and aggregated data, feeds off the growing ambiguity about what is and what is not personally identifiable information.
2. The "data minimization" principle, which implies that organizations are required to limit the collection of personal data to the extent that is necessary to achieve their legitimate purposes and to delete what doesn't conform to those purposes. The business model of Big Data is antithetical according to these principles.
3. *A clear definition and transparent communication about the purpose for which personal data is being processed*. The indicative power of analytics presumes that new purposes will and should be found for personal data if the technology promise is to be realized, therefore also contradicting this principle.

The resulting debate about the regulatory flexibility for the Big Data analytics is thus one of the core data privacy debates of our time. Different alternatives which are complementary to the current privacy paradigm approaches and solutions, have been suggested, including: 1) an agent-based approach instead of patient-based approach [252]; 2) a shift from a negotiation of the time of

collection, in terms of specified, legitimate purposes, towards a focus on data use and management practices and consequent flexibility of re-use of data across context [251]; 3) a set of solutions that de-emphasize the role of individuals at the point of data collection by losing data collection and minimization of restriction with a shift towards empowering the individuals, e.g. allowing them to engage with the benefit of Big Data for their own particular purposes and introducing the obligation for the organizations to not only reveal the existence of their databases but also the criteria (not necessarily the mechanisms) used in their decision-process [253]; 4) development of practical and effective privacy guidelines for the private sector based on a crucial distinction between knowledge discovery and application. It is argued that the former comprises acquisition, pre-processing, integration, analysis and interpretation and in each phase, algorithms perform a variety of classificatory, associative and sequential tasks. Since the most part of the knowledge discovery phase doesn't involve analysis of a particular individual's data (which may be de-identified or pseudonymized), individuals are implicated but not affected and protection of privacy can be warranted [254][255]. The debates related to the revision of the European privacy framework and the open issues of interpretation and enforcement of the new General Data Protection Regulation (GDPR) well represent the complexity of the issue and the search for an appropriate balance between privacy and multiple competing interests. It is obvious that adapting the FIPPs and the regulatory privacy framework to address the Big Data challenges will take some time.
Meanwhile, it was argued that from the perspective of technical measures of privacy protection introduced by the current and changing laws and regulations, it is not in principle different from other aspects of data privacy and other uses of larger data sets. In particular, the challenges published in the list named "Top 10 Big Data Security and Privacy Challenges" (See Figure 53 below) by Cloud Security Alliance, are all of the challenges derived from general IT challenges. While it's imperative to master all these challenges for successful Big Data application, they do have a broader impact.

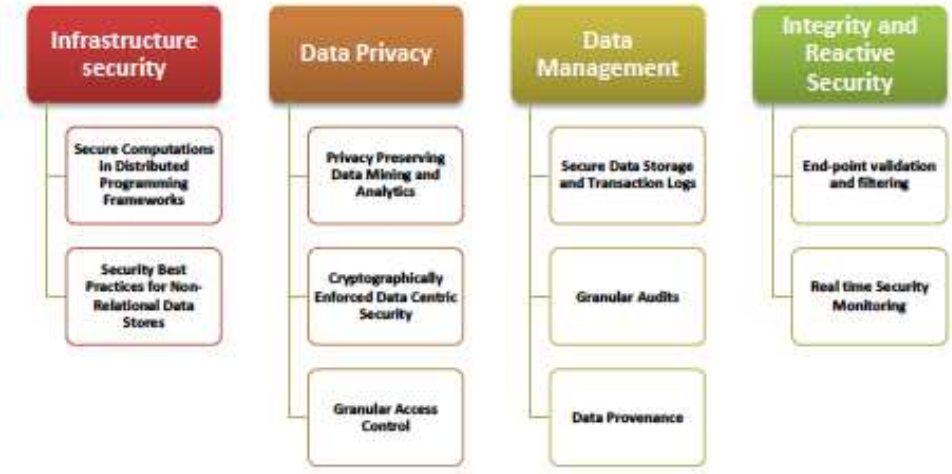

Figure 53 Classification of the Top 10 Big Data Privacy and Security Challenges [256]

However, from a technical perspective in the healthcare domain, the following aspects need to be taken into account:

- the existing data privacy enhancing methods need to be assessed for whether they satisfy all requirements within the scope of Big Data;
- the existing data privacy enhancing methods need to be adapted and enhanced to meet the requirements demanded for Big Data applications or new approaches need to be developed. In particular, the technology advancements are needed in hush algorithms, secure data exchange and de-identification algorithms.

Transnational provenance
One notable particularity of Big Data privacy and security is the growing trend to gather data from international sources and consequently a need to address data privacy regulations of multiple countries. This ranges from determining the applicable regulations in the first place to the ability to provide varying data privacy guarantees, depending on the nationality of data providers, users, processing etc.

1.1. Privacy and Security Regulation

1.1.1. Regulation in Europe

1.1.1.1. EU Directive 95/46/EC and accompanying Opinion 4/2007 on the concept of personal data

The main purpose of EU Directive 95/46 of 24 October 1995, that continues to be the key regulatory act on the matter until May 2018 when new harmonised data protection framework across the EU will enter into application, is to protect the fundamental rights and freedoms of natural persons and in particular their right to privacy, with regard to the processing of personal data. Personal data is the *"data related to an identified or identifiable individual"* and the directive sets the legal ground for the circulation and use of personal data along the following perspectives [257]:

- Fair and lawful processing
- Processing for limited purposes (no further incompatible processing)
- Adequate, relevant and not excessive
- Accurate and up to date
- Preservation no longer than is necessary
- Data subjects' rights (information and access)
- Secured processing (technically and organisationally)
- No transfer to third countries without adequate protection
- Notification to relevant regulator

The Directive also contains specific minimum requirements in terms of the processing of personal health information, which is categorised as a "special category of data" that requires special and additional protection in terms of obtaining, processing, security and disclosure (Article 8). As a summary:

Member States shall prohibit the processing of personal data unless:

- Explicit consent of the data subject is available for data processing; or
- Processing is necessary for the purposes of carrying out the obligations and specific rights of the controller; or
- Processing is necessary to protect the vital interests of the data subject or of another person where the data subject is physically or legally incapable of giving his consent; or
- Processing is carried out in the course of its legitimate activities with appropriate guarantees by a foundation, association or any other non-profit-seeking body and that the data are not disclosed to a third party without the consent of the data subjects; or
- Processing is necessary for preventive medicine, medical diagnosis, and treatment or healthcare services, with supervision by a health professional bound by professional secrecy.

The "Opinion 4/2007 on the concept of personal data" published by the Data Protection Working Party set up under Article 29 of Directive 95/46/EC [258] presents further clarification for the definition of personal data and the processing of personal data under certain circumstances

including clinical research. The aim is to establish the appropriate balance between protection of the data subject's rights on the one side, and on the other side the legitimate interests of data controllers, third parties and the public interest.

The definition of anonymous data is given as "*any information relating to a natural person where the person cannot be identified, whether by the data controller or by any other person, taking account of all the means likely reasonably to be used either by the controller or by any other person to identify that individua*l". In line with this definition, it is clearly presented that, *"the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable".* In this respect, if within a clinical research study, subject data is collected in an anonymous manner in line with the anonymous data definition provided by Opinion 4/2007, it is clear that the data protection rules set in Directive 95/46/EC shall not apply, i.e. explicit consent of data subject is not mandatory in this case[3].

The Opinion 4/2007 also elaborates on the case of pseudonymization. The definition of pseudonymization process is in line with that of ISO/TS 25237:2008: *"Pseudonymization is a particular type of anonymization that both removes the association with a data subject and adds an association between a particular set of characteristics relating to the data subject and one or more pseudonyms".* The definition of retraceable pseudonymization is provided where it is possible to re-identify the subject by using correspondence lists for identities and their pseudonyms or by using two-way cryptography algorithms. It is presented that retraceably pseudonymised data may be considered as information on individuals which are *indirectly identifiable,* and in this respect, data protection rules apply, yet it is presented that *"the risks at stake for the individuals with regard to the processing of such indirectly identifiable information will most often be low, so that the application of these rules will justifiably be more flexible than if information on directly identifiable individuals were processed".*

Regarding irreversible pseudonymization where no re-identification is possible, it is presented that "pseudonymization achieved by one-way cryptography algorithms generally creates anonymous data". It is presented that in cases where "the identification is not supposed or expected to take place under any circumstance, and appropriate technical measures (e.g. cryptographic, irreversible *hashing) have been put in place to prevent that from happening, the information processed by the original controller may not be considered to relate to identified or identifiable individuals taking account of all the means likely reasonably to be used by the controller or by any other person and hence its processing may thus not be subject to the provisions of the Directive".*

As briefly summarized, the Opinion 4/2007 document provides further clarifications, yet there are still grey areas, especially related with the decision of whether a data can be considered as anonymous data and hence can be exempted from the data protection rules. In this respect, in the document the essential role of National Data Protection Supervisory Authorities is emphasized in the framework of their missions of monitoring the application of data protection law, which involves providing interpretation of legal provisions and concrete guidance to controllers and data subjects.

[3] There are well-reported inconsistencies in the use and definitions of the terms "anonymization" and "de-identification". This document bases its terminology on the Technical Specification of the International Organization for Standardization (ISO/TS) 25237:2008 on the pseudonymization of health data. Though as some experts point out there is a problem with these definitions that some anonymization attempts have resulted in data being re-identified, implying that the data thought to be anonymized actually wasn't, for practical reasons in this section of this document the terminology used in the relevant legislation will be used in the same way, while in other sections the terms "de-identification" and "anonymization" will be used interchangeably with the understanding that sometimes de-identified information can be re-identified, and sometimes it cannot be.

Based on these guidelines, the Data Protection Officers in respective EU countries publish guidelines on how clinical data can be used for research purposes. The alternatives to be pursued in terms of precedence during secondary use of personal medical data are as follows:

1. Work on anonymous data,
2. If impossible to achieve the scientific purpose with the previous, work on pseudonymized data (key-coded data),
3. If impossible to achieve the scientific purpose with the previous, work on non-pseudonymized data (personal data).

As an example, the guidelines that are provided by the Data Protection Commissioner of Ireland [259] are also in line with Article 29 Working Party guidelines. The flowchart presented in Figure 54 below by the Data Protection Commissioner of Ireland presents the steps to be followed more clearly.

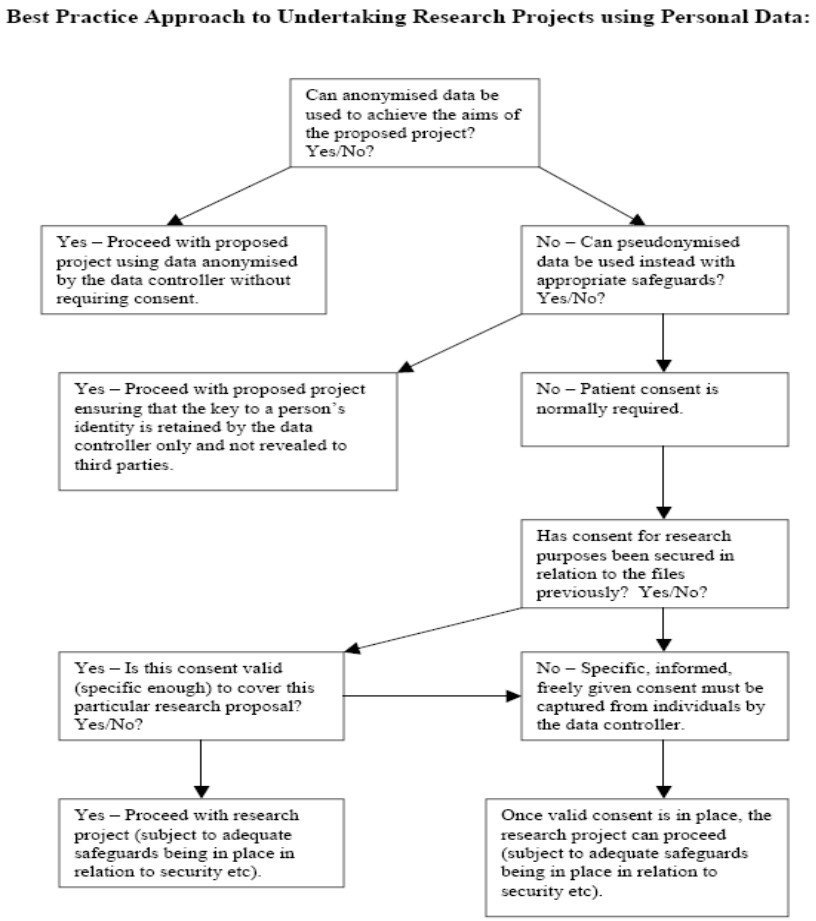

Figure 54 Guidelines provided by Data Protection Commissioner of Ireland [Hawkes 2007]

1.1.1.2. The EU General Data Protection Regulation

On April 27, 2016, the new data protection principles have been released as a Regulation of the European Parliament and of the Council (Regulation (EU) 2016/679) to regulate the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC [260]. It enters into application 25 May 2018 after a two-year transition period and, unlike a Directive it does not require any enabling legislation to be passed by governments [261].

The Regulation updates and modernises the principles enshrined in the 1995 Data Protection Directive to guarantee privacy rights. It focuses on: reinforcing individuals' rights, strengthening the EU internal market, ensuring stronger enforcement of the rules, streamlining international transfers of personal data and setting global data protection standards. .The reform provides tools for gaining control of one's personal data, the protection of which is a fundamental right in the European Union, by introducing new for the European data protection law principles and ensuring the following [262]:

- **A "right to be forgotten":** When an individual no longer wants her/his data to be processed, and provided that there are no legitimate grounds for retaining it, the data will be deleted.
- **Easier access to one's data:** Individuals will have more information on how their data is processed and this information should be available in a clear and understandable way. A right to data portability will make it easier for individuals to transmit personal data between service providers.
- **Breach notification:** Companies and organisations must notify the national supervisory authority of data breaches which put individuals at risk and communicate to the data subject all high risk breaches as soon as possible so that users can take appropriate measures.
- **Data protection by design and by default:** 'Data protection by design' and 'Data protection by default' are now essential elements in EU data protection rules. Data protection safeguards will be built into products and services from the earliest stage of development, and privacy-friendly default settings will be the norm – for example on social networks or mobile apps.
- **Stronger enforcement of the rules**: data protection authorities will be able to fine companies who do not comply with EU rules up to 4% of their global annual turnover.

1.1.1.3. Legal Background on the Security and privacy in mobile e-Health solutions

Mobile Health (mHealth) is a rapidly growing sector stemming out of the convergence between healthcare and ICT (Information and Communication Technology). It includes mobile applications designed to deliver health and well-being services through smart devices often processing large volume of personal information about health, lifestyle and well-being of individuals including the data coming from various different medical sensors. App developers, most of which are individuals, or SMEs, unaware of the data protection requirements, may create unwanted threats to the privacy of users of smart devices. Personal data protection is a fundamental right in Europe, enshrined in Article 8 of the Charter of Fundamental Rights of the European Union, as well as in Article 16(1) of the Treaty on the Functioning of the European Union (TFEU). The current relevant legal framework applicable is composed of the Data Protection Directive [263] and the ePrivacy Directive [264]. The Article 29 Working Party also published an Opinion "n apps on smart devices", which seeks to clarify the legal obligations of each of the parties involved in the development and distribution of apps [265]. The Opinion offers some guidance to all the players, in particular the need to provide clear and unambiguous information about data processing to users (e.g. the types of data processed, the purposes for processing and data retention periods).

The existence of large number of lifestyle and wellbeing apps available with no clear evidence on their safety, quality and reliability has possible negative effects on the ability of consumers to make an informed choice, and also the quality of data to be linked to the electronic health records. Due to this potential lack of trust from the individuals and also from the traditional healthcare delivery organizations, mHealth is currently not used to its full potential in the European healthcare systems.

To remedy this situation, several national and international efforts are ongoing. App certification programmes are already emerging: In UK, the National Information Board is currently in an effort to provide citizens with access to a set of applications which have passed a review to prove their safety and compliance with data protection rule and as a result have been *endorsed* by the NHS [266]. In parallel with this effort, BSI introduced *PAS 277:2015 on Health and wellness apps – Quality criteria*

across the life cycle – Code of practice, as a set of quality criteria for the development, testing and releasing of health and wellness apps [267]. In Portugal, there is an ongoing effort a to provide vision paper for mHealth in Portugal, to identify the main principles about how to classify and evaluate mobile health apps and criteria for prescription of apps [268]. In Catalonia, through the mHealth Master Plan [269] approved in February 2015, the government aims to build a mHealth marketplace of accredited mobile applications in the field of health and social services.

In parallel with these national efforts, on 10 April 2014 the European Commission published a Green Paper on mHealth [270] which launched a public consultation for identifying the existing barriers and issues related to mHealth deployment and the right way forward to unlock mHealth potential. The responses to the consultation revealed the need for a strengthened enforcement of data protection and the rules applicable to mHealth devices possibly through certification schemes or quality labelling of lifestyle and wellbeing apps. Following this, a number of mHealth stakeholder meetings took place interactively addressing ongoing and potential future policy actions in the field of mobile health. To address the issue of legal clarity, the Commission has started preparations to develop a pro-innovation legal framework aiming to clarify the legal status of health and wellness apps as consumer products. The Commission also announced the intention of facilitating the development of a European standard on quality criteria for the development of health and wellness apps, taking as a basis the publicly available specification PAS:277 from BSI. On Certification, the Commission explained that it intends to build on and support the existing initiatives on voluntary certification rather than set up new mechanisms at the EU level. The Commission proposed for discussion a possible future collaboration between interested Member State public authorities, to develop common assessment methodologies, to facilitate mutual recognition or to build a common platform for certified health apps. The Commission also introduced the idea of developing guidelines at the EU level for assessing validity of data for the purposes of linking apps to the electronic health records (EHR). The aim is to narrow the scope and focus efforts to public service use, i.e. not to assess all apps but only those which declare the intention to be linked to the EHR. In parallel with these efforts, the draft privacy Code of Conduct on mobile health apps [271], has been prepared and being discussed through the stakeholders meetings at EU level. The aim of the code is to facilitate data protection compliance and to promote good practices in this field, by providing specific and accessible guidance to app developers on how European data protection legislation should be applied in relation to mHealth apps.

1.1.2. Regulation in the United States

In the United States, personal data privacy regulation within the healthcare environment is governed by the Security and Privacy Rules of the Health Insurance Portability and Accountability Act (HIPAA) [272]. The complete suite of HIPAA Administrative Simplification Regulations can be found at 45 CFR Parts 160, 162, and 164, and includes:

- Privacy Rule
- Security Rule
- Enforcement Rule
- Breach Notification Rule

The HIPAA rules and subsequent guidelines published by the Office of Civil Rights (OCR) at the US Department of Health created a set of standards for establishing when health-care information was no longer “individually identifiable” – or would be considered “de-identified”. The Privacy Rule set national standards for the protection of individually identifiable health information by three types of covered entities: health plans, health care clearinghouses, and health care providers who conduct the standard health care transactions electronically. The Privacy Rule created particular situations where patients consent was presumed (in the area of treatment, payment, and health-care operations), as well as certain public policy areas where disclosure of patient information was permitted for other public goals (such as establishing rules for disclosure in connection with fraud investigations, litigation, certain public health activities, and otherwise). The Security Rule sets

national standards for protecting the confidentiality, integrity, and availability of electronic protected health information. The Enforcement Rule provides standards for the enforcement of all the Administrative Simplification Rules. HHS enacted a final Omnibus rule that implements a number of provisions of the Health Information Technology for Economic and Clinical Health (HITECH) Act Omnibus Rule to strengthen the privacy and security protections for health information established under HIPAA, finalizing the Breach Notification Rule. The last one requires HIPAA covered entities and their business associates to provide notification following a breach of unsecured protected health information. The OCR and Human Services (HHS) administer and enforce the Privacy Rule and the Security Rule.

1.1.3. NIST Risk Management Framework

The need for controls, policies, and procedures is somewhat dependent on the different national regulatory requirements for the MEDOLUTION countries. A useful foundation to begin with is the Risk Management Framework developed by the National Institute of Standards and Technology (NIST), which is a non-regulatory agency of the United States Department of Commerce. The NIST develops and issues standards, guidelines, and other publications to assist federal agencies in implementing the Federal Regulation. Given the proposed collection and retention of data determined as being subject to protection as outlined under the HIPAA Security rule in North America, prudent consideration needs to be given to system security implementation. As prescribed by the NIST Special Publication 800-66: An Introductory Resource Guide for Implementing the Health Insurance Portability and Accountability Act (HIPAA) [273], at a foundational level an entity must:

1. Ensure the confidentiality, integrity and availability of the electronic personal health information (EPHI) that it creates, receives, maintains, or transmits;
2. Protect against any reasonably anticipated threat and hazards to the security or integrity of EPHI; and
3. Protect against reasonably anticipated uses and disclosures of such information that are not permitted by the HIPAA Privacy Rule.

In order to comply with the HIPAA Security Rule, entities must go about implementing effective risk management procedures to mitigate threats related to:

- Confidentiality, defined as “the property that data or information is not made available or disclosed to unauthorized persons or processes”;
- Integrity, defined as “the property that data or information have not been altered or destroyed in an unauthorized manner”; and
- Availability defined as “the property that data or information is accessible and useable upon demand by an authorized person”.

(See Figure 55 below).

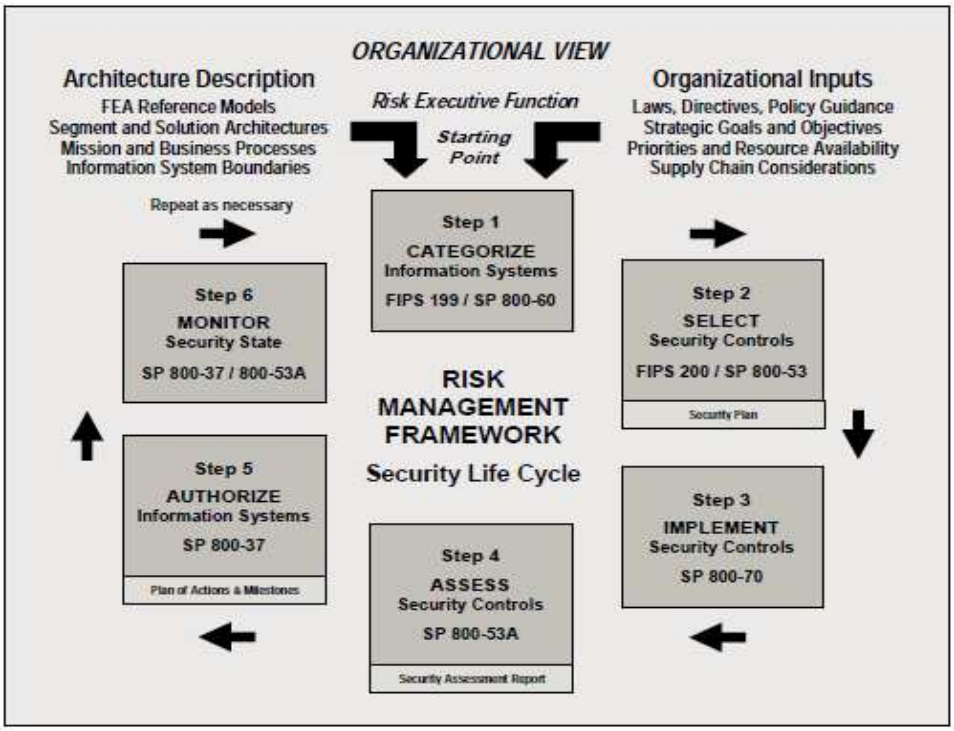

Figure 55 NIST Risk Management Framework [273]

In response to these criteria and in order to facilitate compliance, the following security requirements outline the most rudimentary non-functional requirements (NFRs) that are compulsory to achieving administrative and technical HIPAA compliance.

1.1.3.1. HIPAA Security Rule Administrative Safeguards

HIPAA Security Rule Administrative Safeguards can be summarized as follows:

NFR.01 Security measures shall be sufficient to reduce risks and vulnerabilities to an appropriate level (164.306(a)) of compliance.
- a. Safeguards in place to ensure confidentiality, integrity, and availability of EPHI
- b. Protected against all reasonably anticipated threats or hazards
- c. Protection against reasonably anticipated disclosures not permitted

NFR.02 Audit information must reside on a separate server, apart from the rest of the system.
- a. Audit information to be stored on a separate server from EPHI

NFR.03 The System must create and maintain retrievable exact copies of EPHI.
- a. System maintains exact copies

NFR.04 The system must continue to operate business processes critical to data security while operating in emergency mode.

1.1.3.2. HIPAA Security Rule Technical Safeguard

HIPAA Security Rule Technical Safeguards can be summarized as follows:

NFR.05 Each system user shall be assigned a unique name and/or identifying number.
- a. Ensure all activity is traceable to a specific individual user

b. Ensure necessary data is available in system logs to support audit
c. Every system user has a unique user name and password

NFR.06 The system shall include a mechanism to corroborate that EPHI has not been altered or destroyed in an unauthorized manner.
a. System provides safeguards to authenticate data against access and modification

NFR.07 System includes security measures to ensure that electronically transmitted EPHI is not improperly modified without detection.

NFR.08 System includes a mechanism to encrypt data during transmission.
a. HTTPS must be used if the application is accessing and presenting user data.

NFR.09 System includes a mechanism to encrypt data at rest.
a. All files and database tables containing EPHI must be encrypted to prevent unauthorized tampering or access.

These requirements frame the constraints, which the system must ultimately adhere to without detailing how these objectives are to be achieved through design. It is recommended that a thorough risk assessment be conducted as a part of the system design by administering the NIST Risk Management Framework (see Figure 4). As prescribed by the framework, the standards outlined in FIPS 199 Standards for Security Categorization [274] and FIPS 200 Minimum Security Requirements [275] should be referenced in order to properly categorize information risk and subsequently select the appropriate security controls. In order to duly stress the potential impact that this assessment may have to both the scope of security protocols and system design, the first two steps of this risk assessment have been outlined. It should be noted that the minimum security requirements necessary to achieve compliance cover seventeen security related areas, outlined in Appendix D. Specifications for Minimum Security Requirements. In response to these criteria and in order to facilitate compliance, the security requirements contained above outline the most rudimentary non-functional requirements that are compulsory to achieving administrative and technical HIPAA compliance. These requirements frame the constraints, which the system must ultimately adhere to without detailing how these objectives are to be achieved.

1.1.3.3. HIPAA Privacy Rule

A key aspect of HIPAA is the principle of “minimum necessary” use and disclosure. This principle ensures that covered entities make reasonable efforts to use, disclose, and request only the minimum amount of protected health information needed to accomplish the intended purpose of the use, disclosure, or request [245][276].

Additionally, Section 164.514 of the HIPAA Privacy Rule stipulates that “health information that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual is not individually identifiable health information”. Section 164.514 (b) of the Privacy Rule contains the implementation specifications that a covered entity, or affiliated business associate, must follow to meet the identification standard. In particular, the Privacy Rule outlines two routes by which health data can be designated as de-identified. These are illustrated in Figure 56 below. Neither method promises foolproof method of de-identification with zero-risk re-identification. Instead, the methods are intended to be practical approaches to allow de-identified healthcare information to be created and shared with a low risk of re-identification.

1. "Safe Harbor" method (Section 164.514 (b)(2))**.** Safe Harbor requires the manipulation of 18 fields in the dataset as described in Table 2 below. The Privacy Rule requires a number of these data elements to be "removed". However, there may be acceptable alternatives to actual removal of values as long as the risk of reverse engineering the original values is very small. (See Figure 56 below).

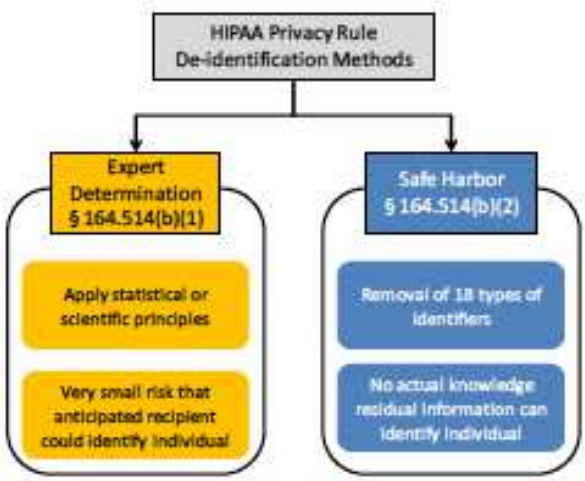

Figure 56 The two de-identification standards in the HIPAA Privacy Rule [277].

Table 2 The Safe Harbor De-Identification Standard

1. Names;
2. All geographic subdivisions smaller than a state, including street address, city, country, precinct, zip code, and their equivalent geocodes, except for the initial three digit of a zip code if, according to the current publicly available data from the Bureau if the Census:
    a. The geographic unit formed by combining all zip codes with the same three initial digits contains more than 20,000 people; and
    b. The initial three digits containing 20,000 or fewer people is changing to 000.
3. All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older;
4. Telephone numbers;
5. Fax numbers;
6. Electronic mail addresses;
7. Social Security Numbers;
8. Medical record numbers;
9. Health plan beneficiary numbers;
10. Account numbers;
11. Certificate/ licence numbers;
12. Vehicle identifiers and serial numbers, including license plate numbers;
13. Device identifiers and serial numbers;
14. Web universal resource locators (URLs);
15. Internet Protocol (IP) address numbers;
16. Biometric identifiers, including finger and voice prints;
17. Full face photographic images and comparable images; and
18. Any other identifying number, characteristics, or code.

Assumptions of the Safe Harbor method are the following:

- There are only two quasi-identifiers that need to be manipulated in a data set: dates and zip codes.
- The adversary does not know who is in the data set
- All dates are quasi-identifiers.

The Safe Harbor method is heavily influenced by Sweeney's research (it cites her work and pays specific attention to the quasi-identifiers that she identified for generalization) and appears designed to strike a balance between a risk of de-identification and the need to retain some utility in the data set. Additionally, the application of Safe Harbor is straightforward and it offers the promise of a known result, namely, a data set is legally de-identified. However, there are instances in which dates and more fine-grained geographic information are necessary. In practice the Safe Harbor standard would remove critical geospatial and temporal information from the data (see items 2 and 3 in the Table 1), potentially reducing the utility of the data. Many meaningful analyses of health data sets require the dates and event order to be clear. For example, in a Safe Harbor data set, it would not be possible to include the dates when adverse events occurred. It should also be taken into account that there is disagreement regarding the effectiveness of the HIPAA Safe Harbor [242].

2. Expert Determination method (Section 164.514 (b)(1))**:** In recognition of the limitations of de-identification via Safe Harbor, the HIPAA Privacy Rule provides an alternative in the form of the Expert Determination method. This method has three general requirements:

- *The identification must be based on generally accepted statistical and scientific principles and methods for rendering information not individually identifiable*. This means that the coved entity needs to ensure that there is a body of work that justifies and evaluates the methods that are used for the identification and that these methods must be generally known (i.e., undocumented methods or proprietary methods that have never been published would be difficult to classify as "generally accepted").
- *The risk of re-identification needs to be very small, such that the information could not be used, alone or in combination with other reasonably available information, by an anticipation with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information*. However, the mechanism for measuring re-identification risk is not defined in the HIPAA Privacy Rule, and what would be considered very small risk also is not defined. Therefore, the de-identification methodology must include some manner of measuring re-identification risk in a defensible way and have a repeatable process to follow that allows for the definition of *very small* risk.
- Finally, *the methods and results of the analysis that justify such determination must be documented*. The basic principles of de-identification are expected to be consistent across all clinical trials e.g., but the details will be different for each study, and these details also need to be documented [278].

Apart from this section, neither Privacy Rule nor the implementation guidelines provided by OCR specify the standards or qualifications for the expert, not they specify requirements for organizations using experts to release the expert's determination or even to acknowledge that the expert determination has been made.

Under either methods, *"the covered entity performing the de-identification must not have actual knowledge that the information could be used alone or in combination with other information to identify an individual who is a subject of the information" (*Section 164.514 (c)*).* However, HHS has

specifically clarified that simply knowing about the existence of re-identification techniques does not meet the “actual knowledge” standard [277].

Where information met the regulatory requirements for de-identification, health-care information was considered de-identified by law, and therefore was no longer subjected to the HIPAA restrictions on the uses of the individually identifiable information. Because the “individual” component of this information had been removed, this de-identified information could then be used and disclosed for a wide variety of purposes (including research and public health purposes, as well as commercial purposes), without creating meaningful privacy risks for any individuals. The HIPAA Privacy Rule acknowledges and supports the benefits of these uses of de-identified information while at the same time recognizes that any material privacy interests have been eliminated through that material privacy interests have been eliminated through this de-identification process. The HHS, in developing the HIPAA standards specifically wanted to ensure “the Privacy Rule would not be a disincentive for covered entities to use or disclose de-identified information wherever possible” [279].

There is no legislative or regulatory requirements to obtain consent from participants to share their de-identified data [278]. From a risk management perspective in situations in which consent has been provided (e.g. by trial participants or notice has been given to the participants) multiple levels of notice and consent can be used for disclosure of de-identified data with different degree of invasion of privacy. However, it should be considered that retroactively obtaining participant consent to de-identify data and use them for secondary analysis may introduce bias in the data [280]. If de-identification is a permitted use under the relevant regulations, then de-identification can proceed without seeking participant consent. Whether that is the case will depend on the prevailing jurisdiction. Thus, HIPAA and extensions under HITECH Act Omnibus Rule, de-identification is a permitted use by a covered entity. However, business associate can de-identify a data set only if the business associate agreement explicitly allows for that. Silence on de-identification in a business associate agreement is interpreted as not permitting de-identification [278].

1.2. De-identification, Re-identification, and Data Sharing Models

As it was discussed, various laws and regulations recognize the importance and utility of data de-identification, considering it as a technical control that can be applied to data, removing personal information and allowing the data that remains to be used in a way that will not disclose the identities of the data subject. However, de-identification approaches based on suppression or generalizing specific fields in a database cannot provide absolute privacy guaranties, because there is always a chance that the remaining data can be re-identified using auxiliary datasets. Besides that, additional privacy issues might result from the disclosure of specific attributes of that dataset linked to the identities. Technical details of the de-identification and re-identification methods for Big Data systems are reviewed in more details in the Section 8.3 on Data Privacy Strategies.

1.2.1. Models for Privacy-Preserving use of Private Information

Two distinct models for using personal information while protecting the privacy of the data subject have been identified in the literature. Both models are considered to be “privacy preserving” in that they are intended to allow the release of some information (e.g. aggregated information, statistical results, classifiers, or synthetic information) without revealing information that can be attributed to a specific individual within the original dataset [242].

Privacy Preserving Data Mining (PPDM): In this model, data are not released, but instead of that are re-used for statistical processing or Machine Learning. The results of the calculations maybe released in the form of statistical tables based on summarization and aggregations, classifiers that implement Machine Learning algorithms, and other kinds of results.

- *Statistical Disclosure Limitations* "is the discipline concerned with the modification of statistical data in order to prevent third parties working with these data to recognize individuals in the data"[281]. Techniques developed for disclosure limitation include generalization of reported information to broader categories, swapping data between similar entities, and the addition of noise in reports [281].
- *Differential Privacy* is a set of technics based on a mathematical definition of identity disclosure and information leakage from operations on a dataset. Differential privacy prevents disclosure by adding non-deterministic noise (usually small random values) to the results of mathematical operations before the results are reported [282].

Privacy Preserving Data Publishing (PPDP): In this model, data are processed to produce a new, de-identified or synthetic data product that is distributed to users. The goal of PPDP is to provide data that have high utility without compromising the identity of the data subjects.

- *De-identification* is designed to protect individual identity, making it hard or impossible to learn of the dataset related to a specific individual, while preserving some of the dataset's utility for other purposes. This method is reviewed in more details in the Section on Privacy and security solutions for IoT and Big Data systems.
- *Synthetic data generation* uses some PPDM techniques to create a dataset that is similar to the original data, but where some or all of the resulting data elements are generated and do not map to actual individuals. As such, synthetic data generation can be seen as a fusion of PPDM and PPDP [242].

1.2.1.1. Release models and data controls

One way to limit the chance of re-identification is to place controls on the way that the data may be obtained and used. These controls can be classified according to different release models. Several models have been proposed in the literature, ranging from no restrictions to tightly restricted. The following models are of particular interest:

- **The Data Use Agreement (DUA) model** [278]: The de-identified data may be made available to under a legally binding data use agreement that details what can and cannot be done with the data. Typically, DUAs prohibit attempted re-identification, linking to other data, from sharing the data without permission or other redistribution of the data. A DUA will typically be negotiated between the data holder and qualified researchers (the "qualified investigator model"), although they may be simply posed on the Internet with a click-through license agreement that must be agreed to before the data might be downloaded (the "click-through model").

  Under the HIPPA Privacy Rule, DUAs are required when sharing limited datasets, which are de-identified to a lesser standard, and not necessarily when sharing datasets that have been de-identified according to the Privacy Rule (e.g. under the Expert Determination method, the expert may require a DUA) [242] Some experts suggests, that in order to strengthen a DUA consenting parties could also add to their DUA a promise from the data provider that the data had been stripped of personal identifiers but still might be re-identifiable. Recipients would then face civil and criminal penalties if they attempt to re-identify [283]

- **The Enclave model** [278] [284]: The de-identification data may be kept in some kid of segregated enclave that restricts the export of the original data, and instead accepts queries from qualified researchers, runs the queries on the de-identified data, and responds with results.

From practical perspective it should also be noted that, given a possibility to re-identify individuals from very limited personal data, no data set can be considered truly anonymous even when devoid of all elements of personal health information. Therefore it is recommended in order to minimize risk to individuals, that all sharing agreements, including data use agreements and agreements to share de-identified data, should include provisions requiring that the recipient to promise not to make any attempt to re-identify the individuals in the data set [278].

1.3. Resources ownership

In healthcare domain the transition from paper to electronic records has created new opportunities for sharing information among healthcare providers, between physicians and patients, and with third parties. The mobile revolution and the rise of wearable tech, namely devices with sensors measuring the user's daily activities and habits, have further heightened consumers' expectations about data sharing and have posed new legal challenges. But this shift is happening quickly, in many ways too quickly for either physicians or the laws and regulations pertaining to medical records and data to keep up.

The rights in relationship to the data and the right to use data are covered by various regulations, in particular by the IPR law (copyright, database law and confidentiality), contracts, data privacy and regulations. For Big Data technologies, it is particularly important to understand when and how further processing of Big Data sets creates a new ownership. It is argued that the collection, curation and combination with other datasets and eventually analysis of data sets derive new rights to the resulting data that must be asserted and enforced. Therefore, there has been an ongoing debate about potential "ownership" for such resulting data, which has been of economic interest for data processing companies.

Data and related rights are often treated differently based on private use versus public use. In the public sector, including the healthcare domain, the main issue related to data ownership is that normally the legislation and applicable regulations do not allow the use of data for purposes other than those regulated and for which the data was collected. Thus, reuse of public sector data in Big Data processes has to be carefully checked with data protection in the public sector. Additionally, data ownership among public bodies is also an important issue closely related with the data protection in the public sector. As for the data not generated by the public sector, there is often a lack of legislation concerning accessing them, which creates grey areas and uncertainty [285]. It should be also considered the related regulatory approach may vary significantly from country to country. Some examples from the main jurisdictions are given below.

Who owns the records in the US [286]?

There is no consensus on who owns medical records. The HIPAA does not specify ownership, and state laws are inconsistent. According to an analysis of state laws by Health Information & The Law, a project of the George Washington University's Hirsh Health Law and Policy Program and the Robert Wood Johnson Foundation, in 2015 only New Hampshire had a law stating that patients own their medical records. In 20 other states, providers own them. By that time, the rest of the states had no legislation addressing the matter.

Many states have specific laws addressing how providers must maintain, protect, and dispose of records, as well as laws giving patients, providers, and others access to medical records, regardless of ownership status. In addition, patients in all states have many rights with respect to their medical records under the HIPAA Privacy and Security Rules, as it was discussed earlier. (See Figure 57 below).

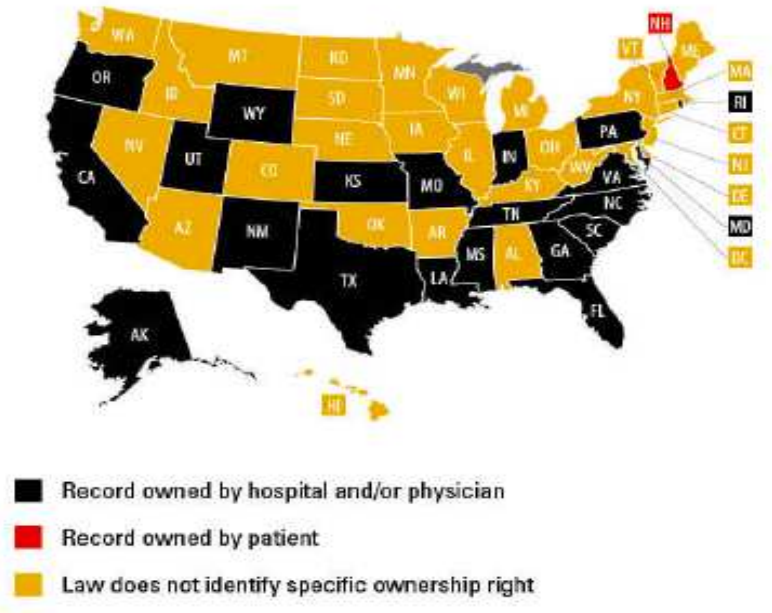

Figure 57 Medical record ownership in the US, by State (State: August 2015) *[286]*

Who owns the records in the UK?

In the United Kingdom, ownership of the NHS's medical records, in the past, has generally been described as belonging to the Secretary of State for Health [287] and this is taken by some to mean that copyright also belongs to the authorities [288].

Who owns the records in Germany?

In Germany, a relatively new law, established in 2013, strengthens the rights of patients. It states, amongst other things, the statutory duty of medical personnel to document the treatment of the patient in either hard copy or within the electronic patient record (EPR). This documentation must happen in a timely manner and encompasses each and every form of treatment the patient receives, as well as other necessary information, such as the patient's case history, diagnoses, findings, treatment results, therapies and their effects, surgical interventions and their effects, as well as informed consents. The information must include virtually everything that is of functional importance for the actual, but also for future treatment. This documentation must also include the medical report and must be archived by the attending physician for at least 10 years. The law clearly states that these records are not only memory aids for the physicians, but also should be kept for the patient and must be presented on request.

In addition, an electronic health insurance card was issued in 2014 which is applicable in Germany (Elektronische Gesundheitskarte or eGK) [289], but also in the other member states of the European Union (European Health Insurance Card). It contains data such as: the name of the health insurance company, the validity period of the card, and personal information about the patient (name, date of birth, sex, address, and health insurance number) as well as information about the patient's insurance status and additional charges. Furthermore, it can contain medical data if agreed to by the patient. This data can include information concerning emergency care, prescriptions, an electronic medical record, and electronic physician's letters. However, due to the limited storage space (32kB), some information is stored on servers.

Who owns the records in The Netherlands [290]?

The Netherlands is one of the pioneer countries when it concerns patient rights. The rights of patients have acquired their place in the Dutch legal system by the codification of the general rights of the patient as part of the Dutch civil code in 1995. The main purpose of the Act is to clarify and strengthen the legal position of the patient. The scope of the legal provisions on patient rights also extends to medical actions that are not performed in the frame of a contract in as far as the nature of the situation allows for the application of the provisions.

The right of access to medical records is derived straight from the right of privacy, which is practically without clauses. Access may be denied only if the granting would result in injuring the privacy of a third party. Even if a medical practitioner fears a patient may be damaged by the information about him contained in his medical records, he is not justified in denying access. On the patient's request access is granted at the earliest opportunity as well as a copy of the medical record. The right of access to medical records is of major importance whenever a patient considers taking legal action against a doctor or an institution. The patient must allege and prove, according to the main rule of Article 177, Civil Legal Procedure Code, unless any other regulation or reasonableness and fairness demand a different partitioning of the burden of proof. Not accessible are the doctor's personal work notes, that is, notes outside the scope of communication. The doctor can charge the patient for a copy of his medical records. The doctor will keep the records for at least ten years from the time they were made. They may be kept for a longer period, if, in reason, the care extended by, a good medical practitioner requires this.

Who owns the records in France?

France disposes of extensively developed patient rights legislation with the Act No. 2002-303 concerning the rights of patients and the quality of the health system approved in 2002, followed by the Act No. 2005-370 concerning the rights of patients and the end of life which came into force in 2005. Both Acts amended in the first place the Public Health Code.

The Code of Public Health provides for everyone the right to access to the data concerning their health which is kept in a medical file by a health care providers or health care institutions: more specifically this data includes the results of researches, notes of consultations, interventions and hospitalizations, protocols and therapeutic regulations, surveillance-files (“feuilles de surveillance”), correspondence between health care providers, with exception of information which was obtained of third parties who were not involved in the treatment or information related to these third parties. Access can occur directly or indirectly through a physician. An answer to a request of access must be given at last within eight days after the request and at earliest after a waiting period of 48 hours. When the requested data is more than 5 years old, this waiting period is brought up to 2 months.

“Quantified self” Datasets

With the recent tendency towards self-track and quantify there is an interest and new legal challenge related to the right of the user to the commonly created data sets, even though there are still corporations acting as a data controller by providing tools for data analysis and storage on their servers. Although the law has not yet offered a concrete answer to the issue of ownership of such “quantified self” datasets co-created by the users and the corporations, there is a growing tendency to allow the user for more control and ownership rights over his data with techno-legal solutions and alternative market models [291]. Personal Data Vaults (PDS) are currently one of the main technical solutions put forth in order to allow the user to gain control of his data back from the various corporations acting as info-mediaries in the big data market. The idea is to develop a privacy enhanced architecture enabling the user to access, control and trace their data once shared online [292]. In this trend, there are many suggestions employing technical means for the user to reclaim control over his/her data. One example might be the MIT Open PDS app, which allows the user to see third-party requests for his/her data and make informed decisions [293]. Another example is Cozy cloud, a French company that provides users with open sourced private clouds to store their personal data. Besides that, there is a rising number of start-ups, such as “Personal”,

"Reputation.com" and "Datacoup", whose aim is to help the user monetize and control own data. Though at present data controllers have the most control over data under the database right protection and are thus the primary beneficiaries of the value extracted from Big Data, there seems to be a slight shift and more calls on both sides of Atlantic towards empowering the user to control and perhaps "own" his data [255]. However, it should be considered that law is still admittedly lagging behind in terms of providing user with more control over his data. There are barriers in the related established laws and regulations that impede such a potential "data ownership" right and a debate on a need and way of modernisation of these fields of law is ongoing.

The issues of potential Data Ownership are controversial and need to be addressed on various levels, in particular on national and/vs. Regional (European) regulations and in the context of trans-border flow of data. From a practical perspective, as long as the current process of regulatory review is ongoing, Big Data companies and Medolution project partners must deal on a case by case basis as to whether the particular solution or use case operates within the legal bounds as defined by the applicable data privacy and data copyright law and contract law [285][294][255], including approaches suggested under data sharing models discussed earlier.

1.4. Medical and security constraints

1.4.1. Medical devices certification

Medical device manufacturers are subjected to national and international laws and norms. Only medical devices which comply with them can be placed on the appropriate market.

1.4.1.1. EU - CE marking certification for medical devices

To place a medical device on the EU-market, it has to be certified with the CE-marking (See Figure 58). The CE-certification is not a quality mark. Instead, it indicates that the medical device meets all requirements of the appropriate EU-directive. To achieve the CE-certification, a conformity assessment procedure has to be carried out to determine, that all CE-requirements are satisfied. The conformity assessment procedure is conducted and controlled by a notified body, which is an organization accreted by an EU-Member State. Figure 59 Basic Steps to achieve CE-Certification illustrates the process. Additionally, every EU-Member State has further regulations, which have to be fulfilled [295], [296].

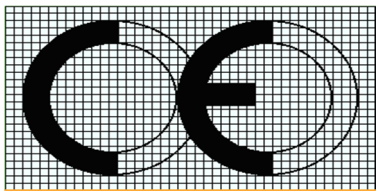

Figure 58 CE-Certification Marking

[297]

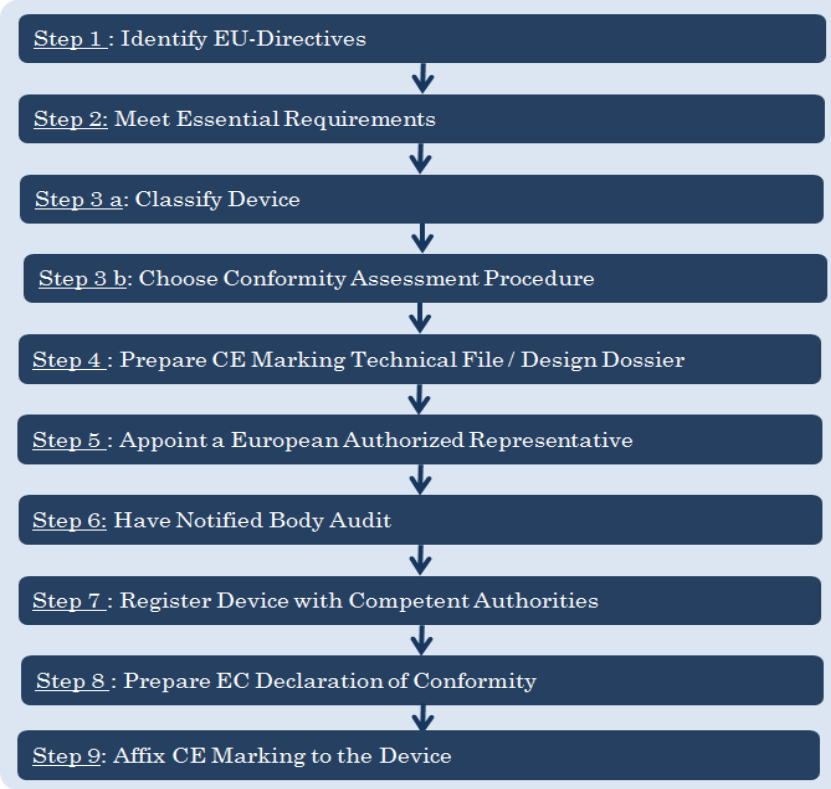

Figure 59 Basic Steps to achieve CE-Certification

[295], [296], [298], [299], [300], [301]

1.4.1.1.1. Step 1: Identify EU-Directives

The European Union provides three directives for the market approval. Each medical device has to be assigned to one of these directives:

1. Directive 93/42/EEC - Medical Device (MDD) [296] describes the essential requirements for medical devices to achieve the CE-marking.
2. Directive 98/79/EC - In-Vitro Diagnostic Devices (IVD) [302] describes the essential requirements for in-vitro diagnostic devices to achieve the CE-marking.
3. Directive 90/385/EEC - Active implantable medical devices (AIMD) [303] describes the essential requirements for implantable medical devices to achieve the CE-marking.

1.4.1.1.2. Step 2: Meet Essential Requirements

To achieve the CE-certification, the medical device must also meet the essential requirements, which are set out in Annex I of Directive 90/385/EEC. The other directives have similar essential requirements.

Harmonized European Standards:
To demonstrate conformity with the essential requirements, it is desirable to use harmonized European standards. The harmonized standards are technical specifications, which defines requirements for the product, the production process, services, and test-methods. The usage of the harmonized standards is voluntary. The following harmonized standards might be the most interesting for most medical device and medical software. A list of all harmonized European standards can be found in [298]:

- *EN ISO 13485:2012 - Medical devices - Quality management systems - Requirements for regulatory purposes* (ISO 13485:2003 QMS) [295], [304]: A QMS ensures that a product conform all consumer and legal requirements. Further, it describes procedures, processes, responsibilities and requirements a medical device has to meet. In particular, QMS defines requirements in regard to the documentation, management responsibility, resource management, product development phases as well as measurement, analysis and improvement.

- *EN ISO 14971:2012 - Medical devices - Application of risk management to medical devices* (ISO 14971:2007, Corrected version 2007-10-01): The purpose of risk management is both to ensure the quality of the product and to ensure the compliances with the official requirements. Further, the medical product has to be validated concerning its criticality and based on that, arrangements have to be made to control and minimize those criticalities. ISO 14971 “Risk management to medical devices” defines international requirements of risk management systems for medical devices. The results of a risk management will show if the medical device is adequate safe, for its intended purpose, to place it on the market.

- *EN 60601-1:2006 - Medical electrical equipment - Part 1: General requirements for basic safety and essential performance:* This standard defines general requirements for basic safety and essential performance of electrical devices that are used for the diagnostic, treatments, and monitoring of patients. This standard is only usable for medical devices which have physical contact to the patient [305] .

- EN 62304:2006 - Medical device software - Software life-cycle processes: This standard is the basic for medical software integrated in medical devices or as a stand-alone-medical device. It describes the requirements for the development and the maintenance of medical software. It is assumed that software for medical devices are developed and maintained within a QMS and a risk management system. During the risk analysis of the risk management process, hazards that could be caused by software have to be considered. The medical software has to be assigned to one of three safety classes as presented in Table 3 below [295], [306].

Table 3 Safety Classes for Medical Software [295]

Safety Class	
A	No injury or damage to health is possible
B	No severe injury is possible.
C	Death or severe injury is possible.

Clinical Evaluation

To demonstrate the compliance with the essential requirements, a clinical evaluation is conducted. The clinical evaluation assesses the clinical data of the medical device to verify the clinical safety and performance of the medical device [307].

The clinical evaluation requires the manufacturer to conduct the following steps:

1. Identification of the essential requirements, which requires support from relevant clinical data
2. Identification of available clinical data, which are relevant to the medical device and its intended use

3. Evaluation of the data in terms of its suitability for establishing the safety and performance of the medical device
4. Generation of any clinical data, which are needed to address outstanding issues
5. Bringing all clinical data together to reach conclusion about the clinical safety and performance of the device [307].

1.4.1.1.3. Step 3: Classify Device and Choose Conformity Assessment Procedure

After the medical device is assigned to a directive, it must be classified into a risk class and a conformity assessment procedure must be identified. Each directive provides different risk classes with different conformity assessment procedures.

Directive 93/42/EWG - Medical Device:
The Directive 93/42/EWG for Medical Devices [296] provides six different risk classes as shown in Table 4 below. The medical device must be classified in one of these classes appropriate to its risk level. The classification rules can be found in Annex IX Classification Criteria of this directive.

Table 4 Risk Classes of MDD [296]

Risk class of MDD	Example
I	crutches, wheelchair
Is (sterile)	disposable products (e.g. injection needle), sterile dressing material
Im (measurements)	thermometers, manual sphygmomanometer
IIa	contact lenses, hearing aid devices
IIb	ventilator, dialysis machine, anaesthesia machine
III	artificial hip, cardiac catheter

After the medical device is classified into a risk class, a conformity assessment procedure must be chosen. MDD describes possible procedures for medical devices of risk class I, Is, and Im. (See Figure 60 below).

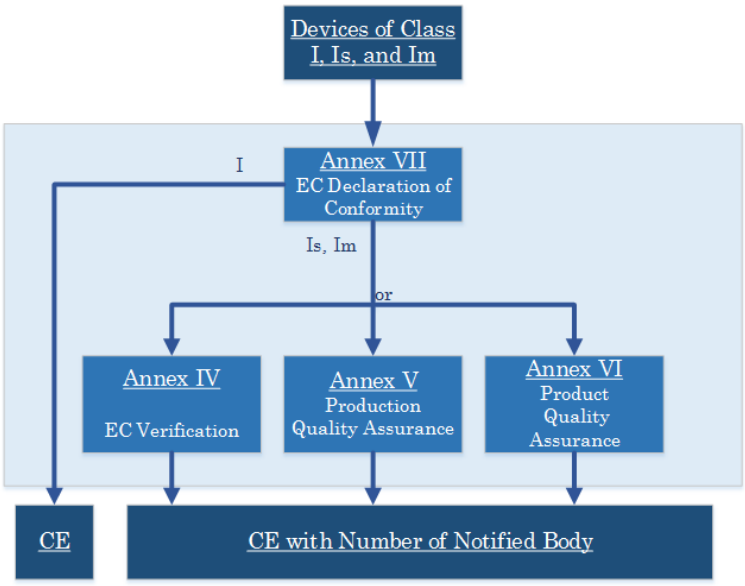

Figure 60 Conformity Assessment Procedure for MDD Devices (class I, Is, Im) [295]

Figure 61 belowFigure 60 describes possible conformity assessment procedures for medical devices of risk class II and III.

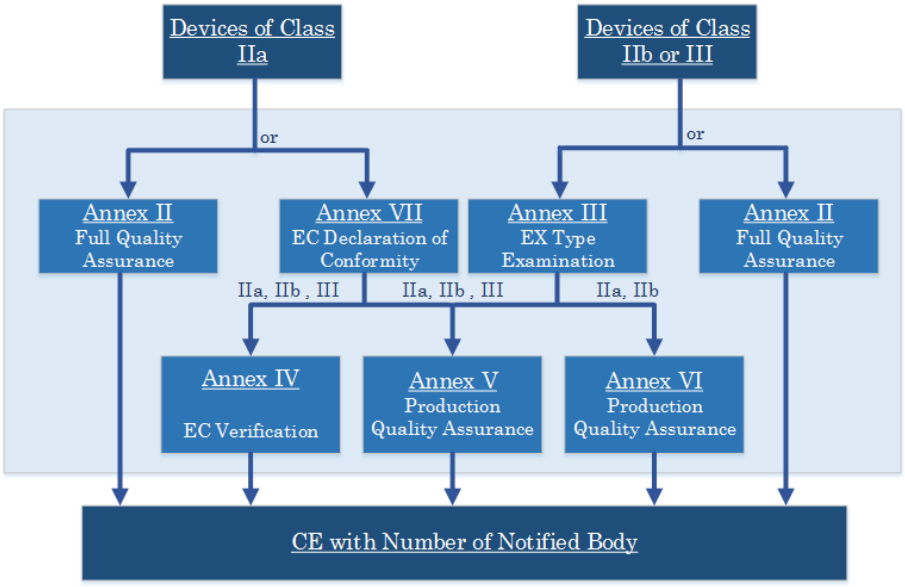

Figure 61 Conformity Assessment Procedure for MDD Devices (class IIa, IIb, III) *[295]*

Depending on the chosen conformity assessment procedure, specific procedures, described in the annexes of the directive, have to be followed [308]:

Directive 98/79/EG - In-Vitro Diagnostic Devices:
The Directive 98/79/EG for In-Vitro Diagnostic Devices [302] provides four different risk classes as shown in Table 5 below. The medical device has to be classified in one of these classes appropriate to its risk level and usage. The classification rules can be found in Annex II of this directive.

Table 5 Risk Classes of IVD Devices *[302]*

Risk Class of IVD	**Examples**
List A	reagents and reagent products for: • determining blood groups • the detection of infections with HIV
List B	reagents and reagent products for: • the detection of infections with cytomegalovirus or chlamydia • self-diagnosis devices for the measurement of blood sugar
Devices for self-testing	pregnancy test
Other IVD products	all products that are not listed in List A and B and are not listed in the List for self-testing, e.g. clinical chemistry tests

Appropriate to the risk class, a certification procedure must be determined. Figure 62 below shows possible certification procedures for all risk classes.

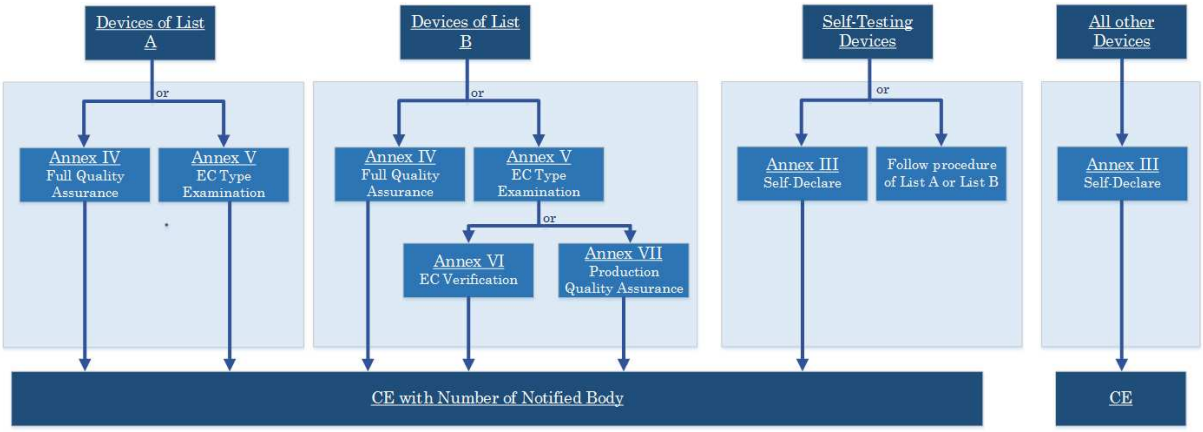

Figure 62 Conformity Assessment Procedure for IVD Devices [302]

Depending on the chosen conformity assessment procedure, specific procedures, described in the annexes of the directive, have to be followed.

Directive 90/385/EWG - Active implantable medical devices
Since active implantable medical devices are subjected to the highest risk class, they have to run a similar procedure as MDD-devices of class III. Additionally, the same procedure has to be applied for equipment of active implantable medical devices. [303]

1.4.1.1.4. Step 4: Prepare CE Marking Technical File / Design Dossier

All medical devices of class I, IIa, and IIb have to create a Technical File. The Technical File is a set of documents with evidence that the medical product complies with the requirements of the EU-directive. Medical Devices of class III have to prepare a Design Dossier. The appropriate directive decides which documents have to be included in the Technical File. The following documents are the most essential ones [299]:

- General Information like Product Description / EC Authorized Representative
- Classification Determination
- Essential Requirements
- Risk Analysis
- Labelling
- Product Specifications
- Clinical Evaluation
- System Test Reports
- Manufacturer's Declaration of Conformity

1.4.1.1.5. Step 5: Appoint a European Authorized Representative

If the company has no physical location within the EU, the CE Marking Directive requires the company to appoint a European authorized representative. The representative serves as a primary contact for all EU participants. Furthermore, the representative is responsible:

- to make a current copy of the Technical Files/Design Dossier to be available for inspections by Competent Authorities
- to assist with Incident and Field Safety Corrective Action (FSCA) reporting
- to assist with device registrations
- to authorize the manufacturer to place the representative name and address on the medical device labels, packaging, and user instructions [300].

1.4.1.1.6. Step 6: Have Notified Body Audit

The technical file / design dossier is controlled by the notified body. The notified body audits the quality system and reviews the medical device if it complies with the EU-directive. If the audit and the review were successfully, the notified body issues a CE certification [309].

1.4.1.1.7. Step 7: Register Device with Competent Authorities

When the medical device shall be placed on the market for the first time, a competent authority has to be informed. The medical device must be register, if the device is a class I device. Devices of class IIa, IIb, III or active implantable medical devices need not to be registered in most EU-member states, since they are already subjected to stricter conformity assessments [301].

1.4.1.1.8. Step 8: Prepare EC Declaration of Conformity

The manufacturer has to prepare a Declaration of Conformity, which states that their medical device complies with all requirements of the appropriate directive [301].

1.4.1.1.9. Step 9: Affix CE Marking to the Device

After the Declaration of Conformity, the manufacturer can affix the CE marking on the medical device [301].

1.4.1.2. Germany MPG

In Germany, every medical product and medical equipment is subjected to “German Act on Medical Devices (Medizinproduktgesetzt - MPG). The MPG adopts the EU-directives and adds national-specific regulations and requirements. The purpose of the MPG is to care about the safety, aptitude, and performance of medical devices. Further, it takes care about the protection of patients, users and others. Additionally, there are some regulations, which belongs to the MPG:

- Medical Devices Ordinance (MPV / Medizinprodukte-Verordnung)
- Ordinance on Clinical Investigations with Medical Devices (MPKPV / Verordnung über klinische Prüfungen von Medizinprodukten)
- German Institute of Medical Documentation and Information Regulation (DIMDIV / DIMDI - Verordnung)
- Medical Devices Operator Ordinance (MPBetreibV / Medizinprodukte - Betreiberverordnung)
- Medical Devices Fees Regulation (BKostV-MPG / Medizinprodukte - Gebührenverordnung)
- Safety Plan Ordinance (MPSV / Medizinprodukte-Sicherheitsplanverordnung)
- Medical Devices Implementation Regulation (MPGVwV / Medizinprodukte-Durchführungsvorschrift)
- Medical Devices Prescription Regulation (MPAV / Medizinprodukte - Abgabeverordnung) [310].

1.4.1.3. USA - FDA Regulatory Process for Medical Devices

In the US, The Food and Drug Administration (FDA) is mainly responsible for the regulation of medical devices. At the beginning, it has to be considered if the product is a medical device (according to US CODE Title 21 Food and Drug s, Subchapter II – Definitions, Section 321). Figure 63 below shows the general steps for medical devices to obtain the market approval.

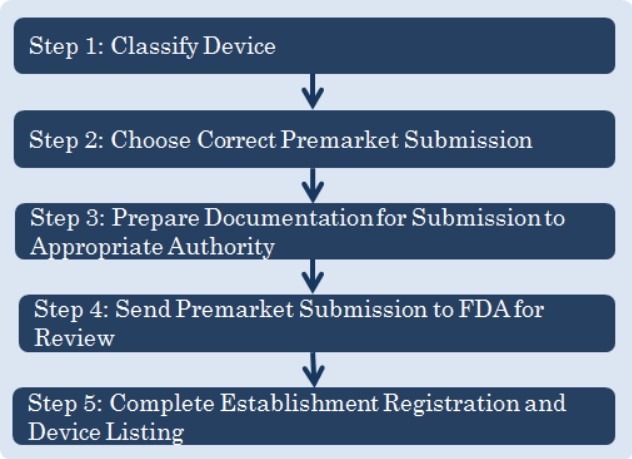

Figure 63 General Steps to Mark Medical Devices on US Market [311]

1.4.1.3.1. Step 1: Classify Device

The first step is to classify the medical device into one of three risk level classes. The higher the risk level, the more regulatory control is required. Table 6 shows the three risk level classes with some examples [311].

Table 6 Risk Level Classes [312]

Risk Level Class	Example
I (low risk)	elastic bandages, examination gloves, mechanical wheelchair
II (moderate risk)	infusion pump, bone fixation screw, blood pressure kit
III (high risk)	pacemakers, dental lasers, heart valves

The FDA provides a Product Classification database, which contains exemplary many medical devices with their appropriate risk class to support the manufacturer to classify their medical device.

1.4.1.3.2. Step 2: Choose correct Premarket Submission

The second step is to choose the correct premarket submission (explained below). In most cases a 510(k) or a Premarket Approval Application (PMA) submission is required. There are some exemptions, stated in the CFR, which are, in most cases, class I devices. Those exemptions do not require one of them. Medical device of Class I and II usually requires 510(k) submission and class III devices usually requires PMA.

510(k)

In 1998, the FDA developed a guidance documentation, called “The new 510(k) Paradigm”, to increase the efficiency of the Premarket Notification (PMN) Review process. There are some preconditions, which have to be fulfilled to apply this procedure:

- the medical product must be of class I or class II
- to determine substantially equivalent (SE), which means that the new medical device is at least as safe and effective as an existing legally marketed medical device, which is equivalent to the new medical device
- a clinical study is not necessarily required

Additionally, to the traditional 510(k) submission, two alternatives methods exist, the “Special 510(k)” and the “Abbreviated 510(k)”. Another submission method is the “de novo” submission method.

- *Traditional 510(k):* may be used for any type of 510(k) submissions.
- *Special 510(k):* is useable for device modification of the company’s own device, which have already been cleared under the traditional 510(k) process.
- *Abbreviated 510(k)*: can be used for medical products when a guidance document exists, a special control has been established, or a relevant consensus standard has been recognized by the FDA.
- *De novo:* applicable to new medical devices with low and moderate risk level (class I and class II). It is also used when the 510(k) is deemed as a non-substantial equivalence (NSE). In addition, one of the 510(k) submissions must be conducted. [311]

Figure 64 below illustrates “The New 510(k) Paradigm” and the decision procedure for the appropriate methods.

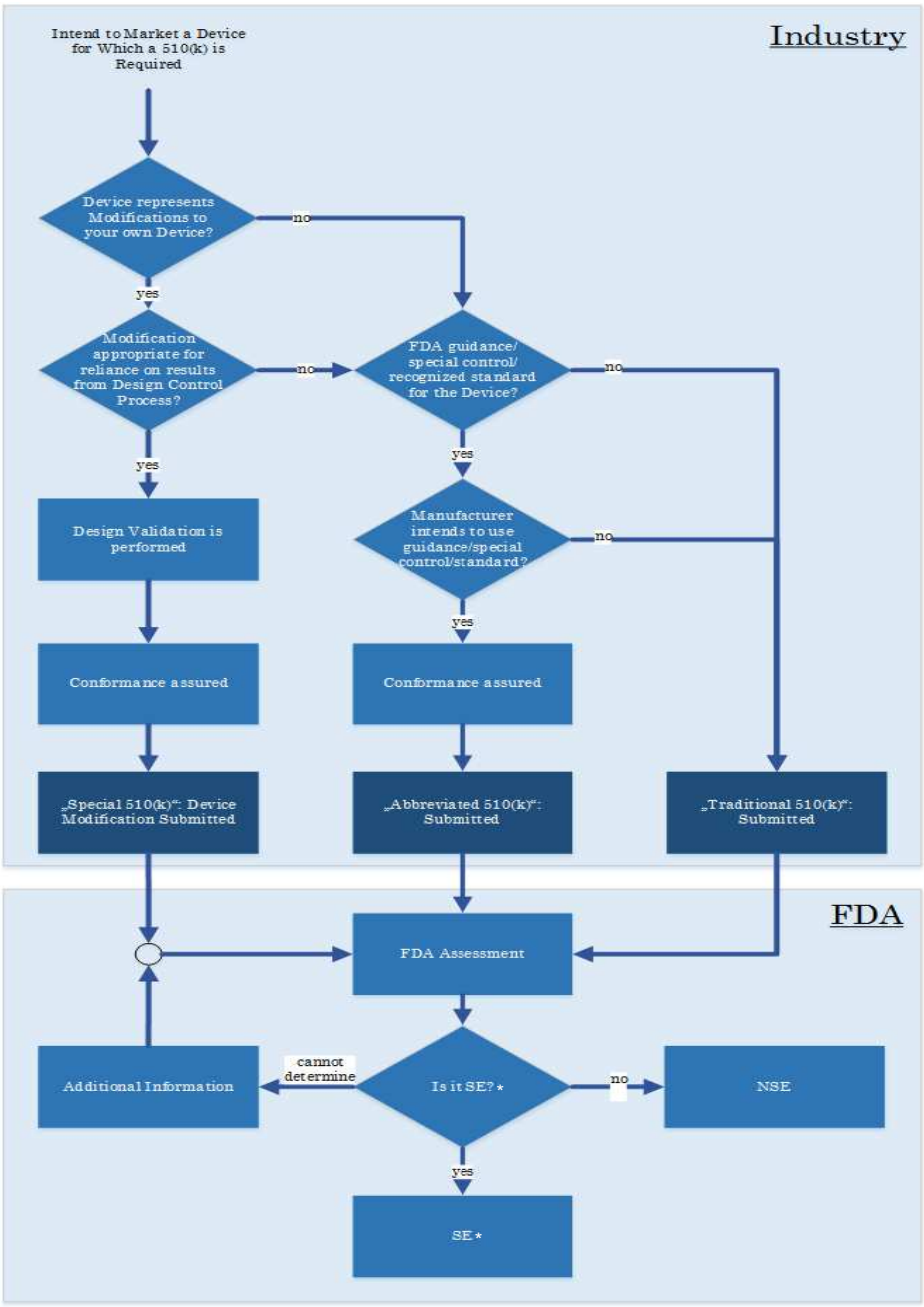

Figure 64 The New 510(k) Paradigm [311]

The Premarket approval (PMA)

PMA is the FDA process of scientific and regulatory review to evaluate the safety and effectiveness of Class III medical devices and of medical devices, which do not fall in 510(k)-procedures. There are different types of PMA applications to increase the flexibility and for more stringent controls.

- *Traditional PMA:* This is the original complete method. All supporting documents, clinical studies, and information have to be submitted to the FDA. This method, for example, is applied when the medical device has already gained approval in another country with established regulations.

- *Modular PMA:* This method allows the manufacturer to submit the application in modules to the FDA. Normally, a complete PMA will be compiled later. This method is especially suitable for devices, which are in an early stage of clinical study.
- *Streamlined PMA:* This method is for medical devices that are well known by the FDA. Because of previous knowledge, previous dealings with similar devices, or existing FDA guidance documents, the PMA submitted will be reviewed in a streamlined and more efficient process [311].

1.4.1.3.3. Step 3: Prepare Documents for Submission to appropriate Authority

After selecting the correct premarket submission type, appropriate information is needed. The FDA provides several types of resource, for example web-based regulatory assistance or the CDRH Pre-Submission Program, to support the premarket submission procedure. When the applicants prepare the premarket submission, they have to consider different information like: design control, nonclinical testing, clinical evidence and labelling [313]

1.4.1.3.4. Step 4: Send Premarket Submission for Review

After completing the premarket submission, it is sent to the FDA. During the review by the FDA, the applicant is interacting with the FDA. [313]

1.4.1.3.5. Step 5: Complete Establishment Registration and Device Listing

The device facility must register its establishment and list their device with the FDA. For devices which are not exempt, the device facility must wait until it receives the FDA clearance or approval. [313]

1.4.2. Service Level Agreements

Service Level Agreements (SLAs) in the cloud are a kind of contract to ensure that the cloud meets the requirements of the enterprise/client. An SLA sets expectations for both cloud providers and consumers by establishing a set of ground rules to deal with challenges, constraints, and changes that can come from networks, security, storage, processing power, database/software availability or even legislation or regulatory changes.

SLA is used to formally define the minimum quality of service required by a customer (i.e. hospitals) throughout the term of the agreement and includes the SA Performance Framework for the delivery of services within agreed Key Performance Indicators (KPIs). SLAs function as a:

- Communication tool: Information and communication channels must be defined and improved to ensure that consumer's staff has access to the information they need to carry out the necessary controls
- Support tool: The rights and duties of parties must be clarified, to protect them, as well as to make the system function and help to avoid or alleviate disputes.
- Measuring tool: SLAs ensure that both parties use the same criteria to assess the service safety and service quality.

The main criteria generally included in SLAs include availability, performance, disaster recovery, change management, problem resolution, dispute mediation, exit strategies, in addition to data security and privacy, location, access, and portability. The SLA may specify thresholds for criteria and the financial penalties associated with the violation of the defined thresholds.

In Big Data applications, SLAs for performance, availability, security, governance, compliance, monitoring, auditing, etc., depend on each application specifications and the priorities of what is considered critical for it. In the Healthcare domain, cloud service providers are required to meet

stringent SLAs, clarify their liabilities and risks, and provide means for auditing and reporting. Here again, specific SLA criteria are modelled according to their criticality for a given Healthcare application.

Up to date there are no relevant standards or guidelines for SLA for healthcare domain as one can find, for example, in Cloud computing area in form a Web Service Level Agreement for web service monitoring. Some theoretical research work has addressed this issue. One example is (a) [314] that present a Healthcare Trusted Cloud Computing (HTCC) framework with a customized SLA considering healthcare requirements. Another example is (b) [315] that deal with medical image processing in the cloud by adopting a genetic algorithm-based approach for data and application (virtual machine) placement to avoid SLA violations. There is no public implementation available for these two models.

Moreover, two examples of SLAs related to healthcare products or services are discussed below.

As the first example we can use an SLA, developed by ZH Healthcare, which is a US based software publisher of software modules for serving the Revenue Cycle Management process of medical providers. It developed the Blue EHS product, which is a Health IT as a Service and a SaaS platform. Its web portal empowers healthcare service providers and innovators to quickly build and deploy their heath IT solutions using a set of tools and modules available on the cloud with the minimum effort. The Blue EHS SLA is provided, which mentions different metrics and aspects of service availability between the Blue EHS as a service provider and clinic as a service consumer [316]. It stipulates, for example, that during its term the Blue EHS Covered Services portal will be accessible and operational to customers at least 99% of time after excluding scheduled downtime. However, If ZH Healthcare does not ensure terms mention in the Blue EHS SLA, and if the client meets its obligations, client is eligible to receive the Service Credits described in Table 7 below.

Table 7 Service Credits according to the BlueEHS SLA

Monthly Uptime Percentage	Days of Service added to the end of the Service term, at no charge
< 99.9% - ≥ 99.0%	3
<99.0% - ≥ 95.0%	7
<95.0%	15

Another example can be found in the practice of Central Adelaide Local Health Network, which is a part of the South Australian public health system and frequently use SLAs. These agreements often include the terms related to the management of available services, amendment windows, resolution process, key performance indicators and targets and SA Health service priorities [317] [318].

Medolution is aimed to be a cloud-based solution, offered as a service. The fact that it deals with IoT, Big Data, and dependability of systems and sub-systems contributes to the complexity of defining an SLA for this solution. In particular, there is a need to match functional and non-functional SLA requirements in Medolution to the services provisioned by underlying clouds with different service prices and quality.

In this respect, Medolution use cases will create an opportunity to define different types and scopes of service to be offered to its customers. In fact, for each use case, SLA terms would be determined throughout the project as it progresses.

1.4.3. Other medical and security considerations

Monitoring and traceability
In medicine, monitoring could be defined as the examination of a disease, condition or one or several medical parameters over time. It can be performed by using medical devices (for example, by continuously measuring vital signs), and/or by periodically performing medical tests (such as blood glucose monitoring with a glucose meter in people with diabetes mellitus). Transmitting data from a monitor to a distant monitoring server is known as telemetry or biotelemetry. Use of in-house sensor and mobile monitoring systems has proven to enhance diagnostics and treatment and also reduce the cost of associated medical care [319] [320], therefore the use of various medical monitoring sensors and devices is planned for Medolution use cases.

Traceability could be defined as the ability of verifying the history, location and status of an item using tools or ways for recorded identification. While developing medical software devices or applications, safety and effective traceability method should be applied. This process is difficult and very complex due to disastrous results in case of defective medical device software. It could lead to death or serious injury. Nowadays, traceability is very important while architecting medical device software. Some traceability assessment and improvement methods have been developed, including so called Med-Trace method [321].

Latency considerations
Latency refers to the time interval between the entry of a signal into a system until its departure or the system reaction. Especially in tele-medicine applications the knowledge about latencies is of paramount important when real-time requirements are to be fulfilled. Thus the availability of latency data is a key performance indicator for real-time and near real-time processing. Details about the regulations can be found in [322].

1.5. Conclusions

The shift created by the transition from electronic records, revolutionary development of mobile and wearable technologies opening new opportunities for sharing information among healthcare providers, between physicians and patients, and with third parties is happening quickly, in many ways too quickly for either physicians or the laws and regulations pertaining to medical records and data and medical devices to keep up. These are essential requirements to be considered by Medolution however. Consequently, relevant regulatory data privacy and security constraints analysed in details in this Appendix constitute an important framework for the standards and solutions surveyed in the main document, in particular Chapter 8, their planned application and enhancement throughout the project.

Appendix B: Relevant European research projects in healthcare data processing

This section provides an overview of selected prior and current collaborative research projects that focus on various technologies and applications that are relevant to Medolution and, along with the conceptual and industrial domains discussed such far, constitute an important part of the state of the art to be addressed.

1. MEDUSA Project

The Medusa project is an ITEA3 project that started in January 2013 and ended in December 2015. It involved 12 partners from the Netherlands and France and in some way is a predecessor to Medolution project. MEDUSA's purpose was to enhance quality of diagnosis and decision making in acute and/or critical situations in a patient's condition by introducing a new service concept in healthcare based on three pillars:

- Advanced imaging as a service, by which MEDUSA: 1) Ensures the timely and efficient exchange of medical information with an aim to satisfy the high levels of safety, security and privacy required for a secure environment 2) Supports and encourages the appropriate use of both remote and local advanced medical image processing in real-time, such that the medical professional obtains the required detail of information 3) Supports transparency of distribution with respect to geographical location, both fixed and mobile
- Secure virtual workspaces as a service, by which MEDUSA: 1) Supports a collaborative workspace in a professional medical situation, where knowledge, capabilities and patient data are brought together and made available for clinical processing, respecting medical regulations and clinical procedures and protocols. This is an area where traceability, latency and interoperability are crucial issues.
- Medical diagnosis support as a service, by which MEDUSA: Enables physicians to virtually group around a patient for diagnosis and treatment decision making.

The added value of the virtual collaborative environment, that allowed the described above new service concept, can be illustrated through three use cases.

Acute trauma care
Time and the immediate availability of specialized expertise are crucial for this use case. For an optimal treatment of trauma patients, it is very important to get the right patient in the right time to the right hospital. To reduce over triage and under triage and improve the quality and speed of diagnosis, instant collaboration is needed between ambulance personnel, physicians at nearby hospitals, and experts in specialized trauma centres.

Acute ischemic stroke
In this case the rapidly developing loss of brain function is caused by a blockage of a cerebral artery. Here, consultation among a team of experts is made possible by MEDUSA through the use of remote collaboration technology. Furthermore, advanced image processing, which may not be available in workstations at the treating hospitals, is required to determine the optimal treatment.

Cancer treatment
In this use case physicians collaborate in order to determine the patient diagnosis and plan the best treatment. One crucial step in radiation oncology treatment is the delineation of the areas to be irradiated (tumours) and to preserve organs at risk. This step involves the inputs of radiotherapists and nuclear medicine physicians, respectively specialists of anatomical and functional series, who

possibly practice in different health centres. Thanks to MEDUSA, the experts can access remotely the contouring application, manage information from multi-modality imaging and collaborate in real-time on the patient case.

Additional use cases on virtual microscopy, post-traumatic coma assistance and cross-disciplinary medical meetings were also considered in the project.
(See Figure *65* below).

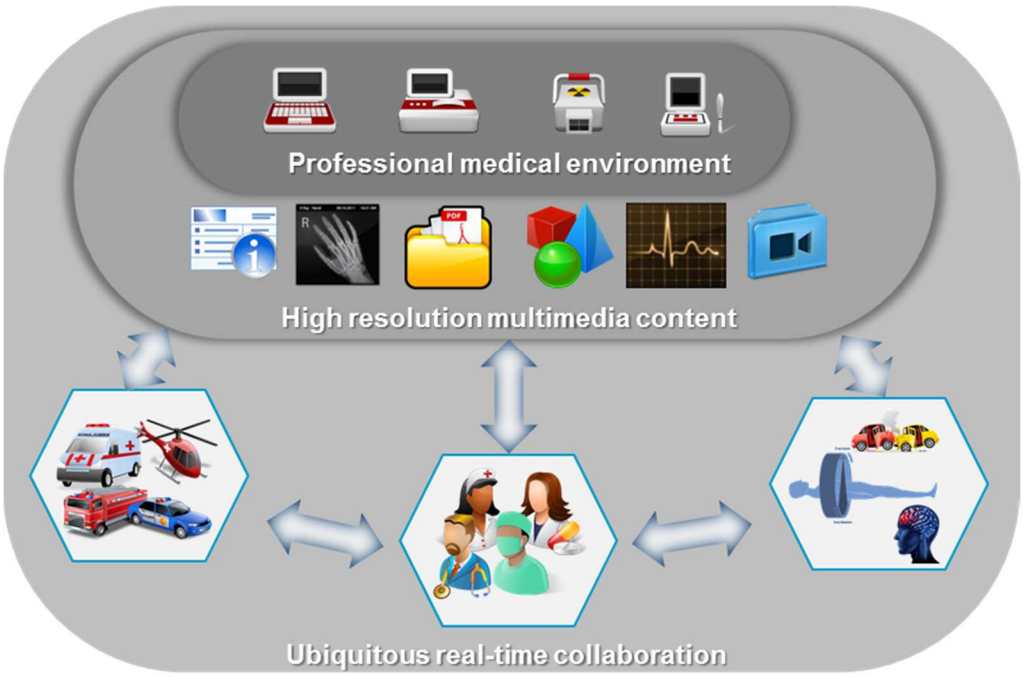

Figure 65 *MEDUSA virtual collaborative environment: an integrated and interoperable set of tools to unlock medical information and functionalities [323].*

The outcomes of Medusa project can be presented as a set of interoperable tools to unlock medical information functionalities:

Advanced medical image processing as a service
In 5 applicative scenarios, high-throughput processing tools have been developed and evaluated. All of these solutions share the common Cloud-based framework. Furthermore, MEDUSA has compression-based techniques for optimized image transfer and image processing, in addition to advanced processing capabilities to extract required information from the images for optimization of treatment and clinical decision support. Examples include intracranial haemorrhage quantification, infarct volume detection, and tumour contouring for oncology. Furthermore, remote and high-performance analysis of in-vitro images has been implemented and validated. NICo-Lab, an AMC spin-off for automated Neurovascular image analysis supporting Multi-Center Trials, has been created so as to leverage the market access for the MEDUSA cutting-edge research results. (See Figure 66 below).

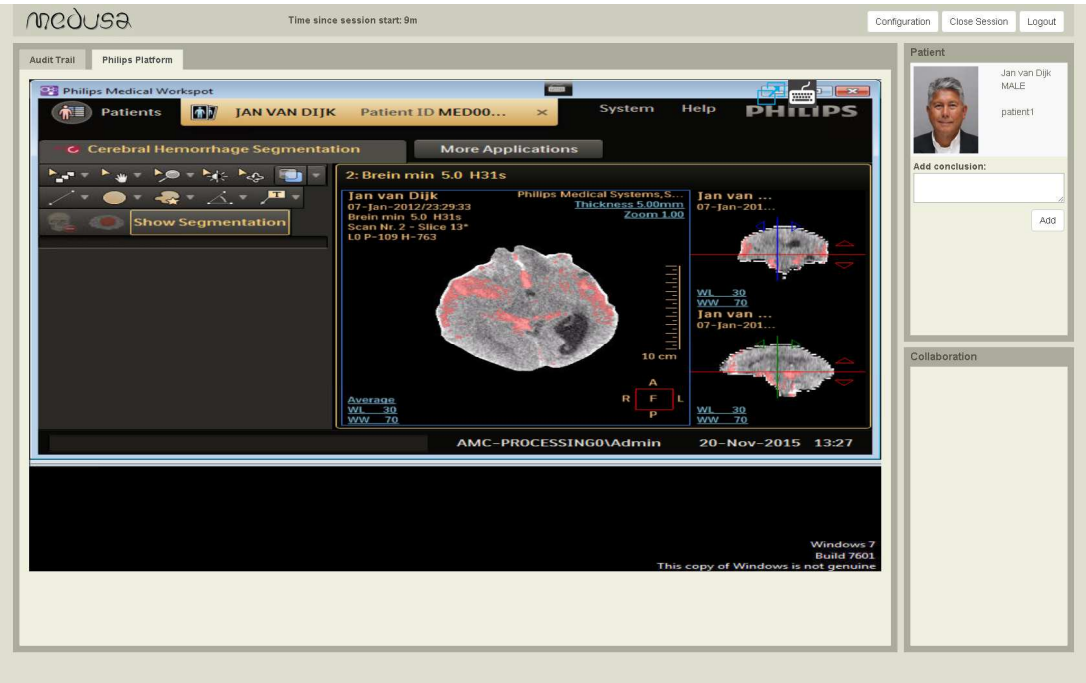

Figure 66 Advanced medical imaging algorithms running inside the MEDUSA framework [324].

Medical system end-to-end protection and defence

An overall security architecture meeting privacy requirements in virtual medical collaborative workspaces has been defined. It includes resources ensuring authentication, confidentiality, integrity, availability and content tracking. The main innovations are provided by (1) a Central IAM (Identity & Access Management) and firewall, (2) a multi-level encryption strategy, (3) traceability and integrity proof of technologies to medical uses.

Cloud-based virtual collaborative framework

Thanks to its architecture, infrastructure, support tools and functionalities, the MEDUSA virtual collaborative framework represented a true prime. First, infrastructure components were deployed over high performance physical resources (computing, storage, etc.), thus providing the infrastructure as service (IaaS) relying on an open-source Cloud management system. This high performance real-time Cloud allowed for deployment and execution of application components with optimal performances, meeting the MEDUSA use cases and requirements. This infrastructure also benefited from meta-Cloud deployment and management. Resources were allocated on this infrastructure through a Cloud management platform, which handles their deployment, lifecycle management, monitoring, orchestration, and access in a dynamic, provider-independent way. Finally, the collaboration functionality was offered, for the first time, as a service: it was no longer pre-programmed with the application but could be dynamically updated so as to take into consideration the actual working conditions, professional access rights, user privacy, etc. Both legacy (mono-platform) and Cloud-based applications can be deployed in the MEDUSA virtual collaborative space, thanks to optimized, cross-standard virtualization solutions. (See Figure 67 below).

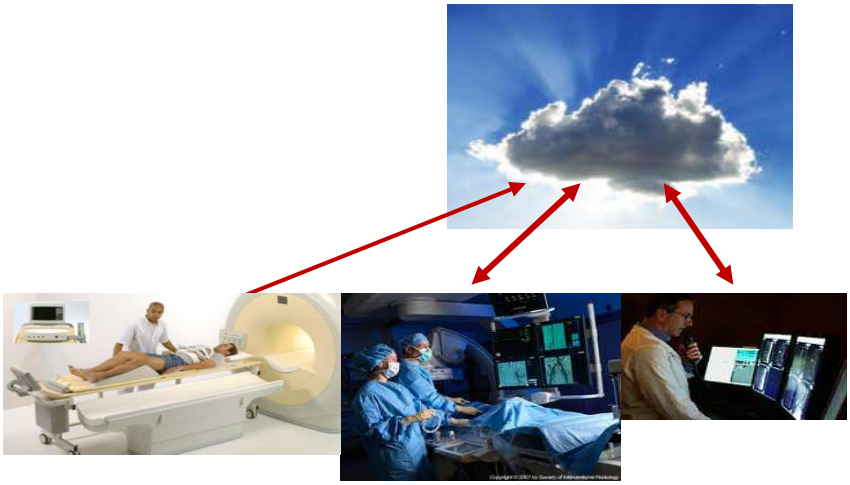

Figure 67 MEDUSA virtual collaborative framework [323].

Real-time decision support with sensors
The decision support functionality provided the doctors in a collaborative session with alerts about dangerous trends in a patient's condition. These trends were captured by sensors on the patient that were connected to the Cloud through mobile applications to be interpreted in real-time by rules as defined in approved medical protocols.

Relevance to Medolution
This brief presentation of the MEDUSA project brings to light at least four PoC (proof of concepts) that can be reconsidered and enlarged in order to support the Medolution development:

- the PoC for bringing together, in a common collaborative working environment, applications of various types (native OS, business culture, execution constraints etc.), with no initial collaborative resources; under the Medolution framework, this collaborative working environment is expected to be extended so as to accommodate applications intimately related to the IoT (processing, control, etc.) and Big Data processing;
- the PoC for flexible, dynamic, on-demand Cloud deployment according to some pre-established business or technical constraints; in Medolution, this dynamic Cloud management is expected to be adapted to the peculiarities of the IoT and Big Data as well as to the national rules and regulations related to location of the medical data processing and storage;
- the PoC for integrating live generated data from medical sensors into medical DSS (decision support systems); this will be core issue in Medolution, where the complexity and heterogeneity of the sensors combined to the amount of data they create is a particular constraint;
- the PoC for preserving a prescribed level of security (similar to the one provided in non-Cloud infrastructure) when deploying medical applications in Cloud; this generic tool box is expected to be enriched according to the Medolution use cases.

Additionally, MEDUSA resulted into a set of software tools for medical content traceability based on novel methodological frameworks, in particular those for watermarking and fingerprinting. The software is available as stand-alone components, developed in C and integrated with the reference software for basic image/video file formats; both robust and semi-fragile functionalities are available. These tools will be used and further advanced in the Medolution project.

2. SALUS Project

The SALUS project [325] (Scalable, Standard based Interoperability Framework for Sustainable Proactive Post Market Safety Studies) is an R&D project co-financed by the European Commission's 7th Framework Programme (FP7), that stated in February 2012 and ended by April 2015. It was coordinated by SRDC, and involved 10 partners from 8 countries, including Germany, France, Sweden, Switzerland, Italy, Netherlands, Belgium and Turkey:

SALUS has explored new ways of accessing and analysing data found in electronic health records to provide an infrastructure that will enable execution of safety studies for mining and analysing real-time patient data. The aim is to ensure safety through early detection of rare adverse events; to provide the pharmaceutical industry faster medication innovation by decreasing time to market for new, safe and effective drugs; and to reduce the load of overwhelmed medical practitioners at the same time.

SALUS project's main objective was to provide a comprehensive solution supported with ready-to-use tools in order to enable the secondary use of the already available Electronic Health Record (EHR) data in patient care domain, for clinical research purposes. SALUS particularly aims to strengthen the spontaneous reporting process by automated adverse drug event (ADE) detection on disparate EHR systems; enable standards based ADE reporting by automatically extracting the available information from the EHRs; realize the execution of post-market analysis and effectiveness queries for different subpopulations selected from multiple, distributed EHRs as target cohorts; contribute to the signal detection processes; and facilitate wide scale outcome and effectiveness research by enabling to observe selected cohorts of patients over an extended period of time screening multiple, distributed, heterogeneous EHR systems to identify long term safety issues of a product.

Additionally, SALUS provided a mechanism for anonymizing data by requiring stakeholders define and analyse the identifying data elements on a per-case basis and assign each identifying data element a de-identification algorithm. The identifying data elements are then anonymized according to these algorithms to produce the de-identified data set. Each identifying data element is de-identified independently of the other identifying data elements. To aid in the decision making process about which data elements are identifying and the de-identification algorithm that should be used, SALUS requires the Data Protection Offices of the EHR to conduct a risk analysis of their data using a k-anonymity approach. SALUS provides tools and guidance to assist in this effort (section "3.7 Safety Analysis Tools" in SALUS Deliverable D3.4.1 Conceptual Design of the SALUS Architecture and section "5.4 K-anonymity Approach" of SALUS Deliverable D5.4.1 Interoperability Profiles and Open Source Toolsets for Security and Privacy).

SALUS provided a comprehensive solution supported with ready-to-use tools in order to enable the secondary use of the already available Electronic Health Record (EHR) data in patient care domain, for clinical research purposes. Using the SALUS system, it is now possible to:

- detect Adverse Drug Events on patient summaries using predefined detection rules
- prefill Individual Case Safety Reports (ICSRs) with available patient data
- submit ICH E2B based case safety reports to regulatory bodies
- perform effectiveness studies on available EHR data by submitting various queries for different purposes such as case series characterization and temporal association screening
- perform post market drug surveillance on selected cohorts coming from different EHR sources and make analytical calculations

To achieve this, SALUS Project has developed comprehensive semantic and technical interoperability framework to seamlessly access heterogeneous EHRs [326]. It provides a system which enables the available Electronic Health Record (EHR) systems to serve their data anonymously and securely for the sake of clinical research purposes.

SALUS developed end-user tools with the goal of enabling post market safety studies on top of the already available EHR data of different systems. SALUS toolset included the Safety Analysis Tools [327] (Case Series Characterization Tool (CSCT), Patient History Tool (PHT), Temporal Association Screening and Temporal Pattern Characterization Tool (TAS/TPC Tool)), and Post Marketing Safety Study Tool (PMSST) [328]. The common purpose of these tools are to provide web based, easy-to-use interfaces for the clinical researchers so that they can monitor the available EHR data through different statistical methodologies. In general, they send their queries to the Care Zone and they receive statistical results or limited patient data in secure and de-identified fashion. A number of statistical methods used by the Research Zone are executed on an Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) [329] based clinical data repository. This database is populated by SALUS routines and kept up to date according to the original data sources by the Subscription Mechanism of SALUS. The initial setup moves all available data from the EHR data sources to these repositories through the Semantic Interoperability Layer. Afterwards, a scalable subscription mechanism starts running which monitors any updates in the EHR sources and updates the OMOP database accordingly. Apart from the tools developed for the clinical researchers, SALUS implements two dedicated end-user tools for clinicians. These are the ADE Notification Tool [330] and the ICSR Reporting Tool.

SALUS followed the international standards for data interoperability and for that reason the Technical Interoperability Layer extended two already available IHE profiles, namely Query for Existing Data (QED) [331] and Care Management (CM) [332] with the HL7 Health Quality Measures Format (HQMF) [333] constructs. As a result, SALUS opens up a standard web service interface on top of the existing data warehouses.

Lastly, any kind of information exchange between the Care Zone and the Research Zone is being audited according to the IHE Audit Trail and Node Authentication (ATNA) [334] profile guidelines. SALUS uses OpenATNA [335] which is an open-source implementation of the IHE ATNA profile and maintains an audit record repository both in Research Zone and Care Zone separately.
A full description of SALUS components can be found in SALUS Final Report [121].

Relevance to Medolution
In the SALUS Project, data analytics methods have been built for pharmacovigilance to detect and quantify adverse drug event signals by examining vast number of EHRs. During the deployment phase of SALUS, several performance problems have been encountered while processing vast amounts of EHRs for pharmacovigilance data analytic operations that have been addressed through multiple threads processing EHRs in parallel. Based on this experience, Medolution will set-up a scalable cloud infrastructure that can cope with large computing loads on demand.

In addition to this, in SALUS project several technical and semantic interoperability solutions have been built to consume population based Electronic Health Records based on international standards like IHE Query for Existing Data (QED) and Care Management (CM) profiles. These results can be utilized in Health Data Ingestion stack layer of Medolution for addressing the interoperability issues in communication with EHR systems.
As for the SALUS approach for data anonymization, it requires subject matter experts to decide which anonymization technique should be applied to each data attribute in the schema without considering the context of the data queries which might be coming. This static mechanism potentially results in data being anonymized more stringently than may be necessary depending on the query and the data set being returned. This may hamper the possible analytics, which could be possible on the data sets. Medolution seeks a more dynamic approach, which considers each data attribute within the context of the data query that is being performed and selects an anonymization mechanism which ensures that the resulting data set satisfies the re-identification risk threshold set by the data's custodians.

3. iCARDEA Project

iCARDEA (An Intelligent Platform for Personalized Remote Monitoring of the Cardiac Patients with Electronic Implant Devices) [336] is an R&D project co-financed by European Commission FP7 seventh framework program, which ran from 2010 to 2013. It was coordinated by SRDC and involved 8 partners from 5 countries, namely Turkey, Germany, Austria, Spain and Greece.

iCARDEA Project's main objective was to expose CIED (Cardiovascular Implantable Electronic Device) data through standard interfaces to develop an intelligent platform to semi-automate the follow-up of CIED patients with context-aware, computer interpretable clinical guideline models. The heterogeneous, disperse system components (HIS, PHR, EHR, CIEDs etc.) were made interoperable by selecting and implementing the relevant standard interfaces.

Using the platform developed by the iCARDEA Project [337], it is possible to semi-automate the follow-up of CIED patients with context-aware, adaptable computer interpretable clinical guideline models by exposing CIED data through standard interfaces. The computer interpretable guideline models were designed from re-usable building blocks to facilitate personalization of the patient care and follow-up workflow. The CIED data was exposed through standard interfaces based on the HL7, ISO/IEEE 11073 standards and the IHE IDCO Profile [338]. EHR interoperability was achieved by exposing legacy EHR systems through standard HL7 CDA interfaces so that information about patients' medical history such as the non-cardiac conditions denoting contraindications to the proposed therapies can be obtained from the patient EHR data and used in the clinical follow-up workflow. The clinical guidelines semi-automate the care process and hence support medical professionals by automatically assessing the situations of the patients. The patients were also empowered with Personal Health Records (PHR) to enable informed and responsible participation in the process and for their education. Additionally, iCARDEA platform provided comprehensive security and privacy mechanisms that are all validated in a hospital in Austria (SALK) with CIEDs from two major CIED vendors, namely St. Jude and Medtronic.
The final results are deployed at SALK Premises, serving the tele-monitoring service to the cardiac patients to manage automated remote monitoring of implantable cardioverter defibrillators (ICDs) for secondary prevention of life-threatening arrhythmias.

The following overall objectives were achieved as a result of iCARDEA R&D Project through the project lifetime:

- **Remote monitoring for implantable cardiac devices**: An intelligent platform was developed to semi-automate the follow-up of the CIED patients with adaptable computer interpretable clinical guideline models which access data in EHR data resources, CIED data and PHRs using standard interfaces. In order to create computer interpretable guideline models, firstly, clinical guideline definitions were specified with the medical partners as flowchart definitions. Then, these definitions were converted into the machine processable format by using Adaptive Care Planner component from re-usable building blocks. Then these guideline models were ready to be used as executable clinical workflows which perform the follow-up activities and semi-automate the care process and hence support medical professionals by automatically assessing the situations of the cardiac patients.
- **Integrating remote cardiac monitoring with EHR Systems:** An integration of EHR system with CIED Module to assess patient status better in remote monitoring process was conducted. The integration was provided via EHR Interoperability Framework component implementing standard profiles to retrieve patient EHR data from the hospital data sources in an interoperable way.
- **Leveraging the potential of CIEDs as widespread, ambient intelligent devices:** Due to CIEDs' limited processing capabilities restricted by their size, they need to be supported

with software running on the servers. Thus the server side processing was standalone with their custom software and proprietary interfaces. By using international standards provided by CIED Interoperability module, iCARDEA has exposed this information to be used to semi-automate the care and follow-up processes based on computer interpretable clinical guidelines. Furthermore, by using standard interfaces and interoperability utilities provided by iCARDEA CIED module, CIEDs from different manufacturers have become interoperable.

- **Addressing the under-utilization of clinical guidelines because of lack of integration into EHR Systems:** Clinical guidelines can automate the healthcare processes, which need to be achieved as routine follow-ups or remote monitoring of patients with CIEDs. Despite the potential benefits of the clinical guidelines, they have been underutilized in clinical practice due to interoperability problems of healthcare data sources to retrieve data seamlessly from the EHR data sources. The iCARDEA platform has provided EHR interoperability so that information about patients' medical history such as history of non-cardiac conditions; more detailed information about severity of each condition (e.g., record of prior hospitalizations or specifics of therapy for the condition); the medications being taken at the time of spontaneous arrhythmia occurrence or the non-cardiac conditions denoting contraindications to the proposed therapies can be obtained from the patient EHR data in an interoperable way and used in the clinical workflow. There were two major challenges to address related with EHR interoperability: the legacy EHR systems and the interoperability of the code systems used (semantic interoperability). For the EHR legacy system interoperability, iCARDEA has exposed these systems through standard interfaces using HL7 Clinical Document Architecture (CDA). To be able to map different code systems, HL7 Common Terminology Services (CTS) [339] has been developed.
- **Integration of data analysis for the quality of service in health care**: The iCARDEA Data Analysis components support healthcare professionals at hospitals. This aim was addressed by making it easier to access the patient data in a structured and harmonized way (Patient Parameter Monitor-PPM component) and by using long-time-harmonized data acquired over longer periods to generate patient-specific warnings and suggestions based on statistically valid patterns extracted using state-of-the-art data analysis techniques applied to long-time reference case knowledge bases (Data Analysis and Correlation Tool-DACT component).
- **Patient specific adaptive care:** Personal Health Record (PHR) component is built for patients to report observations of daily living, medications, life style and edit their profile. PHR also provides feedback and education to patients. Through the PHR interoperability provided with the IHE Care Management (CM) Profile, patient data could be shared between the iCARDEA components for the patient specific adaptive care in an interoperable way.
- **Validation for effectiveness, privacy, trust and security:** The iCARDEA platform was validated through a pilot validation study in Austria at SALK by demonstrating the cost and time effectiveness, clinical validity and safety with statistical analysis. Comprehensive identity management, trust and privacy mechanisms have been provided through the iCARDEA platform.

Relevance to Medolution

In iCARDEA Project several technical and semantic interoperability solutions have been built to interact with Health Information systems to collect EHRs, with CIEDs, and with Personal Health Record systems based on international standards such as HL7 CDA, IHE CM, QED and IDCO profiles. These results can be utilized in Health Data Ingestion stack layer of Medolution for

addressing the interoperability issues in communication with EHR systems, implantable cardiac devices and PHRs.

4. EASI-CLOUDS Project

EASI-CLOUDS (Extendable Architecture and Service Infrastructure for CLOUD-computing Software) is an ITEA2 project involving 6 countries (Denmark, Egypt, Finland, France, Germany and Republic of Korea) and 30 Partners. The Project started in September 2011 and ended in March 2015 [340].

EASI-CLOUDS aims at overtaking the barriers for Cloud adoption (such as vendor lock-in, quality of services guarantees, data protection and financial impact of Cloud-based business models) by proposing solutions, for a sustainable Cloud ecosystem, enabling new business models for the benefits of both Cloud consumers and providers. EASI-CLOUDS targets the creation of a comprehensive Cloud computing infrastructure featuring the three classical categories of Cloud computing offerings – Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS) – with superior reliability, elasticity, security and ease-of-use characteristics at all levels, and enabling the run of applications on top of hybrid (private/public) or multi-Cloud architectures and as well as federation of Clouds. EASI-CLOUDS also targets the creation of state-of-the-art Cloud-aware service offers in a number of domains (Healthcare, Multi-media, and Gaming), as well as the creation of a proof-of-concept marketplace with the available services and suppliers in the project.

The project selected OpenStack as the most promising solution for virtual infrastructure management and the basis of the EASI-CLOUDS Infrastructure. Another main component is the EASI-CLOUDS Platform, the system that acts as an intermediary between Consumers, SaaS Services providers, and heterogeneous IaaS providers. The EASI-CLOUDS Platform leverages the results of the French Compatible-One project (the descriptive model and execution environment based on OCCI Standard), and extends it with innovative feature such as negotiation of SLA terms between consumers and providers, automated provisioning of virtual resources and software configuration, as well as scalable monitoring mechanism to achieve elasticity and high availability, support of federation of Clouds. In addition the project did build specific Cloud services on top of a standard IaaS Cloud-stack which simplified the realization of complex and scalable services, by integration a SLA-based resource management and automatic scaling according to predefined rules [341] [342].

6 Demonstrators (Cloud based brain image processing, Photo-stitching SaaS, Evaluation of game server performance, distributed Mesh Cloud for Web Media, simple development of new services, federated placement of Cloud resources under SLA Control) were implemented leveraging the EASI-CLOUDS and illustrating target business case implementation such as (to automate elastic Software as a service delivery, to extend the resources of private Clouds with public providers, to act as a Cloud broker i.e. “one-stop shop” for several providers, or to establish a Cloud federation in which several providers agree to share workload.

The German medical scenario of EASI-CLOUDS has been the 3D-reconstruction of MRT scans of the human brain to evaluate the application for brain disease diagnostics. This reconstruction is very time consuming and, the application on a number of patients’ data has been only possible by using the flexible computing power of a Cloud. The software had to be enabled to become a Cloud-aware service achieved through creation of an integration layer and a topology in the Cloud, which allowed the use of independent worker processed to achieve dynamic scalability. The realized solution showed the complete usage lifecycle from job definition and selection of different pricing models through the billing services. Using complex event monitoring and processing and Cloud-

management the processing service could be managed to deliver the agreed service level for the neurologist in the correct time and cost frame.

Relevance to Medolution
From the technical point of view, the relevance with regard to Medolution is the lifecycle of Cloud aware services (development, deployment, operations: monitoring & adaptation, termination) with SLA and target business case constraints.

From the healthcare domain point of view, the relevance with regard to Medolution is the complete treatment of a medical use-case with all constraints and SLA information. Technically the configurable scalability, which added computing resources and automatically set up the internal network infrastructure, has enabled an automated SLA-management and SLA-based Cloud-management. The interface and events had been designed in a problem-oriented fashion, and thus interfaced with the management component and the SLA-monitoring. Another part of EASI-CLOUDS provided a Cloud orchestration of services which might be interesting Medolution for distributing services according to their criticality and data security concerns between a private hospital Cloud and external healthcare service providers in the Cloud.

As a conclusion EASI-CLOUDS provides templates and significant knowhow directly for image processing setups as Cloud services in combination with security constraints as well as a blueprint for complex Cloud services, which also concern billing, Cloud management and general SLA management in self-service Cloud applications.

5. OSAMI-Commons Project

The **OSAmI** (**O**pen **S**ource **AM**bient **I**ntelligence) was a European ITEA *2* research project. The aim of the project was the design of a basic, widely applicable SOA-oriented component platform, its development, test and its provision as open source. The project consists of a number of national subprojects, each focusing on a certain field of application. The German subproject OSAMI-D, funded by the BMBF (reference number 01 IS 08003), contributed to the e-Health domain. The main objectives were interoperability, maintainability, reliability, as well as automated configuration and management of medical devices and services to provide new forms of healthcare to diseased and convalescent people. The advantages of these technical contributions were demonstrated by means of an e-Health application, which supports ambulant cardiologic rehabilitation.

The software component platform specified by the Open Service Gateway initiative (OSGi Alliance) formed the technical basis of the OSAMI platform. It provides lifecycle management for software components as well as local service interactions as defined in service-oriented architectures and were combined with the Web Services approach, in particular Device Profile for WebServices / WebService for Devices (DPWS/WS4D) in order to implement distributed, dynamically configurable, vendor-neutral and device-independent solutions. The OSAmI e-Health system adopts a distributed architecture with integrated vital sensors (See Figure 68 below). The system transmits events through OSGi and streams of data from the home gateway to the clinic.

For distributed communication within OSAmI, services on remote frameworks were used in accordance with the OSGi Remote Service specification. The underlying communication technology of the Remote Service implementation used in OSAmI is based on Web Services, in particular, DPWS. The DPWS implementation used is Java Multi Edition DPWS Stack (JMEDS) v2. Software developers do not need to work with the JMEDS v2 API directly, because it is hidden by the Remote Service implementation. The Remote Services specification does not offer special interfaces. Remote services are used transparently, just like local services.

The data streaming component supports to transport data streams between devices and the OSAmI platform and between several instances of the OSAmI platform using the remote service specification. This component is designed as extension for the device integration component and the remote service implementation based on DPWS. This component does not provide any specific interfaces. It is relevant for OSGi services that use the features of the remote service or the device integration specification. It transparently plugs into transmission mechanisms of such services that offer Java ObjectStreams and is responsible for the transmission of these ObjectStreams. Morever, this component does not provide any specific interfaces. It is relevant for OSGi services that use the features of the remote service or the device integration specification. It transparently plugs into transmission mechanisms of such services that offer Java ObjectStreams and is responsible for the transmission of these ObjectStreams.

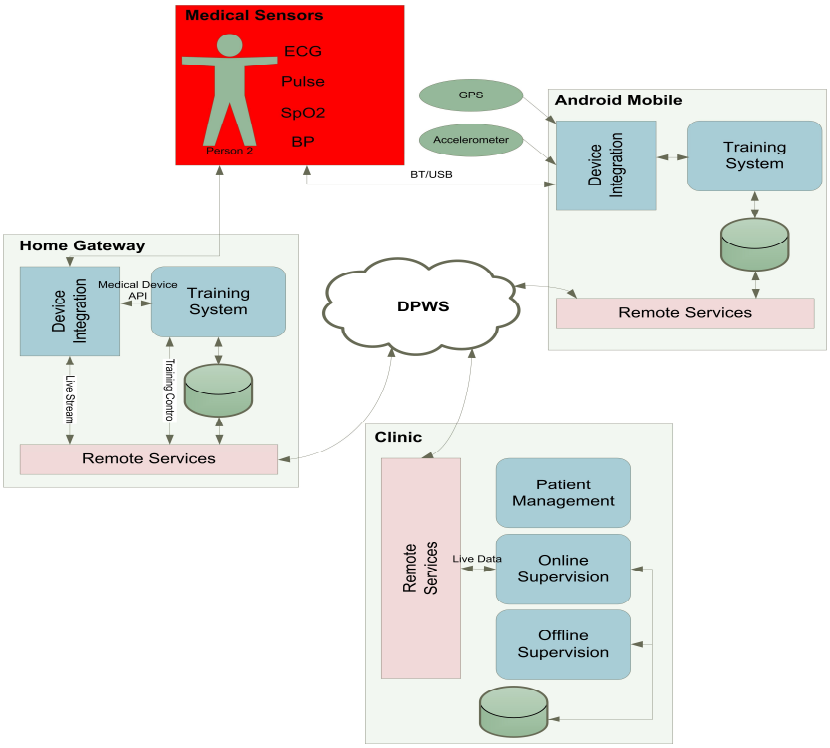

Figure 68 OSAmI Architecture [343]

Home-based medical assistance applications like the OSAmI-D system incorporates different devices like medical sensors as also needed for Medolution project. The ability of changing devices from different vendors with different communication protocols in an application without restructuring main parts of the application makes a device abstraction required. A mechanism for decoupling the application from the device driver was already implemented in the OSAmI-D project. To provide a usage of devices, independent from their protocol, an API of the data and service of medical sensors is described. This device API specifies the functionality of the medical sensors without any vendor or communication specific information. The medical API distinguishes between data and services. The data represents the information that needs to be used by devices (services). The services represent the devices which are described by interfaces [343].

Relevance to Medolution

This abstraction layer independent from any communication technology and vendor might be a good basis for Medolution device layer as exactly such mechanism is needed for device virtualization. Also reliability and dependability was in focus of OSAmI and is therefore integrated into the API, but needs to be extended by Medolution as medical device certification and modelling of dependability was not in the scope of OSAmI.

6. CloudPort Project

CloudPort [344] is an R&D project financed by the French government as part of the Call for Project n.1 on Cloud Computing, which started in November 2011 and ended by April 2014. The project involved the following four French industrial and academic partners:

- Prologue (coordinator), R&D partner developing the core services and interfaces.
- The research labs of the Institut Télécom, developing application visualization protocols.
- CityPassenger, R&D partner specialized in networks performance and security.
- The Groupe MASSA, industrial partner responsible for different use cases.

CloudPort is an original project that aims at developing a software platform allowing companies to migrate their existing applications to the Cloud without risks and with a guaranty of interoperability among Clouds (private or public) and of reversibility. These companies will therefore be able to develop their activity dynamically in SaaS mode, and thus respond to the market's evolution.

A CloudPort platform offers **unique tools and interfaces** aimed to provision, operate and manage Cloud resources, whatever is their provider. Currently, each Cloud provider proposes its own tools and interfaces, so that there are as many tools and interfaces as Cloud providers. The tools and interfaces brought by the CloudPort platform allow users to access the resources of all Cloud providers the platform supports. The platform also includes tools for billing consumed resources and services. Companies can migrate and deploy their applications for themselves exclusively, or place them into the **application catalogue** provided by the CloudPort platform. Applications in the catalogue are available for rent and deployment to the other companies registered on the platform. The CloudPort project also includes development of technologies which allow end users to **access applications from many types of terminal**s, including PC, thin client, smartphone and tablet.

At all levels (access, data, application, infrastructure) **security** seems to be one of the first obstacles for companies to adopt Cloud SaaS services. A CloudPort platform implements several ways, as certificates usage or VPN network management, to improve confidence in the Cloud computing solutions security.

Brief overview of CloudPort Technical architecture
As depicted in Figure 69 below, the diverse modules and services of a CloudPort platform are gathered into five main components:

- The Customer Management Framework (CMF)
- The Application Management Framework (AMF)
- The Terminal Management Framework (TMF)
- The Infrastructure Management Framework (IMF)
- The Service Management Framework (SMF)

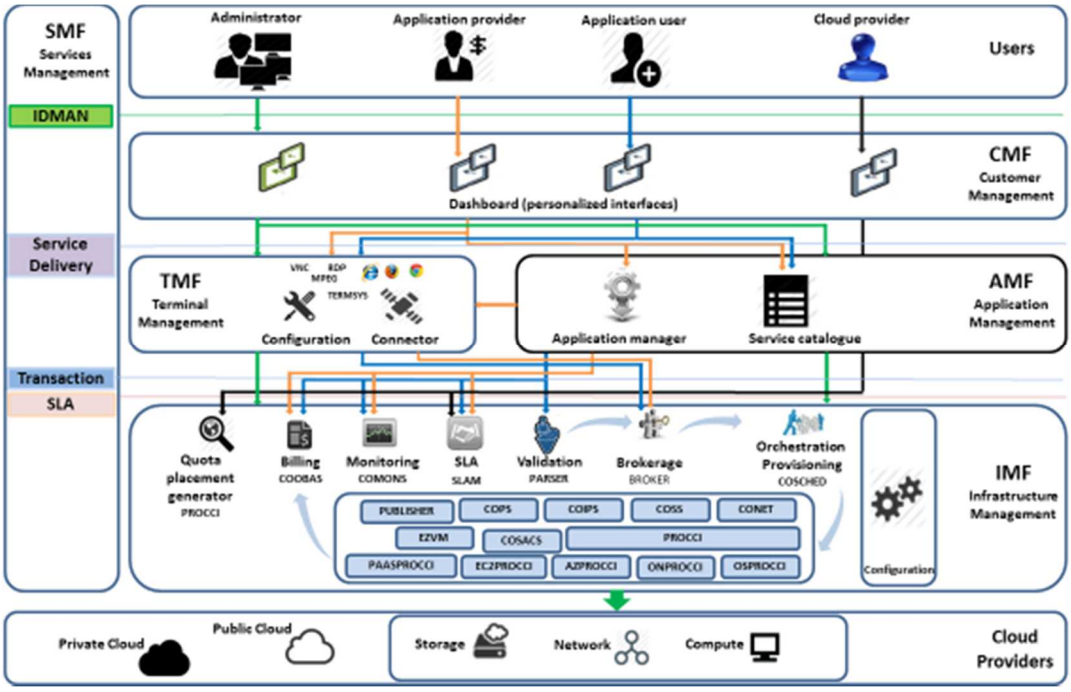

Figure 69 The components of a CloudPort platform [345]

The Customer Management Framework (CMF) provides end users with a WEB 2.0 portal (dashboard) access to the CloudPort platform. The CMF brings tools to manage user accounts, view application catalogue or initiate rent and deployment of applications. Each type of user has a personalized interface to perform their needed operations. The CMF also allows the end users to launch and display their applications.

The Application Management Framework (AMF) provides the software environment to manage entirely the application images that will be deployed on the Cloud via the CloudPort platform. The application catalogue is managed by the AMF.

The Terminal Management Framework (TMF) provides all the necessary tools to manage and control efficiently the various types of end user terminals connected to the applications running in the CloudPort platform. These tools are in charge of declaration, configuration, connection, monitoring and supervision of the terminals. The TMF also provides services based on MPEG4 and HTML5 technologies which allow applications to be displayed on various types of terminals, such as thin client, smartphone or tablet.

The Infrastructure Management Framework (IMF) provides the other CloudPort frameworks with a collection of services used to book, instantiate, monitor and manage Cloud resources (compute, storage and network). Through the integration of several specific connectors the IMF allows these resources to be provisioned on various public or private Cloud providers (AWS, Windows Azure, Cloudwatt, Openstack, etc.). The IMF was originally built on a technological platform based on the results of the CompatibleOne research project (Open Source Cloud Management and Brokering project) [346]. It comprises several services whose structure is compliant with the Open Cloud Computing Interface (OCCI) specifications and APIs [347]. Finally, the Service Management Framework (SMF) provides services to technically and commercially manage the CloudPort platform itself.

Relevance to Medolution

Since its end, the CloudPort project developments have been improved by Prologue into a Cloud management platform named “Use it Cloud Broker”. Use it Cloud Broker has already been used in

the MEDUSA project as a base for the deployment and the display of various Medical applications (see Section 7.1). Use it Cloud Broker could be used in the Medolution project for the same purpose, with additional functionalities including complex deployments of Big Data and IoT platforms, and the management of these platforms.

7. OpenIoT Project

OpenIoT is a 287305 FP7 project [348] that ran from 2012 to 2014, involving 9 prominent open source contributors from 8 countries. It implements an open source middleware for getting and processing information from sensor networks, with a main goal to facilitate the use of sensors in ICT based services for smart cities, manufacturing and agriculture.

The OpenIoT project offers services which allow the efficient usage and management of Cloud environments for almost any IoT resources (actuators, smart devices, sensors processing algorithms, etc.). It also offers pay-as-you-go IoT services. Deploying these services in Cloud environments enables the concept of “Sensing-as-a-Service” through the middleware platform, which finally results in open large-scale intelligent IoT applications.

An OpenIoT platform provides IoT tools and services for:

- Dynamically discovering sensors and their data.
- Collecting, filtering and processing the information stemming from the internet-connected objects (sensors). For this purpose OpenIoT integrates and enhances results from the Global Sensor Networks (GSN) project [349]. GSN is an extensible software infrastructure for rapid deployment and integration of heterogeneous wireless sensor networks, tested with Mica2, Mica2Dot, TinyNodes, Wisenode, Wired & Wireless cameras, several RFID readers, etc.
- Streaming and storing the processed data into an optimized Cloud computing infrastructure.
- Annotating data and linking data from multiple sensors. For this purpose OpenIoT integrates and enhances results from the Linked Sensor Middleware (LSM) project [350]. LSM brings together the live real world data from sensors and the Semantic Web.
- Dynamically querying sensors and their data.
- Visualizing produced data based on appropriate mashups (charts, graphs, maps, etc).

Brief Overview of OpenIoT Technical architecture

The OpenIoT Architecture is an instantiation of the reference architecture of the European Research Cluster on the Internet of Things (IERC). As depicted in Figure 70 below, the diverse components of an OpenIoT platform are gathered into three different logical layers; the Utility-Application Plane, the Virtualized Plane and the Physical Plane.

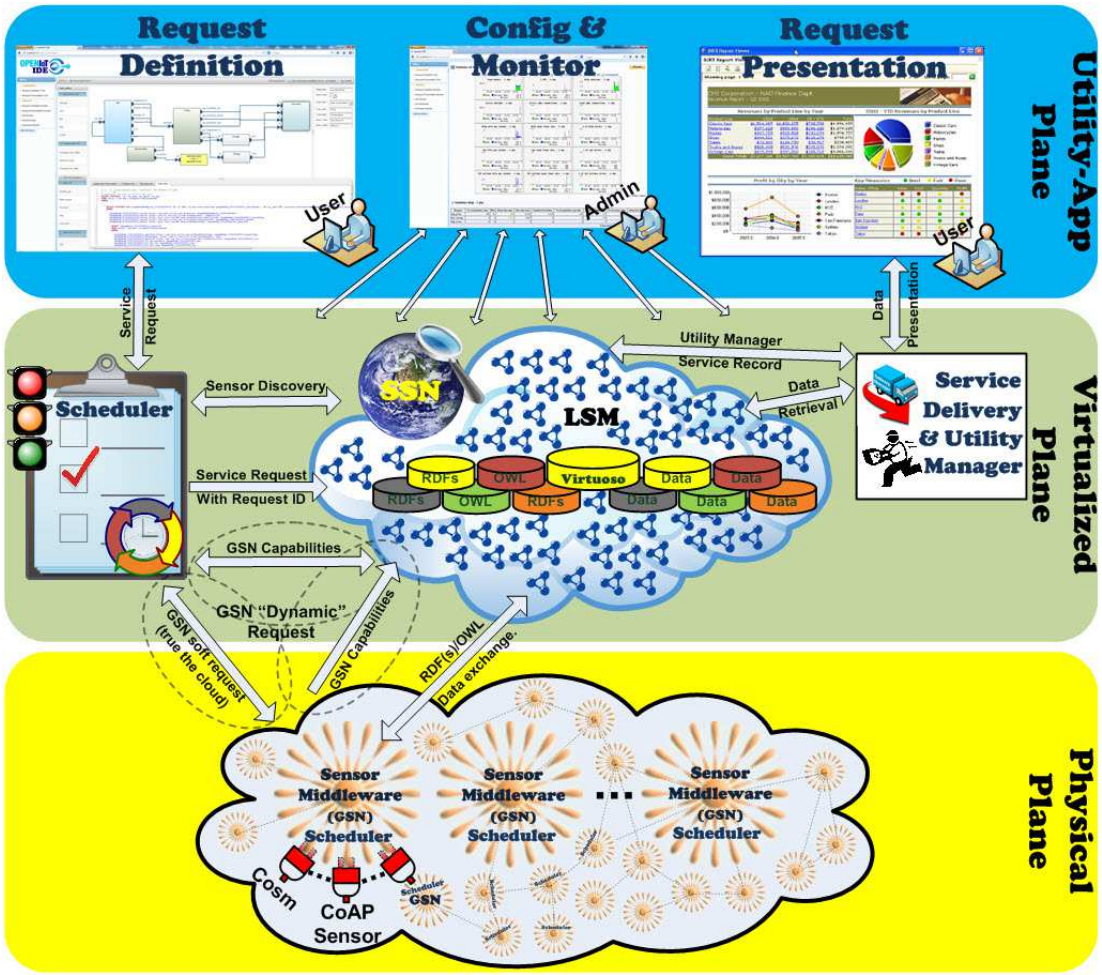

Figure 70 The components of an OpenIoT platform [351]

On the **Utility-Application layer** there are three tools:

- For on-the-fly definition of requests for services of the OpenIoT platform. Via a Web 2.0 interface, users can specify, formulate requests and then submit them to the Scheduler component of the platform.
- To visualize the data resulting from the requests in a Web 2.0 interface. A set of libraries helps to get chart, graph or map views from the data. The tool uses the Service Delivery & Utility Manager component in order to retrieve the relevant data.
- Dedicated to the administrators, who can configure and manage the sensors and the OpenIoT services. They can also monitor the health of the different services deployed into the platform.

The **Virtualized layer** comprises these main components:

- The **Scheduler component** accepts the various service requests from the request definition tool and insures that the required resources (data streams for example) are accessible. It first tries to discover the relevant sensors (and their data) to setup the service, then manages the service and its resources. The usage of semantically rich descriptions about sensor data and metadata (according to the W3C Semantic Sensor Networks (SSN) specifications) facilitates interoperability features and dynamically discovering.
- The Service Delivery & Utility Manager component is in charge of:

- Delivering each requested service to the presentation tools (internal or third party tool). To this end, it uses service workflows and combined data streams (via the SPARQL request language included into the Scheduler).
- Keeping track of metrics for these services in order to bring functionalities such as accounting, billing or resource optimization.
- In the **Cloud Data Storage component,** the data streams from Sensor Networks are stored and the above components access to it. The Cloud infrastructure also stores the functional data of the OpenIoT platform itself. Using such Cloud storage has also involved adding Cloud-based streaming processing capacities to the Linked Sensor Middleware (LSM) project modules.

The **Physical layer** includes the **Sensor Middleware**, which links the OpenIoT platform to the (virtual or physical) sensors. Based on an extended version [352] of the Global Sensor Network (GSN) project modules, it is deployed on one or several nodes (depending on the amount of sensors) to collect, filter and process data streams.

Relevance to Medolution
In the OpenIoT architecture, Cloud computing implementations offer dynamic and on-demand IoT resources and capabilities following a Cloud/utility based paradigm. This capability might be used in the framework of Medolution project especially on the subproject related to Cloud management resources. Furthermore, OpenIoT proposes a mechanism for optimizing resources within Cloud computing infrastructure. This ability could improve cost effectiveness of the solution and is developed in particularity respecting the pay-as-you-go model. Moreover, OpenIoT architecture proposes the means of collecting and processing data from any virtually available sensor in the world, including physical devices, sensors processing algorithms, etc. This work could be relevant to Medolution project especially on subproject work package 6 and task 6.1 – resource virtualization, connectivity, and collaboration.

8. Cloud4Health Project

Cloud4Health [353] is a German Cloud computing project that focuses on solutions for secure processing of personal medical data in healthcare analytics. It has been carried out from 2011 till 2014 by Rhön-Klinkum AG, Frauenhofer SCAI, Universitätsklinikum Erlangen and Averbis GmbH. The project has been funded by the German Federal Ministry of Economics and Technology as a part of the funding program “Trusted Cloud”.

Brief Overview of Cloud4Health Architecture
During the in-patient treatment, many accompanying documents emerge. Generally, the documents include a lot of unstructured information. This information exhibits a great potential for studies and medical research, but before it can be used, the information needs to be structured. As the amount of documents increase, the procedure of information structuring cannot be handled by a single clinic infrastructure. Using Cloud computing solutions has been proved to be a promising approach provided that the patient’s data confidentiality is guaranteed throughout the whole lifecycle of data processing [354].

Cloud4Health system offers several services on the SaaS basis. A service of a special interest is the text mining service, which applies Natural Language Processing (NLP) techniques for analysing clinical data.

Three main security-relevant areas of research are addressed by the Cloud4Health project. The first one is the pre-process and anonymization of the patent data before it is processed by the text mining service. The second one includes clinic-internal pseudonymization techniques of the patients' data applied. This allows tracing patient's documents after analysing them in the Cloud. The third one concentrates on connecting all documents of one patient with each other while applying pseudonymization techniques them across clinics.

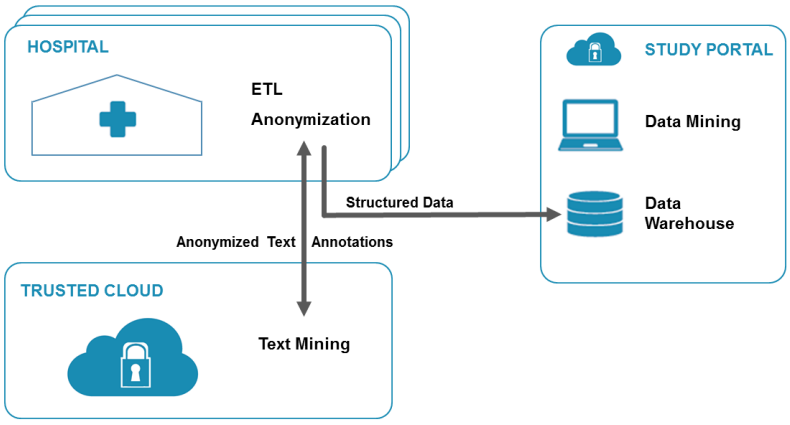

Figure 71 Cloud4Health Architecture [353], [354]

The Cloud4Health architecture is presented in Figure 71.

The developed components have been introduced to an in-depth risk analysis that was oriented at the procedures of the IT Baseline Protection standard created by the German Federal Office for Information Security (FOIS). According to [354], the following security-relevant aspects were handled by the system within the test bed:

- ***Secure data transfer over public networks***
  The data transferred from the clinic to the text mining services is encrypted with OpenVPN. The definition of key lengths and the selection of cipher suites follow the guidelines of the FOIS [355], [356].
- ***Exclusive text mining services for each user***
  The text mining services are running on virtual machines. Each user has its own virtual machine and has no authorization to access any other virtual machine.
- ***Limited lifetime of virtual machines***
  After the text mining process ends, the virtual machine terminates immediately. The patient data resides in the Cloud as long as it is needed.
- ***No persistence of personal patient data***
  Temporary files such as e.g. log files are deleted when the virtual machine is shutdown. No patient data is stored on the image of the virtual machine or elsewhere in the Cloud.
- ***Secure virtual machine image storage***
  The template images of the virtual machines are stored in a secure central storage and on-request can be copied to a respective execution node via an encrypted channel (SCP). The key length and encryption techniques are adjustable.
- ***Separation of Cloud-internal communication***
  Multiple text mining services can be used to enhance the performance. The communication between the text mining virtual machines is supported by setting VLANs up dynamically.

Relevance to Medolution

The security relevant aspects approached by Cloud4Health project are also to be addressed within the Medolution project.

9. TRESOR Project

TRESOR [357] (TRusted Ecosystem for Standardized and Open Cloud-based Resources) project has been funded by the German Federal Ministry of Economics and Technology as a part of the funding program "Trusted Cloud". TRESOR has been carried out by Deutsches Herzzentrum Berlin (DHZB), Paulinen Krankenhaus, Medisite, T-Systems, Technical University of Berlin and Bitplaces from 2011 till 2015.

Within the project, a Cloud ecosystem has been developed, which "provides a modern, secure, and legal compliant way in consuming and trading Cloud services and focuses especially on sensitive sectors, such as the health care industry while minimizing lock-in effects [358]. Different components build up the Cloud ecosystem: the Cloud broker, the Cloud proxy, and an open PaaS platform. The architecture of TRESOR is shown in Figure 72 below.

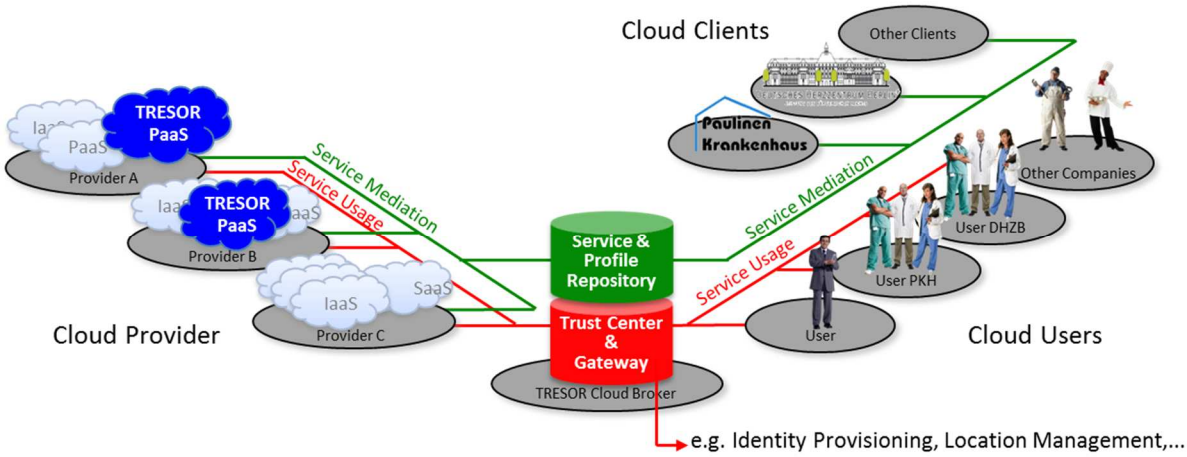

Figure 72 TRESOR Architecture [359]

The Cloud broker provides services from the marketplace to the customer in accordance with legal, company and security policies. These policies can depend on the location of the consumer (e.g. privacy requirements). The location management component is used to gather and exchange the location information with services and applications. The technologies used to gather the information are: GPS, Cell-ID, WLAN and IP-based Positioning. The location information can also be used to define an access model for the Cloud [360], [361]. E.g. the user can only login to the Cloud if he/she is in the hospital. The Cloud providers gain additional support from the Cloud broker with a PaaS solution in order to provide TRESOR services. But there is also the possibility to connect their solution without entering the PaaS.

Relevance to Medolution
The security relevant aspects approached by TRESOR project are also to be addressed within the Medolution project.

10. BaaS Project

BaaS (Building as a Service) [362] is an ITEA 2 project that started in November 2013 and will end in December 2016. The consortium of the BaaS project shown in it involves 17 partners from 4 countries. (See Figure 73 below).

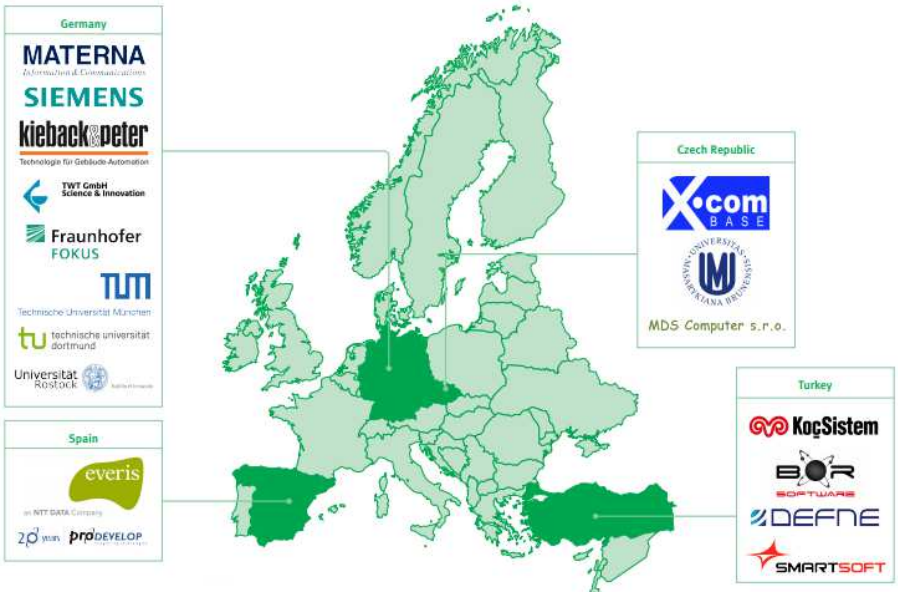

Figure 73 BaaS Consortium [362]

The motivation of BaaS is to create cross-domain management and control functions for smart buildings. Today, the building automation domain suffers from a separation of automation disciplines and a lack of integration of all available sources of information. Building automation systems on their own are separated into largely independent control systems, e.g. for lighting control, HVAC (heating, ventilation and air-conditioning), safety and security. To fill this gap, the BaaS project establishes a generic service platform for smart commercial buildings that integrates traditional building automation and management systems with ICT infrastructures. In order to attain this objective, the BaaS project is aimed at the design and development of:

- a flexible open building service platform comprising basic building services and facilitating the easy specification, generation and deployment of value-added building services at considerably lower cost compared to the state of the art;
- a BaaS data model providing additional meta-information to simplify the engineering of value added services and applications for the BaaS system and the integration of legacy systems;
- mechanisms for annotation of data providers and data consumers with meta-information regarding the functional semantics and non-functional properties of their data offerings or demands, respectively;
- model-based mechanisms for analysis, aggregation and transformation of data according to the meta-information provided in the BaaS data model;
- concepts for a “building information sphere” for facility managers and building occupants; all stakeholders may be actively involved as producers and consumers (“prosumers”) of information;
- a building system complying with established standards of today’s IT infrastructures, e.g. data transport protocols and IT security.

Relevance to Medolution

In the BaaS project, functions and components for the automated high level management of heterogeneous devices and services are developed which might be of use also in device-based Medolution systems. Furthermore, some dependability aspects were considered through fault tolerance patterns and corresponding implementation components. Particularly, fault tolerance patterns and components supporting the automated reconfiguration of the system and its services

were designed and developed. Moreover, a useful aid is the Management Tree that was developed through the BaaS project by Materna and TU Dortmund. The management tree forms a virtual data access structure supporting the homogeneous access to distributed and heterogeneous management data. It offers a hierarchical view on the whole system. The management data is accessed remotely by means of management agents. In addition, the Management Tree facilitates the verification of the system and the visualization of modifications on the service layer.

11.I-Treasures project

I-Treasures (Intangible Treasures - Capturing the Intangible Cultural Heritage and Learning the Rare Know-How of Living Human Treasures FP7-ICT-2011-9-600676-i-Treasures) is an Integrated Project of the European Union's 7th Framework Programme 'ICT for Access to Cultural Resources'. Cultural expression is not limited to architecture, monuments or collections of artifacts, but also includes fragile intangible live expressions, which involve knowledge and skills such as music, dance, singing, theatre, human skills and craftsmanship. These manifestations of human intelligence and creativeness constitute Intangible Cultural Heritage (ICH) [363].

The main objectives of i-Treasures Project were to develop an open and extendable platform, to provide access to the ICH resources, enable knowledge exchange between researchers and contribute to the transmission of rare know-how from Living Human Treasures to apprentices. The project also aimed to propose novel methodologies and new technological paradigms for the analysis and modelling of ICH, in particular a methodology based on multisensory technology for capturing the hidden/never analysed information for the creation of information (intangible treasures) that will be transmitted to new generations. High level semantics were extracted enabling researcher to identify possible implicit or hidden correlations between different ICH expressions or different interpretation styles of the same ICH and study the evolution of a specific ICH through its transmission from generation to generation or to other communities. Combining conventional learning procedures and advanced services, such as Singing Voice Synthesis and sensorimotor learning through an interactive 3D environment, i-Treasures made a significant input in education and knowledge transfer of ICH.

Relevance to Medolution
In i-Treasures project, visualization and simulation of derived information was created via high-technology 3D interfaces which was an important part of the overall project. Human–computer interaction (HCI) researches the design and use of computer technology, focused on the interfaces between people and computers. Researchers in the field of HCI both observe the ways in which humans interact with computers and design technologies that let humans interact with computers in novel ways [364]. Hence, much of the research in the field seeks to improve human-computer interaction by improving the usability of computer interfaces.

Throughout the Medolution project, big data visualization in a pictorial or graphical format will be the main interaction platform between human and the display interfaces. Therefore, decision makers will see data analytics visually that will help to grasp difficult concepts or identify healthcare situations with more concrete and understandable visual patterns. Using charts or graphs to visualize large amounts of complex data is easier than poring over spread sheets or reports. The HCI experiences from the i-Treasures project with different types and volumes of big visual data, will guide to have better visual analytics platforms for the Medolution project. The effectiveness metrics of the visual data representations will be modelled and monitored with the methodologies of the i-Treasures measures.

Throughout Medolution project, the HCI experiences from the i-Treasures project can be used to define better visual applications in all aspects.

12. OFERTIE Project

OpenFlow Experiment in Real-Time Internet Edutainment (OFERTIE) project is the EC FP7 programme project run in 2013-2014, which aimed to use software-defined networking (SDN) approaches to improve delivery of an emerging class of distributed applications for the Future Internet known as Real-Time Online Interactive Applications (ROIA). OFERTIE aims to enhance and use the OFELIA Testbed for OpenFlow Programmable Networking to run experiments to establish how programmable networks that can be used to support technical solutions such as multicast and managed QoS, as well as investigate what business models and value chains would be able to use these solutions in an economically sustainable fashion [365]. Some of the advantages of SDN are listed as below:

1. The ability to allow networks to keep pace with the speed of change.
2. SDN creates a framework to support more data-intensive applications like Big Data and modelling.
3. By abstracting Cloud resources using software defined networking, it's easier to unify Cloud resources.
4. The networking components that make up massive data centre platforms can all be managed from the SDN controller.
5. Also, enterprise networks have to set up new applications and virtual machines on-demand to accommodate new processing requests such as those for Big Data.

Relevance to Medolution
The Medolution project will involve the use of devices and high volume of data, i.e. Big Data, and its processing/modelling. As described above, the biggest promise of software defined networking (SDN), which was the main subject of OFERTIE project, is that it will centralize and simplify control of enterprise network management with traffic programmability, greater agility, the ability to create policy-driven network supervision, and implementing network automation. Therefore, the related development realized in the OFERTIE project can be used to create a more robust networking architecture/paradigm for Medolution project.

The above mentioned policy-based approach to the management of medical systems has been studied and applied recently in a few research projects discussed below. Such policy-based management approaches, which will be further researched and applied within the management system to be provided for Medolution project. Particular focus and relevant outcomes of each of the projects are highlighted.

13.AMUSE Project

The project Autonomic Management of Ubiquitous Systems for e-Health (AMUSE) was carried out during 2004 – 2007 as a joint collaboration between the University of Glasgow and Imperial College, London [366], [367]. The main focus of the work was to develop architecture for autonomic management of ubiquitous computing environments in the home healthcare sector.

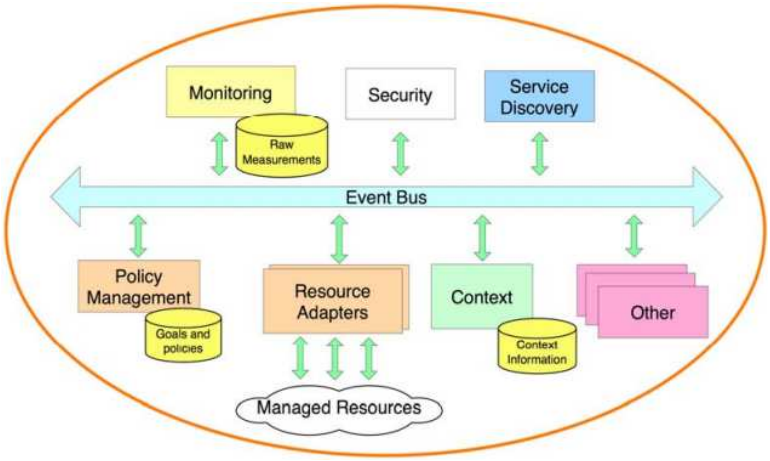

Figure 74 Self-Managed Cell (SMC) Architectural Pattern

The Self-Managed Cell (SMC) was proposed as a policy-driven architectural pattern for implementing autonomic ubiquitous systems (See Figure 74) [366]. A SMC manages a set of managed resources uniformly using resource adapters. The communication with the resources is therefore independent from the used communication protocol and the resource interfaces. The common event bus provides the interaction with the offered services by using a router to forward event notifications from the event publishers to the subscribers. This approach permits to decouple the services, so that the sender does not know the listeners of the event. The advantage of this is that the new services could be added more comfortably without interrupting the others. Furthermore, the concurrent and independent response of multiple services to the same event is facilitated. As well as the communication overhead could be lowered by transmitting only the measured data that exceeds the specified threshold. Self-management and adaptation are performed by means of the policy service that conducts a basic feedback control loop. On changing in the state of the managed objects the corresponding reconfiguration actions in form of events are forwarded to the event bus (See Section 6.2 on model-based management of medical systems).Which actions are to be executed is a subject to obligation policies. These are represented by means of event-condition-action rules. The authorization policies define which actions may be performed on which resources. As an implementation of the policy service the authors present Ponder2 [109] the successor of Ponder [108], a policy definition language and toolkit developed at Imperial College, London. The Ponder2 compounds a general-purpose object management system with a domain service providing a hierarchy for the managed objects, an obligation policy interpreter for handling the obligation policies, a command interpreter performing invocations on the managed objects and an authorization enforcement supporting fine-grained authorizations for the managed objects. The detection of new devices or other SMCs is a task of the discovery service. It is responsible for generating the corresponding component-detected and component-left events as well as for distinguishing between the transient disconnections and permanent device departures. For managing more complex environments several SMCs could be composed or collaborate with each other. The composition of SMCs allows managing more smart diagnostic devices which manage in their turn their own resources. The interaction of multiple SMCs permits scenarios where new policies from other SMCs are to be loaded or updated. The requirements of one SMC for interacting with another are defined within its mission which is a group of policies determining the communication behavior with the other SMC.

14. CareGrid Project

The project Autonomous Trust Domains for Healthcare Applications (CareGrid) has been a collaborative work between the Imperial College, London and the University of Cambridge during 2005 – 2008 [368], [369]. The main aim of the project is developing a middleware for supporting decisions based on trust, privacy, security and context models in a healthcare application domain. A targeted framework should include an architecture, which would consist of diverse services and support their interaction and administration. The integration of the UK electronic health record (EHR) service is to be supported. The framework is policy-based, providing a mechanism for controlling access to the medical data and dynamic adaptation of the system. Monitoring and archiving functions, which comprise also system reliability and performance monitoring, are of a special interest.

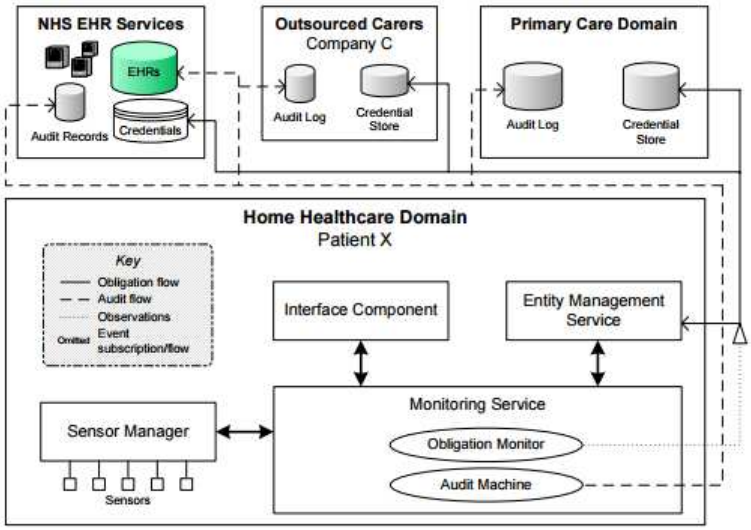

Figure 75 CareGrid Architecture [369].

The home healthcare domain interacts with various domains, including hospitals, homecare providers, specialists, social care providers and many others (See Figure 75). The coordinating domains (e.g., the National Health Service (NHS)) ensure the compliance of the provided services and the collaboration of the domains. The primary care domain, such as hospital, creates a home-based patient care environment. Diverse other domains could provide specific services. To act as an authority for validating and verifying entities, domains require a credential store.

The developed architecture for a home healthcare domain includes the following components: sensor manager, interface component, entity management service and monitoring service. All the aspects concerning the use of sensors including sensor discovery, failure detection, stream management and data capture belong to the tasks of the sensor manager. The sensor manager is also in charge of the evaluation of the captured data and the invocation of the appropriate responding actions. The user interface and the interface for access by devices are provided through the interface component. The entity management service tracks the devices and services within the domain and defines the privileges according to the actual policies. The core of the infrastructure is the monitoring service observing all the interactions between components passing through a monitoring pipeline. The monitoring service offers two components: audit machine and obligation monitor. The first is responsible for transferring relevant information to various audit logs (e.g., to the electronic health record) according to the defined policies. The obligation monitor launches compensatory actions in case of a failure in obligation fulfillment. It also informs the credential services about the performance of the system.

The proposed framework is policy-driven both at the system-level (e.g., defining events, actions and domains) and the user-level (e.g., defining thresholds for relevant parameters). Ponder2 [109] was used as a policy language supporting obligation and authorization policies. The entities to which policies apply are organized in hierarchical domains of managed objects. The managed objects are associated with a set of data, which is used by authentication, authorization and obligations. The domain hierarchy for each component of the system is maintained by the local policy interpreter.
For execution of actions on managed objects from the external domains, proxies are created by the local interpreter and inserted in the local domain structure. The developed conflict resolution strategy used statically and dynamically is a subject of [370].

15.MATCH Project

The research project Mobilising Advanced Technologies for Care at Home (MATCH) has been carried out during 2005 – 2009 by the universities of Stirling as a lead partner, Glasgow, Edinburgh and Dundee [151], [371]. The main aim of the project is to develop advanced technologies in support of social and health care at home, particularly in the area of home network services, lifestyle monitoring, speech communication and multimodal interfaces. OSGi has been selected as an ideal technology for the implementation as a vendor-neutral, device-independent approach to service provision. The management of home networks is to be accomplished using policies that allow multiple stakeholders to configure the system behavior. The use of ontologies enhances the discovery of services and the use for policies managing these services. (See Figure 76 below).

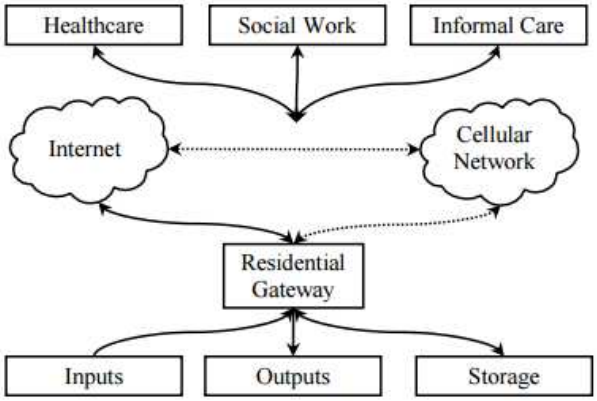

Figure 76 MATCH System Architecture

The proposed architecture (See Figure 77 below involves OSGi residential gateway embedding the home services and device control [371]. The sensors (e.g. physical devices, logical or user-oriented data sources) provide the inputs of the system. The outputs invoke the actuators, which could also be in their term physical, logical or relating to user. The link to the outside world is usually via a broadband connection to the Internet or a direct link to a cellular network. The information is captured and saved in the storage saved information could be forwarded to the care providers (e.g. healthcare centers, social work departments and informal careers).

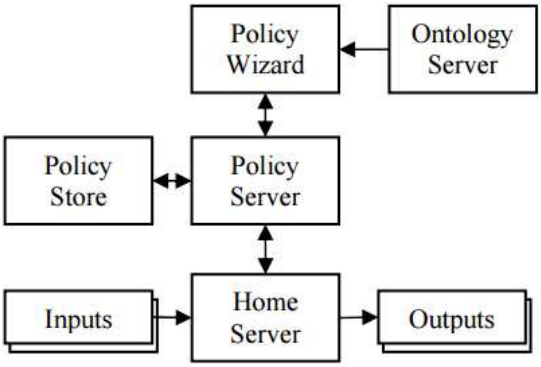

Figure 77 MATCH Policy System Architecture [151]

Figure 77 captures the design of the proposed policy-based management system. The residential gateway houses the management system managing devices and services by means of predefined policies expressed as trigger-condition-action rules. These are formulated in a language APPEL (ACCENT Project Policy Environment and Language) [110] in the form of XML documents. A web-based policy wizard was developed for the remote policy edition and creation. Domain-specific knowledge of concepts and relationship of policies is integrated using an ontology server, a system called POPPET (Policy Ontology-Parsing Program – Extensible Translation). The policy store is used as a repository for holding user profiles, the system configuration and state. The latter two allow the policies to refer to abstract terms and to be interpreted depending on context. The interaction with the policy system is a task of the home server. The communication is performed by sending and receiving events. The home server notifies the policy server about a triggering event. The policy server selects the corresponding policies, evaluates them and responses accordingly. Conflict handling is performed by means of high-level resolution policies which are triggered by a conflicting action and conduct the resolution according to some given high-level criterion [372].

16. SmartHEALTH Project

SmartHEALTH Integrated Project has been carried out during 2005 – 2009 and coordinated by the University of Newcastle upon Tyne. One of the main objectives of this large project was to introduce new SmartHEALTH sensor systems for delivering novice healthcare services and for improving the existing ones [373]. Furthermore, the role of Ambient Intelligent (AmI) medical devices and online services for pervasive healthcare provision is to be demonstrated.

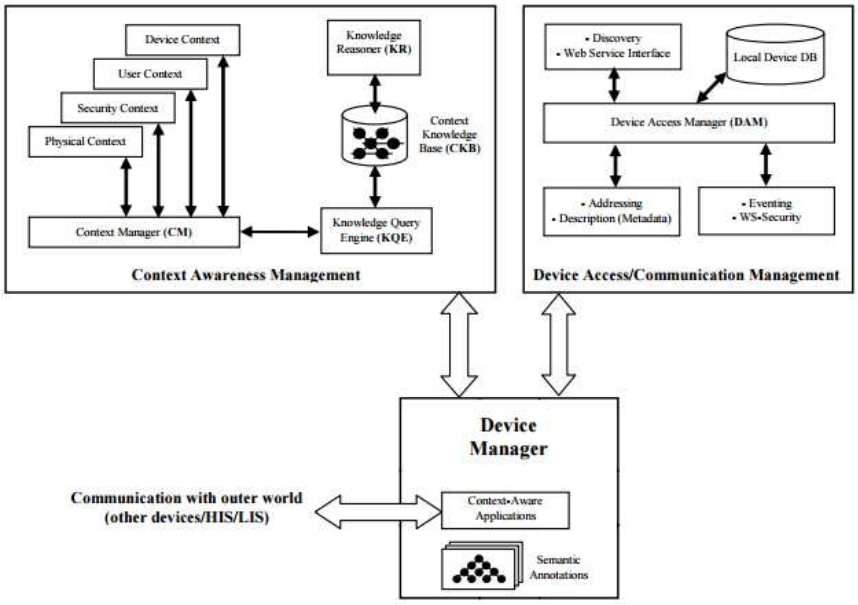

Figure 78 SMDS AmI Device Architecture [373]

An infrastructure for the interoperability of ambient intelligent medical devices, Semantic Medical Devices Space (SMDS), was proposed within the scope of the project. The architecture of an AmI medical device comprises the following components: Context Awareness Management, Device Access/Communication Management and the Device Manager (See Figure 78). The Context Awareness Management manages the context awareness behavior of the medical device. The retrieval of the contextual information is a task of the Context Manager, which communicates directly with the sub-components providing information about the current device, user, security and physical context information. The contextual information is provided in form of context markups (i.e. RDF graph). This helps the Context Manager to query the Context Knowledge Base with the help of the Knowledge Query Engine and reason about the high-level context through the Knowledge Reasoner (implemented by using Jena2 Semantic Web Toolkit [374]). The Context Knowledge Base is a persistent knowledge storage linking the extended context ontology for a particular environment (i.e. hospital) and the context markups gathered from the sub-components. In this way a single semantic model is established. As context query language any RDF Data Query Language (e.g. SPARQL) can be used, because it allows querying, using declarative statements, over semantic models based on triples (<subject, predicate, object>) patterns. The Device Access Manager supports communication with the Addressing, Discovery, Description (Metadata), Web Service Interface and Eventing components. To provide the semantic description of the devices and services, it is suggested to use existing ontologies (e.g. FIPA [111] or CC/PP [112]). The Local Device DB stores the measurement results captured by the device and offers functionalities to retrieve them through Web Services or to send them to the remote Hospital Information System (HIS) and Laboratory Information System (LIS). The Device Manager manages the collaboration of the components.

17. Conclusions

While the Medolution project builds upon the results of the Medusa project that provides collaborative cloud access to medical information relevant in critical situations and addresses security, latency and collaboration related aspects, the outcomes of other European projects in Healthcare data processing, discussed in this Appendix, constitute an important part of the technical state of the art to be considered, utilized and updated according to the Medolution project plan.