A Demonstration of 4D Digital Avatar Infrastructure for Access of Complete Patient Information

Project acronym: MyHealthAvatar

Deliverable No. D3.4
Technical report on the links with external data resources

Grant agreement no: 600929

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

COVER AND CONTROL PAGE OF DOCUMENT	
Project Acronym:	MyHealthAvatar
Project Full Name:	A Demonstration of 4D Digital Avatar Infrastructure for Access of Complete Patient Information
Deliverable No.:	D3.4
Document name:	Technical report on the links with external data resources
Nature (R, P, D, O)[1]	R
Dissemination Level (PU, PP, RE, CO)[2]	PU
Version:	1.0
Actual Submission Date:	30/11/2015
Editor: Institution: E-Mail:	Manolis Tsiknakis TEI-C tsiknaki@epp.teicrete.gr

ABSTRACT:

This deliverable describe MHA architecture to support the linking of the 4D avatar with the external sources as an interface to extract data and information. The purpose of this document is to describe the architectural consideration and design that was used to create the technical solution to allows the access of information from a range of external sources, e.g. individual hospital records, and other

[1] R=Report, P=Prototype, D=Demonstrator, O=Other
[2] PU=Public, PP=Restricted to other programme participants (including the Commission Services), RE=Restricted to a group specified by the consortium (including the Commission Services), CO=Confidential, only for members of the consortium (including the Commission Services)

data/model repositories, and also supports data management, transfer, security and sharing controlled by individuals

External resources include :

1) Social network linkage. This will investigate the infrastructure of the links to social networks. The 4D avatar is superimposed over the social graph of the patients. This task will provide social web mechanisms and encourage the patients to adopt those in order to define their digital avatar. This requires integration with the social web accounts that the patients maintain already and the extraction of the social graph and other information from there.

2) External data warehouse linkage. To increase the interoperability and accessibility, the links to other public databases and data warehouses will be considered.

3) Hospital Information Systems and records linkage. This will look into the feasibility of the links to the hospital systems to allow the exportation of the health related data of the patients from the linked hospital systems. The predominant issues in this task relate to the security and the transformation of the data followed by the proper annotation in order to be compliant with the syntactic and semantic principles of the system. Notably, we do not expect to realize the physical link in the duration of this project due to complex legal issues. This will present as a feasibility study on the technical side, which will be in conjunction with the work on the legal side in WP11. A simulated database can be built to allow the investigation.

We emphasize on standardization and interoperability issues considering all known standards and related work [EHR, HL7, OpenEHR. EpSoS. Etc.] to allow data exchange. Also, the task will be carried out in conjunction with Taks 6.1, in which data collection utilities will be developed to allow information extraction from the social network using Twitter and Facebook APIs.

KEYWORD LIST: Link with external sources, Health Data, Clinical data, Electronic Health Records, Social Networks, Biomedical Research, EPSOS, Patient Summary, Drug databases, Drug Drug Interactions, Security, Legal Requirements, IPR

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 600929.

MODIFICATION CONTROL			
Version	Date	Status	Author
1.0	06/10/2014	Draft	Manolis Tsiknakis Emmanouil G. Spanakis
1.1	05/12/2015	Draft	Stelios Sfakianakis, Kostas Marias
1.2	05/01/2016	Draft	Emmanouil G. Spanakis, Stelios Sfakianakis
1.4	25/01/2016	Draft	Feng Dong, Nikolaus Forgo, Sarah Jensen, Zhikun Deng, Haridimos Kondilakis, Iryna Lishchuk
2.0	15/02/2016	Final	Manolis Tsiknakis , Emmanouil G. Spanakis, Kostas Marias,

List of contributors

- Tsiknakis Manolis (TEI-CRETE)
- Emmanouil G. Spanakis (FORTH-ICS)
- Stelios Sfakianakis (FORTH-ICS)
- Irini Genitsaridi (FORTH-ICS)
- Demosthenes Akoumianakis (TEI-CRETE)
- Vivi Fragopoulou (TEI-CRETE)
- Marc Stauch (LUH)
- Sarah Jensen (LUH)
- Iryna Lishchuk (LUH)
- Marios Spanakis (FORTH-ICS)
- Nikolaus Forgo (LUH)
- Haridimos Kondylaksi (FORTH-ICS)
- Feng Dong (BED)
- Kostas Marias (FORTH-ICS)

Contents

1 Executive Summary

A core objective of the MyHealthAvatar (MHA) architectural specification is to enable the seamless linking of the 4D avatar with the external sources. As defined in the project's DoW external sources for MHA include a) Social networks, b) Hospital Information systems (EHR systems) and c) other external data warehouses relevant to the service delivery scenarios envisaged by MHA. Especially developing an interoperability framework with Hospital records and EHR systems we plan to explore the feasibility of linking to the hospital systems, with the Hospital systems acting a source of patient's health related data for the MHA system. The predominant issues in this task relate to the security and the transformation of the data in order to be compliant with the syntactic and semantic principles of the system, i.e. in order to achieve semantic interoperability.

The objective of this deliverable is twofold:

A) to provide a thorough state-of-the-art review of other international efforts in achieving such unidirectional or bidirectional transfer of health related data, and

B) to describe the provision, specifications, legal, intellectual rights, and other measures embedded into the MHA architecture in supporting the linking of the 4D avatar with appropriate external data sources.

In addition the present document reports all those architectural considerations and core design requirements for creating an efficient linkage to different types of health and health related source. We describe the implemented technical solutions in order to allow a seamless interoperable access to health and health related information from a range of external sources. The deliverable reports on other procedural aspects that are relevant, such as data management and transfer, legal aspects for data security, security measures as well as intellectual property rights for sharing and control of the data.

We emphasize on standardization and interoperability issues considering all known standards and related work (EHR, HL7, OpenEHR. EpSoS. etc.) to allow for data exchange. Also, the task will be carried out in conjunction with Task 6.1, in which data collection utilities will be developed to allow information extraction from the social network using Twitter and Facebook APIs.

Within the context of the MHA external sources include:

1) Social network linkage. This will investigate the infrastructure of the links to social networks. The 4D avatar is superimposed over the social graph of the patients. This task will provide social web mechanisms and encourage the patients to adopt those in order to define their digital avatar. This requires integration with the social web accounts that the patients maintain already and the extraction of the social graph and other information from there.

2) External data warehouse linkage. To increase the interoperability and accessibility, the links to other public databases and data warehouses will be considered.

3) Hospital Information Systems and records linkage. This will look into the feasibility of the links to the hospital systems to allow the exportation of the health related data of the patients from the linked hospital systems. The predominant issues in this task relate to the security and the transformation of the data followed by the proper annotation in order to be compliant with the syntactic and semantic principles of the system. Notably, we do not expect to realize the physical link in the duration of this project due to complex legal issues. This will present as a feasibility study on the technical side, which will be in conjunction with the work on the legal side in WP11. A simulated database can be built to allow the investigation.

2 Introduction

The platform of MHA aims to support the 4D digital representation of a given patient but of course parts of the patient's clinical and social history are already stored and managed by third party systems. For this reason proper mechanisms and infrastructure should be in place for retrieving relevant user information from these external data sources. Whenever it's possible such "linking" with the third party systems should be based on available standard interfaces since they allow the building of generic ports and interfaces and the reuse of existing code bases. Figure 2-1 shows some notable examples for the realization of these links to external resources: Clinical data can be retrieved from Hospital Information Systems (HIS) through the Clinical Document Architecture (CDA[3]) guidelines and set of specifications, clinical trial specific patient data can be acquired using the Operational Data Model (ODM) of the Clinical Data Interchange Standards Consortium (CDISC[4]), whereas cross-border healthcare provisioning is supported by the adoption of epSOS[5] Patient Summary interfaces. Additional well-known and widely supported standards and quasi-standards include Digital Imaging and Communications in Medicine (DICOM[6]) and the "transactions" defined by the Integrating the Healthcare Enterprise (IHE[7]) initiative.

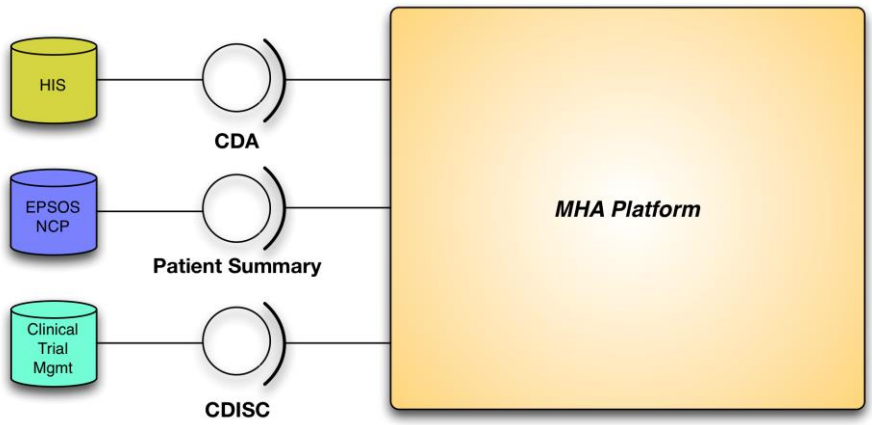


Figure 2-1: Linking MyHealthAvatar with external systems through well-defined interfaces

In the following picture we present the main external components that are linked with MHA platform to support the link of multilevel data to MHA repository in a generic way. A new architectural layer is introduced that hosts the adapters, gateways and other components which are responsible for the linking with the external data sources. The components

[3] http://www.hl7.org/Special/committees/structure/index.cfm
[4] http://www.cdisc.org/
[5] European Patients Smart Open Services (epSOS), http://www.epsos.eu/
[6] http://dicom.nema.org/
[7] http://www.ihe.net/

belonging to this layer interact with the main backbone of the MHA platform i.e. the Cassandra based data repository, the semantic infrastructure, and other repositories. In this figure we emphasize also in the semantic integration layer and the semantic transformation of these data in order to be uniformly accessible, through MHA common information data model, via MHA published APIs.

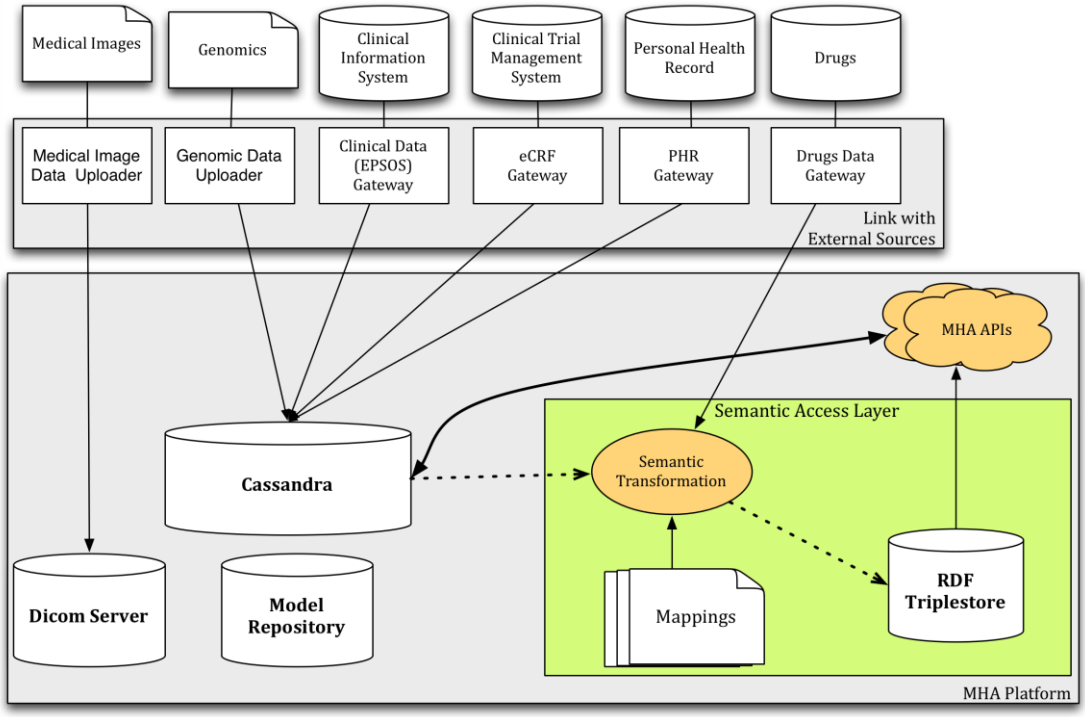


Figure 2-2: MHA external data source linkage

3 Linkage to External Health Data Sources: The Benefits and Barriers

3.1 The Need of Linkage to Health Data Sources: Literature Review

In the last decade there is a vast amount of health data being collected electronically in order to support medical research and healthcare provision[8,9,10]. The further integration of this information from the various health data sources will provide even larger datasets for analysis which have the potential to enhance the quality of healthcare. However, the distribution of data among the sources and the heterogeneity of their representation may prohibit their use and integrative analysis in a seamless way[11].

Linkage to External Health Data Sources is defined as any type of network connection that enables secure and legitimate health information flow, from a variety of external health data sources (HDS) to a recipient health data management system (RS), which also ensures the integrative analysis of the sources' information. Bidirectional linking to external health data sources refers to the ability of the RS to either receive or provide information back to the sources.

Linkage to External Health Data Sources is a realization of the concept of *Secondary Data Use*. Secondary data should be distinguished from primary data[12,13]. Primary data are collected in order to support the primary analysis or care provision while secondary data are reused to support a secondary purpose, other than the initial purpose of their collection. For example, secondary data use is performed when data acquired from a hospital's cardiovascular disease registry, are reused to support a nutrition research study, but that were initially collected to meet the needs of the cardiology department. In addition, secondary data are able to

[8] Hoerbst, A., & Ammenwerth, E. (2010). Electronic health records. Methods Inf Med, 49(4), 320-336.
[9] Ginsburg, G. S., Burke, T. W., & Febbo, P. (2008). Centralized biorepositories for genetic and genomic research. Jama, 299(11), 1359-1361.
[10] Tang, P. C., Ash, J. S., Bates, D. W., Overhage, J. M., & Sands, D. Z. (2006). Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. Journal of the American Medical Informatics Association,13(2), 121-126.
[11] Hasselbring, W. (2000). Information system integration. Communications of the ACM, 43(6), 32-38.
[12] PricewaterhouseCoopers, L. L. P. Transforming Healthcare Through Secondary Use of Health Data. 2009.
[13] Safran, C., Bloomrosen, M., Hammond, W. E., Labkoff, S., Markel-Fox, S., Tang, P. C., & Detmer, D. E. (2007). Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. Journal of the American Medical Informatics Association, 14(1), 1-9.

represent changes and facts of the past which are not available in the present and cannot be reproduced, providing thus an irreplaceable source of information.

Secondary Data Use that is achieved through linking to external health data sources has numerous medical benefits. Integrative analysis of large volumes of data derived from this linkage can facilitate both biomedical research breakthroughs and improve the quality of healthcare practise[14,15,16,17]. Secondary use of health data is able to enhance the three core processes of health management, disease prevention, early diagnosis and personalized treatment which are described below.

- *Disease Prevention* is the aim of preventive medicine[18]. Preventive medicine is concerned with keeping individuals healthy in order to prevent disease occurrences. Essentially it is the field of medicine focused on disease prevention rather than disease treatment. Clinicians with specialty on preventive medicine are able to suggest patient lifestyle changes and recommendations in order to decrease the risk of developing a disease.
- *Early Diagnosis*[19] is the process of identifying early symptoms of a disease in order to diagnose and treat the patient as soon as possible. Early disease diagnosis is a crucial process for positive patient outcomes, since the success of the treatments usually depends highly on the level of disease progression.
- *Personalized Treatment*[20,21] provides more a tailored treatment to the patient, focusing on the personal characteristics of the individual, such as his genetic information.

[14] Botsis, T., Hartvigsen, G., Chen, F., & Weng, C. (2010). Secondary use of EHR: data quality issues and informatics opportunities. AMIA summits on translational science proceedings, 2010, 1.
[15] Linder, J. A., Haas, J. S., Iyer, A., Labuzetta, M. A., Ibara, M., Celeste, M., ... & Bates, D. W. (2010). Secondary use of electronic health record data: spontaneous triggered adverse drug event reporting. Pharmacoepidemiology and drug safety, 19(12), 1211-1215.
[16] Safran, C., Bloomrosen, M., Hammond, W. E., Labkoff, S., Markel-Fox, S., Tang, P. C., & Detmer, D. E. (2007). Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. Journal of the American Medical Informatics Association, 14(1), 1-9.
[17] Elger, B. S., Iavindrasana, J., Iacono, L. L., Müller, H., Roduit, N., Summers, P., & Wright, J. (2010). Strategies for health data exchange for secondary, cross-institutional clinical research. Computer methods and programs in biomedicine, 99(3), 230-251.
[18] Rose, G. (1992). The strategy of preventive medicine. The strategy of preventive medicine.
[19] Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C. R., Jagust, W., ... & Beckett, L. (2005). Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI).Alzheimer's & Dementia, 1(1), 55-66.
[20] Harbeck, N., Salem, M., Nitz, U., Gluz, O., & Liedtke, C. (2010). Personalized treatment of early-stage breast cancer: present concepts and future directions.Cancer treatment reviews, 36(8), 584-594.
[21] Hamburg, M. A., & Collins, F. S. (2010). The path to personalized medicine.New England Journal of Medicine, 363(4), 301-304.

Personalized treatment advances healthcare which is customized for the patient, based on his genetic makeup (DNA, RNA and protein information).

In the following sub-sections, we describe numerous case studies that focus on secondary use of information from external health data sources, and discuss their rationale. The case studies are categorized based on the following distinct types of data sources that were identified in the literature (a) Electronic Medical Records, (b) Biomedical Data Repositories, (c) Personal Health Records and (d) Social Networks. These studies highlight the benefits that can be derived from an efficient linkage to external health data sources.

3.1.1 Linking to Electronic Medical Records

The term Electronic Medical Record[22] refers to a system used by healthcare professionals that supports the management of the longitudinal medical history of the patients in a healthcare facility. The definition of Electronic Health Record (EHR) system expresses the extended functionality of an EMR system that is able to support the management of patient health information in multiple healthcare facilities and enables data access to healthcare professionals across the facilities. Sophisticated solutions of EMR systems provide clinical decision support services, clinician-to-patient communication and health information access and management by the patients[23].

Case studies have been presented that utilize patients' adverse drugs reactions (ADR), recorded by clinicians in EMRs, to facilitate research on drug safety[24,25]. Secondary use of ADR data in EMRs supports the selection of the right drug treatments and the reduction of drug

[22] Boonstra, A., & Broekhuis, M. (2010). Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. BMC health services research, 10(1), 231.

[23] Häyrinen, K., Saranto, K., & Nykänen, P. (2008). Definition, structure, content, use and impacts of electronic health records: a review of the research literature. International journal of medical informatics, 77(5), 291-304.

[24] Gurwitz, J. H., Field, T. S., Harrold, L. R., Rothschild, J., Debellis, K., Seger, A. C., ... & Bates, D. W. (2003). Incidence and preventability of adverse drug events among older persons in the ambulatory setting. Jama, 289(9), 1107-1116.

[25] Bates, D. W., Evans, R. S., Murff, H., Stetson, P. D., Pizziferri, L., & Hripcsak, G. (2003). Detecting adverse events using information technology.Journal of the American Medical Informatics Association, 10(2), 115-128.

morbidity and mortality. In other studies secondary use of information in EMRs is conducted to facilitate clinical research on heart failure[26] and rheumatoid arthritis[27].

Another study presented by Kho et. al.[28] describes secondary use of information in EMRs to identify disease phenotypes with positive and negative predictive values. The study analyses the identification of phenotypes for Dementia, Cataracts, Peripheral Arterial Disease, Type 2 Diabetes and Cardiac Conduction, in the EMRs information.

These studies have shown that secondary use of information in EMR systems can support various research purposes (genetic research, drug safety research, clinical research etc.) to improve the quality of healthcare. Thus, linkage to electronic medical records in hospitals that enables secondary use of their data has the potential to impact healthcare significantly.

3.1.1.1 Realization of Electronic Medical Records Linking

3.1.1.1.1 Non-European International Initiatives

An important international effort in this direction is a relatively recent initiative of CDISC[29]. CDISC open data standards have been developed collaboratively by global volunteers to improve the quality, efficiency and cost effectiveness of clinical research processes from protocol through analysis and reporting.

In 2008, a pilot project was launched between CDISC, CRIX, Pfizer, Brigham and Women's Hospital, Partners Healthcare and Harvard Medical School, entitled ASTER, or "ADE (adverse drug events) Spontaneous Triggered Event Reporting". This project sought to enable automated ADE collection through the EHR using CDISC and IHE's Retrieve Form for Data Capture[30]. The application that was developed was a novel concept – it directly downloaded data held in the EHR and allowed direct submission to the FDA, all in the correct format for the electronic reporting of individual case safety reports.

The results were impressive. In the three-month timeframe that the pilot project was in progress over 200 reports were sent to the FDA, and it was found that the time to fill out a report was reduced from 34 minutes average to less a 1 minute per patient. 91% of the

---

[26] Pakhomov, S., Weston, S. A., Jacobsen, S. J., Chute, C. G., Meverden, R., & Roger, V. L. (2007). Electronic medical records for clinical research: application to the identification of heart failure. The American journal of managed care,13(6 Part 1), 281-288.

[27] Liao, K. P., Cai, T., Gainer, V., Goryachev, S., Zeng-treitler, Q., Raychaudhuri, S., ... & Karlson, E. W. (2010). Electronic medical records for discovery research in rheumatoid arthritis. Arthritis care & research, 62(8), 1120-1127.

[28] Kho, A. N., Pacheco, J. A., Peissig, P. L., Rasmussen, L., Newton, K. M., Weston, N., ... & Kullo, I. J. (2011). Electronic medical records for genetic research: results of the eMERGE consortium. Science translational medicine,3(79), 79re1-79re1.

[29] http://www.cdisc.org/standards-and-implementations

[30] http://wiki.ihe.net/index.php?title=Retrieve_Form_for_Data_Capture

physicians involved had not even submitted ADE reports the prior year, and of those 200 reports that were filed, it was found that 20% of these ADEs were deemed serious.

More recently, there are efforts going on around the globe to increase the use of EHRs for research, including the Innovative Medicines Initiative's EHR4CR project (http://www.ehr4cr.eu/) and several initiatives in Japan, using RFD and CDISC standards. In 2013, the U.S. Department of Health and Human Services Office for the National Coordinator (ONC) of Health IT launched a new initiative, called Structured Data Capture (SDC). This initiative seeks to address the current, limited use of EHR data outside of direct patient care due to "*a lack of uniformity in the terminology and definition of data elements across EHRs*". This is a first step towards Meaningful Use 3 to achieve a Learning Health System. The ONC team has specifically stated that CDISC Retrieve Form for Data Capture (RFD) and CDISC CDASH are to be leveraged to ensure that there is no duplication of efforts. More information about this initiative can be found on the ONC S&I Framework Wiki[31].

The CDISC Healthcare Link Initiative

The CDISC Healthcare Link Initiative is one of the most rapidly developing areas of work being conducted for addressing the technical challenges to enable seamless linking to electronic health record systems. Leveraging standards to improve the methods by which investigative sites can conduct medical research and capture data for clinical research studies is vital for a number of reasons; one very important reason is that clinicians frequently do one research study and no more due to the unwieldy nature of clinical research processes today. The increasing presence of an electronic health record (EHR) at healthcare sites opens new opportunities to integrate the processes of clinical care and clinical research. This will, in turn, expand the capacity for research and increase patient participation.

The overarching goals for CDISC Healthcare Link have been to:

a) Make it easier for physicians to conduct clinical research,
b) Collect data only once in an industry standard format for multiple downstream uses
c) Improve data quality and patient safety.

The Initiative has taken steps to ensure that the link between healthcare and other secondary uses, including research, takes into account existing regulations, privacy and security concerns, and current practices to provide practical pathways to achieve the vision through a stepwise approach.

These enablers were developed in conjunctions with EHR vendors, and respect the limited amount of resource that these vendors can devote to a problem that is secondary to their main concern. These enablers are available now and have already proven to significantly

[31] http://wiki.siframework.org/Structured+Data+Capture+Initiative.

decrease the time and effort to provide data for certain use cases, such as safety reporting, using EHRs.

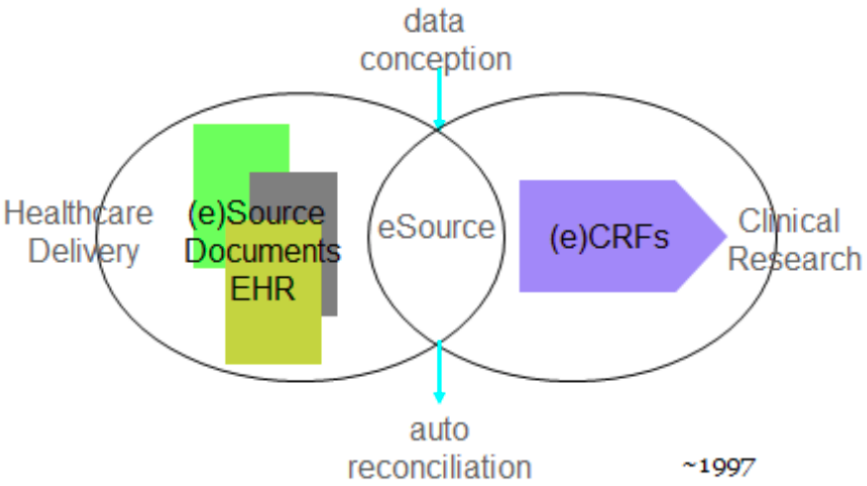


Figure 3-1: Healthcare Link – Optimizing healthcare through secondary use of healthcare data for research or other purposes (picture taken form: http://www.cdisc.org/healthcare-link)

An electronic data capture (EDC) system requires copying and re-entering data from the EHR. One would think, therefore, that re-use of EHR data must be significantly better. However, much of the valuable data in electronic health records is unable to be accessed due to a disconnect between the clinical research and clinical care domains – a disconnect caused by the use of different standards and terminology systems[32].

Existing Standards and Enablers for CDISC Healthcare Link

Well over a decade's worth of work has contributed to a set of standards and enablers to facilitate using EHRs to conduct clinical research (regulated or not). Specifically, these include the following.

1. Electronic Source Data Interchange (eSDI) Document
2. CDISC's Clinical Data Acquisition Standards Harmonization (CDASH) Standard
3. Biomedical Research Integrated Domain Group (BRIDG) Model
4. Interoperability Specifications
5. Integration Profiles

Each of these is briefly described in the following sections; however, there is far more information available through the links and references provided.

[32] Laleci, Gokce, Mustafa Yuksel and Asuman Dogac. "Providing Semantic Interoperability between Clinical Care and Clinical Research Domains", IEEE transactions on Information Technology in Biomedicine, 17(2), September 2012

1. **Electronic Source Data Interchange (eSDI) Document.** The eSource Data Interchange Initiative began in 2004 when the FDA requested that CDISC assist by forming a team to explore the further use of new technologies for research in the context of the existing regulations (i.e. 21CFR11, Good Clinical Practices and Guidances related to electronic source (eSource) documentation, eSource referring to entering the data electronically initially. Global regulations were evaluated and analysed and a set of 12 requirements for processes to support eSource and still meet regulatory requirements was developed. The entire document can be found at: http://www.cdisc.org/esdi-document. The eSDI initiative provided the foundation and basis for the development of the CDISC IHE Retrieve Form for Data Capture (RFD) integration profile.
2. **Clinical Data Acquisition Standards Harmonization (CDASH) Standard.** The rationale for developing CDASH came from the FDA's "Critical Path Initiative: Innovation or Stagnation" article33. The development of CDASH was a global collaborative project that resulted in a minimal core standard dataset that is common across research studies. There were 18 domains core domains (including Medical History (MH), Adverse Events (AE), Concomitant Medication (CM), Demographics (DM), Subject Characteristics (SC), Inclusion/Exclusion (IE), Substance Use (SU), Vital Signs (VS), Disposition (DS), Drug Accountability (DA), Exposure (EX), Protocol Deviations (PD), Comments (CO), Lab (LB), ECG (EG). To these, it is possible to add domains specific to therapeutic areas. In terms of Healthcare Link, CDASH provides a logical target for an EHR to map to a defined set of research standards. One of the former barriers to using EHRs for secondary use (i.e. research) was that each study sponsor expected a unique set of data in a proprietary format, thus CDASH has been instrumental in removing this barrier.
3. **Biomedical Research Integrated Domain Group (BRIDG) Model.** The Biomedical Research Integrated Domain Group (BRIDG) Model[34] was developed for two primary reasons: 1) to ensure that all of the CDISC foundational standards are harmonized among each other (in a model that is can be understood by those who are involved in clinical research), and 2) to provide a bridge from research standards to healthcare standards. BRIDG effectively enables information system interoperability by allowing the CDISC standards to work together, as well as ensuring that developers can develop applications that will work with the CDISC standards. BRIDG is a UML model that was developed collaboratively through CDISC, HL7, FDA and NCI.

---

[33] "Innovation or Stagnation: Challenge and Opportunity on the Critical Path to New Medical Products."
http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/CriticalPathOpportunitiesReports/ucm077262.htm.

[34] The BRIDGE model is openly available at www.bridgmodel.org

4. **Interoperability specifications and Integration Profiles.** As part of Healthcare Link, CDISC has been working closely with Integrating the Healthcare Enterprise (IHE), to develop an entire set of integration profiles to support the realization of the Healthcare Link goals.
   "IHE is an initiative by healthcare professionals and industry to improve the way computer systems in healthcare share information. IHE promotes the coordinates use of established standards such as DICOM and HL7 to address specific clinical need in support of optimal patient care. Systems developed in accordance with IHE communicate with one another better, are easier to implement, and enable care providers to use information more effectively."[35]
   While IHE is not a typical standards organization, it takes a problem-focused approach, using existing standards, and integrating them into a targeted specification called an Integration Profile. These profiles are extensively tested and demonstrated before their release.
   IHE has a number of domains, with the most relevant from the Healthcare Link perspective, is the Quality, Research, and Public Health (QRPH) domain. This domain focuses on the re-use of EHR data for external purposes.
   It is obvious that relevant developments and integration profiles are very important to the work and objectives of MHA and should be taken into consideration.

3.1.1.1.2 Relevant European initiatives

Healthcare systems are increasingly dependent on information and communication to deliver high-quality care to European patients. It is a fact that today we have a plethora of eHealth tools offering services and revolutionizing the health care sector[36]. It is also well known that during the last decades, national and regional health authorities, hospitals, and/or doctors have selected and implemented their own individual systems for health information management.

Promoting eHealth interoperability, ensuring all different, diverse systems are able to communicate with each other, is a top priority for the European Commission in order to offer significant benefits in healthcare across the Europe. The objective is to overcome the fragmentation of the eHealth market in Europe by promoting the use of existing EU standards and fostering the adoption of interoperable solutions, to ensure different eHealth systems can exchange information seamlessly within and between countries. EU policies around eHealth interoperability address the four layers of interoperability: technical, organizational, legal and

[35] IHE Website: http://www.ihe.net
[36] eHealth: Extending, Enhancing, and Evolving Health Care, Carlos A. Meier, Maria C. Fitzgerald, and Joseph M. Smith Annual Review of Biomedical Engineering, Vol. 15: 359 -382 (Volume publication date July 2013)

semantic. This action will contribute to the development of an appropriate EU-wide framework for standards, interoperability testing, and certification of eHealth systems. It should help elaborate a joint vision on interoperability architecture related to eHealth and provide guidance on architecture domains where EU Member States share a common interest.

The main objectives of cross border interoperability are to ensure that healthcare information systems across EU are able to communicate with each other and to contribute to the negotiation and development of a *Framework* which will define a set of standards, profiles and procedures relevant to the electronic provision of healthcare services.

The most important expected benefits are to improve coordination efforts in order to implement national and international such infrastructures able to provide support to Member States and relevant players to deploy/receive interoperable, cross-border health services in a sustainable way and enhance patient safety and continuity of care across Europe.

In this respect there are a number of related actions that already have been developed by existing EU Commission-funded eHealth projects some of which we present briefly in the following section:

- **epSOS**: epSOS[37] aimed to design, build and evaluate a service infrastructure that demonstrates cross-border interoperability between electronic health record systems in Europe. The epSOS Patient Summary is a standardized set of basic medical data that includes the most important clinical facts required to ensure safe and secure healthcare. This summarized version of the patient's medical data gives health professionals the essential information they need to provide care in the case of an unexpected or unscheduled medical situation (e.g. emergency or accident). Though this data is mainly intended to aid health professionals in providing unscheduled care, it can also be used to provide planned medical care (e.g. in the case of citizen movements or cross-organizational care paths).

  The epSOS Patient Summary[38] contains the following data:
  - General information about the patient (e.g. name, birth date, gender).

[37] http://www.epsos.eu/

[38] http://ec.europa.eu/information_society/newsroom/cf/document.cfm?action=display&doc_id=723

- A medical summary consisting of the most important clinical patient data (e.g. allergies, current medical problems, medical implants, or major surgical procedures during the last six months).
  - A list of the current medication including all prescribed medicines that the patient is currently taking.
  - Information about the Patient Summary itself e.g. when and by whom the Patient Summary was generated or updated. This data is also used for protocol and security purposes.

MHA selected this aforementioned standardization set, to be the main building block for exchanging clinical related information between the platform and a clinical information system. In section 4.3 we present in detail the efforts for creating a physical link between MHA platform and a Hospital information system.

- **EURECA:** The goal of EURECA[39] project (Enabling information re-Use by linking clinical REsearch and Care) was to enable seamless, secure, scalable and consistent linkage of healthcare information residing in electronic health record (EHR) systems with information in clinical research information systems, such as clinical trials. Achieving semantic interoperability among EHR and clinical trial systems is at the core of the EURECA project, as it is the basis for enabling many of the software services and tools developed in the project. The EURECA project focused on the definition of a standards based interface layer towards external systems. The goal of the interface layer is to hide the heterogeneity of external EHR, CDW or CDMS data sources from the EURECA services. The interface layer has been defined as a set of abstracted interfaces (e.g. providing functional access through standard EURECA APIs, providing direct data access, etc.).
- **CALLIOPE**: The CALLIOPE[40] Network is part of the Open eHealth Initiative, which is driven by Member States health administrations. It represents a targeted effort aiming to establish an appropriately governed, composed and structured open forum, with the focal goal to support Member States to implement interoperable eHealth solutions, in close collaboration with the key stakeholders, including users, industry and payers. It will offer a portfolio of targeted support services to be provided on request. The primary goal of

[39] http://eurecaproject.eu/
[40] http://www.calliope-network.eu/

Calliope is to contribute to the development of eHealth Interoperability providing a set of guidelines that aim to achieve the interoperability of electronic health records at local, regional, national, and cross-border levels.

- **Trillium Bridge**: Bridging Patient Summaries across the Atlantic. The Trillium Bridge[41] support action extends the European Patient Summaries and Meaningful Use II, Transitions of Care in the United States to establish an interoperability bridge that will benefit EU and US citizens alike, advancing eHealth innovation and contributing to the triple win: quality care, sustainability and economic growth.
- **HITCH**: HITCH[42] is about developing a vision of how interoperability and conformance testing of eHealth systems should be organized in Europe and beyond. This ranges from the analysis of eHealth testing tools, over quality management in interoperability testing, to complete certification and quality labeling scenarios. HITCH defines the European Commission's roadmap on eHealth interoperability testing and therefore will provide recommendations to institutions and authorities interested in establishing cross-vendor interoperability testing events but also to vendors that like to implement an in-house interoperability testing Quality Management System. Additionally, HITCH provides a vision on how a future eHealth quality labelling or certification could look like in Europe. The quality of the roadmap should be guaranteed by the HITCH partners that are already deeply involved in those topics.
- **eHR-Q^{TN}**: EHR-QTN[43] is a Thematic Network project that prepares the health community across Europe for systematic and comparable quality assurance and certification of e-Health products, more specifically of the Electronic Healthcare Record systems. The project promotes certification by organizing national workshops in 27 different European countries, by validating the EuroRec functional statements (over 1.400 statements), translating a substantial set of them in over 20 different European languages and by validating the EuroRec certification tools and certification procedures. The focus functionalities of the validation and the translations to be addressed during the project will be on medicinal product prescriptions, on medication management, on summary

[41] http://www.trilliumbridge.eu/
[42] http://www.hitch-project.eu/
[43] http://ehrqtn.eurorec.org/

records as well as on generic statements regarding reliability and trustworthiness of the systems and on security and access management.

- **NetC@rds**: NETC@RDS44 aims at achieving initial deployment of an online service for the electronic European Health Insurance Card (eEHIC). The central positioning of the project is to serve as an experimental test bed for the electronification of the EHIC. The established online verification infrastructure for cards or entitlement rights is the first real trans-European interconnection of different Member States in the eHealth sector. The ultimate goal of NETC@RDS is to achieve full integration with the existing and emerging national/regional infrastructures for eHealth and eID through the initial deployment of common administrative dataset and process supporting the use of electronic health and insurance cards.
- **Antilope:** ANTILOPE[45] project ambition was to drive eHealth interoperability in Europe and beyond. The challenge for Antilope was to define a comprehensive, usable framework that enables the development of a unified market and improves the quality of the projects and solutions in eHealth. Antilope has given regional, national and international projects practical guidelines to converge their eHealth platforms and practices. These guidance documents are available here: (https://www.antilope-project.eu/resources/).
- **SemanticHealthNet:** SemanticHealthNet[46] will develop a scalable and sustainable pan-European organisational and governance process for the semantic interoperability of clinical and biomedical knowledge, to help ensure that EHR systems are optimized for patient care, public health and clinical research across healthcare systems and institutions. Through a clinically-driven workplan, exemplified in cardiovascular medicine, SemanticHealthNet will capture the needs for evidence-based, patient-centred integrated care and for public health, encapsulating existing European consensus in the management of chronic heart failure and cardiovascular prevention. Experts in EHR architectures, clinical data structures, terminologies and ontology will combine, tailor and pilot their best-of-breed resources in response to the needs articulated by clinicians and public health physicians.

[44] http://www.netcards-project.com/
[45] https://www.antilope-project.eu/
[46] http://www.semantichealthnet.eu/

- **SemanticHEALTH**: The purpose of SemanticHEALTH[47] EU funded project is to describe a short and medium term Research and Deployment Roadmap for Semantic Interoperability in e-health. It started by defining 4 levels and 3 dimensions for Semantic Interoperability. The vision is to reconcile the needs for the direct patient care safety, biomedical and clinical research and for public health by the reuse of direct care data: from gene to individuals and populations. The methodology is presented and preliminary results and milestones for the short and the long term are set. We conclude by statements on the main characteristics and needs of the roadmap to sustain better health for individual and populations in the changing EU health care systems.

Also several other EU funded projects, with more horizontal and generic objectives, whose work, nevertheless, are important for the scope of MHA are :

- **STORK:** The aim of the STORK[48] project is to establish a European eID Interoperability Platform that will allow citizens to establish new e-relations across borders, just by presenting their national eID. Cross-border user authentication for such e-relations will be applied and tested by the project by means of five pilot projects that will use existing government services in EU Member States. In time however, additional service providers will also become connected to the platform thereby increasing the number of cross-border services available to European users. Thus in the future, you should be able to start a company, get your tax refund, or obtain your university papers without physical presence; all you will need to access these services is to enter your personal data using your national eID, and the STORK platform will obtain the required guarantee (authentication) from your government. The role of the STORK platform is to identify a user who is in a session with a service provider, and to send his data to this service. Whilst the service provider may request various data items, the user always controls the data to be sent. The explicit consent of the owner of the data, the user, is always required before his data can be sent to the service provider. The platform will not store any personal data, so no data can be lost. This user centric approach was not taken to satisfy some philosophical preferences, but in line with the legislative requirements of all the various

[47]Lewalle P1, Rodrigues JM, Zanstra P, Ustun B, Kalra D, Surjan G, Rector A, Stroetmann V, Virtanen M., A deployment and Research Roadmap for Semantic Interoperability: the EU semantic health project, Stud Health Technol Inform. 2008;136:635-40, http://ebooks.iospress.nl/publication/11653

[48] https://www.eid-stork.eu/

countries involved that oblige concrete measures to be taken to guarantee that a citizen's fundamental rights, such as his privacy, are respected.

- **e-SENS:** e-SENS (Electronic Simple European Networked Services)[49] e-SENS is a large-scale project that embodies the idea of European Digital Market through innovative ICT solutions. It faces technical and legal challenges by providing solutions for seamless public service delivery across borders. e-SENS consolidates, improves and extends existing technical solutions to develop a coherent and sustainable European Interoperability Architecture. This will in turn affect the quality of public services in the EU making them easily accessible across borders. e-SENS covers different aspects of ICT applied in a number of cross-border cases in domains such as e-Health, e-Justice, e-Procurement and business setup. It enables actual transactions between business/citizens and public administration to test IT components in real environment. The e-SENS solutions based on standards and technical specifications can be combined with each other and integrated with sector-specific applications. The goal of generic and reusable components is to enable any information systems to offer services across national borders and sectors to businesses and citizen. The objective of e-SENS is to make available a comprehensive set of building blocks (BBs) for an interoperable European infrastructure for cross sector services. The project focuses on providing architecture driven solutions and technical specification on the state of the art technologies by engaging in close corporation various domain communities. The goal is to create general purpose components that can be extended to other various policy areas. FORTH-ICS is a participating consortium member of e-SENS.

3.1.2 Linking to Biomedical Data Repositories

Biomedical data repository is a repository that contains data about medical and biological specimens to support future scientific research. In this report, we study biomedical data repositories that contain information regarding human bio-samples (biobanks). In this section, we will describe case studies that support medical research through secondary use of information in biomedical data repositories.

[49] http://www.esens.eu/

The study of Reed et. al.[50] supported by the Oslo University Hospital in collaboration with Akershus University Hospital in Norway describe a data warehouse[51] infrastructure that links a data repository system to the hospitals' laboratory and administrative systems in order to achieve centralized collection of clinical and biomedical data for research purposes. The data warehouse infrastructure supports integration of information from different sources and provides a secure environment for data protection in accordance to Norwegian legislations. Access to data is provided within a legal context and given patients' consent. The data warehouse architecture is also adopted in the study of Chute et. al.[52] that supports integration of heterogeneous medical data derived from linked biomedical repositories and clinical primary care databases at the Mayo Clinic, USA, in order to facilitate biomedical research.

Bowton et.al.[53], describe an infrastructure that links electronic medical records to biobanks, to facilitate pharmaco-genomic research. This infrastructure enables case-control pharmacogenomics studies by utilizing EMRs information for case subjects and biobanks information for control subjects, achieving thus substantial costs reduction in the design of the studies. Similar studies[54,55] study the linking of clinical and biological database systems, in order to enhance cancer clinical trial design. Their methodology is based on ontological semantic technologies[56] to integrate effectively the heterogeneous information provided from the health data sources. These studies have shown that linking to biomedical repositories has the potential to advance medical research significantly. Integration of clinical

[50] Reed, W., Jor, S., & Bjugn, R. (2012). How can clinical biobanks and patient information be adapted for research–Establishing a hospital based data warehouse solution. Norsk epidemiologi, 21(2).
[51] Devlin, B., & Cote, L. D. (1996). Data warehouse: from architecture to implementation. Addison-Wesley Longman Publishing Co., Inc..
[52] Chute, C. G., Beck, S. A., Fisk, T. B., & Mohr, D. N. (2010). The Enterprise Data Trust at Mayo Clinic: a semantically integrated warehouse of biomedical data. Journal of the American Medical Informatics Association, 17(2), 131-135.
[53] Bowton, E., Field, J. R., Wang, S., Schildcrout, J. S., Van Driest, S. L., Delaney, J. T., ... & Karnes, J. H. (2014). Biobanks and electronic medical records: enabling cost-effective research. Science translational medicine,6(234), 234cm3-234cm3.
[54] Maojo, V., García-Remesal, M., Billhardt, H., Alonso-Calvo, R., Pérez-Rey, D., & Martín-Sánchez, F. (2006). Designing new methodologies for integrating biomedical information in clinical trials. Methods Inf Med, 45(2), 180-185.
[55] Tsiknakis, M., Brochhausen, M., Nabrzyski, J., Pucacki, J., Sfakianakis, S. G., Potamias, G., ... & Kafetzopoulos, D. (2008). A semantic grid infrastructure enabling integrated access and analysis of multilevel biomedical data in support of postgenomic clinical trials on cancer. Information Technology in Biomedicine, IEEE Transactions on, 12(2), 205-217.
[56] Bodenreider, O. (2008). Biomedical ontologies in action: role in knowledge management, data integration and decision support. Yearbook of medical informatics, 67.

and biological information can be utilized for various research purposes (pharmacogenomics research, clinical trial research etc.) to improve the quality of healthcare.

3.1.3 Linking to Personal Health Records

Personal health record (PHR)[57] are essentially general-purpose health information systems that are managed by people rather than medical personnel and healthcare providers. The aim of PHRs is to allow people to manage and collect information about their health state. The type of information that is managed by a PHR may range from simple entries of personal health observations, everyday exercise notes and diet habits to complete medical histories including medical exams and clinical diagnoses[58]. A PHR system is able to enhance the quality of healthcare by empowering individuals to maintain their own health record, supporting thus the self-management of their healthcare[59].

Several case studies present the benefits of linking hospital EMR systems to PHR systems[60,61,62,63]. Firstly, clinical information can flow easily between the systems based on their linkage. Patient data, that represent the patient's interaction with the hospital, can be loaded easily from the hospital system to the PHR, which is otherwise a time-consuming process that should be managed by the patient. Hospital laboratory exam results can be provided from the EMR to the PHR automatically. In addition, communications with clinicians,

[57] Kim, M. I., & Johnson, K. B. (2002). Personal health records. Journal of the American Medical Informatics Association, 9(2), 171-180.
[58] Genitsaridi, I., Kondylakis, H., Koumakis, L., Marias, K., & Tsiknakis, M. (2013). Evaluation of personal health record systems through the lenses of EC research projects. Computers in biology and medicine.
[59] Russell-Minda, E., Jutai, J., Speechley, M., Bradley, K., Chudyk, A., & Petrella, R. (2009). Health technologies for monitoring and managing diabetes: a systematic review. Journal of diabetes science and technology, 3(6), 1460-1471.
[60] Halamka, J. D., Mandl, K. D., & Tang, P. C. (2008). Early experiences with personal health records. Journal of the American Medical Informatics Association, 15(1), 1-7.
[61] Tang, P. C., Black, W., Buchanan, J., Young, C. Y., Hooper, D., Lane, S. R., ... & Turnbull, J. R. (2003). PAMFOnline: integrating EHealth with an electronic medical record system. In AMIA Annual Symposium Proceedings (Vol. 2003, p. 644). American Medical Informatics Association.
[62] Sands, D. Z., & Halamka, J. D. (2004). PatientSite: patient-centered communication, services, and access to information. In Consumer Informatics(pp. 20-32). Springer New York.
[63] Genitsaridi, I., Kondylakis, H., Koumakis, L., Marias, K., & Tsiknakis, M. (2013). Towards Intelligent Personal Health Record Systems: Review, Criteria and Extensions. Procedia Computer Science, 21, 327-334.

that are required in the patient's healthcare, can be facilitated. PHRs such as the Eredbook[64,65] are able to support processes of appointment scheduling and prescription renewal which require interaction with the hospital EMR system. Linkage of PHRs to EMRs has also been presented to support the management of patient-clinician communications for type 2 diabetes mellitus[66].

Clinical decision making can also be enhanced by the linkage of the hospital EMRs to the patient's PHRs[67]. Firstly, the EMR is able to retrieve from the linked PHR the patient's health history. The health history of the patient includes his clinical visits to the various healthcare facilities, his personal health observations and everyday life activities notes. Consequently, the clinician will be able to access the health history from the EMR which can assist him in the various clinical decisions that are part of the patient's healthcare process. This new source of information can be proved invaluable in the process of diagnosing and treating the patient effectively. Disease symptoms and treatment adverse reactions may be identified early from the patient's personal health observations. In addition, the clinician can monitor the patient's everyday life activities and motivate him towards appropriate lifestyle changes, improving thus the quality of patient healthcare.

PHRs can also provide a new source of information to facilitate medical research[68]. Medical systems can be linked to PHRs to access research data. Factors related to the etiology or the development rate of diseases can be researched on the patient's everyday life activities and living environment. Predictors of diseases can be identified from the integrative analysis of multiple patients' PHR data.

These studies have shown that linking to personal health records can advance medical research and improve the quality of healthcare significantly.

---

[64] O'Connor, S., Devlin, A. M., McGee-Lennon, M., Bouamrane, M. M., Browne, S., O'Donnell, C. A., & Mair, F. S. (2015). The Experiences of Health Visitors in Implementing a Digital Child Health Record, the Eredbook, in the United Kingdom: Some Challenges and Lessons Learned.
[65] Velázquez, J. E. G. eRedbook: Towards Comprehensible Security A Case Study and Redesign.
[66] Grant, R. W., Wald, J. S., Schnipper, J. L., Gandhi, T. K., Poon, E. G., Orav, E. J., ... & Middleton, B. (2008). Practice-linked online personal health records for type 2 diabetes mellitus: a randomized controlled trial. Archives of Internal Medicine, 168(16), 1776-1782.
[67] Tang, P. C., Ash, J. S., Bates, D. W., Overhage, J. M., & Sands, D. Z. (2006). Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. Journal of the American Medical Informatics Association,13(2), 121-126.
[68] Weitzman, E. R., Kaci, L., & Mandl, K. D. (2010). Sharing medical data for health research: the early personal health record experience. Journal of medical Internet research, 12(2).

3.1.3.1 Relevant European initiatives: The EURECA – INDIVO X case study

A novel approach for linking to aPHR system has been studied and implemented within the recently completed EU funded EURECA project. The technical details of this effort are presented in the following paragraphs. Indivo-X[69] is a personal health platform that tries to enable individuals to own and manage a complete, secure, digital copy of their health and wellness information. Indivo-X integrates health information across sites of care and over time. It is free, open-source and web-based and it uses open standards including those from the SMART[70] platforms project. A screenshot of the platform can be seen in the following figure. In section 4.4 we present the technical details of connecting Indivo-X with MHA platform.

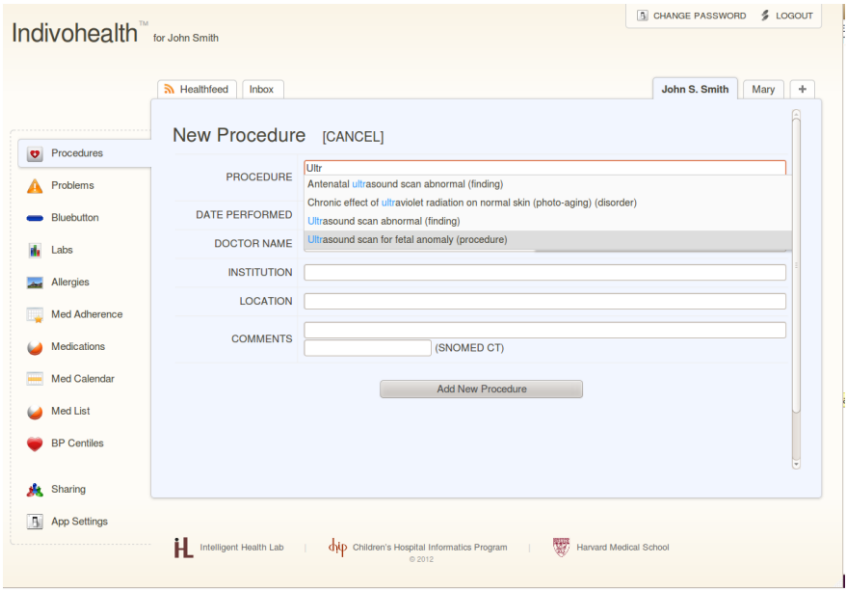


Figure 3-2: Indivo-X health information integration

Data export from IndivoX IndivoX already implements one mechanism for exporting Data, via the pull mechanism and within the EURECA project another mechanism has been implemented via pull mechanism.

Those two methods are explained below:

Push: The push mechanism allows a user to push his own data to the Eureca DWH. In order to achieve this, he has to select the appropriate app, and then to select the data that he/she would like to push to the DWH. The app that allows this functionality is shown in Figure 3-3.

[69] http://indivohealth.org/
[70] http://smartplatforms.org/

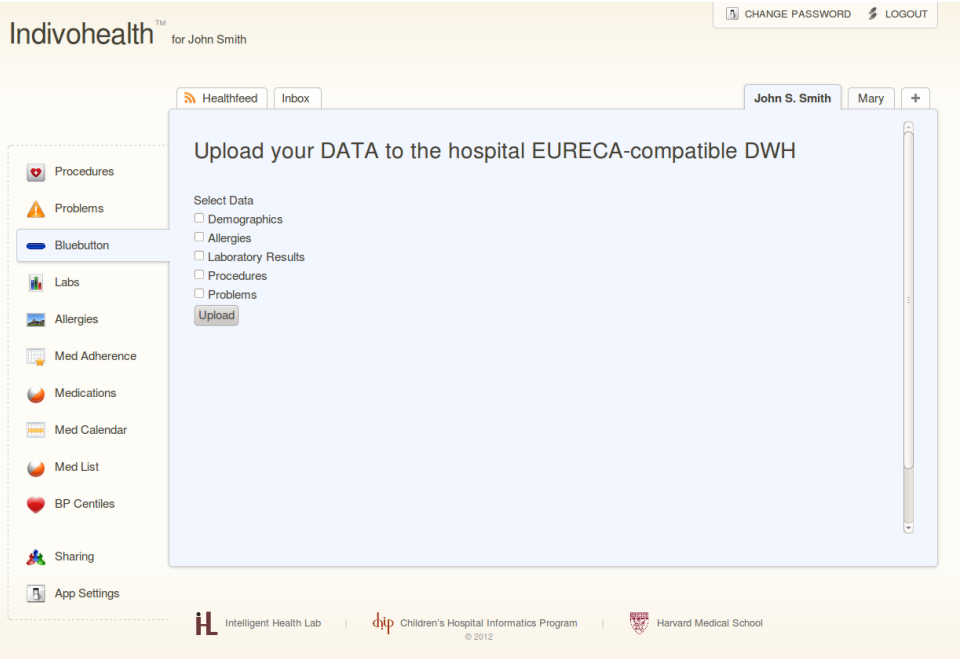


Figure 3-3: Pushing data to the Eureca DWH

Figure 3-4: Transformation of exported XML data to HL7 messages

To achieve that, initially the IndivoX API is used. However, the API allows an authenticated user to export data as JSON, RDF and XML data and the EURECA DWH imports messages as HL7 messages. So data exported from Indivo-X should be transformed in HL7-messages. This transformation is being done on-the-fly each time using a web service that has been created. A screenshot of some of the correspondences between the exported data and the HL7

messages are shown in the following Figure 3-4. After the transformation, another web service pushes the HL7 messages to the DWH.

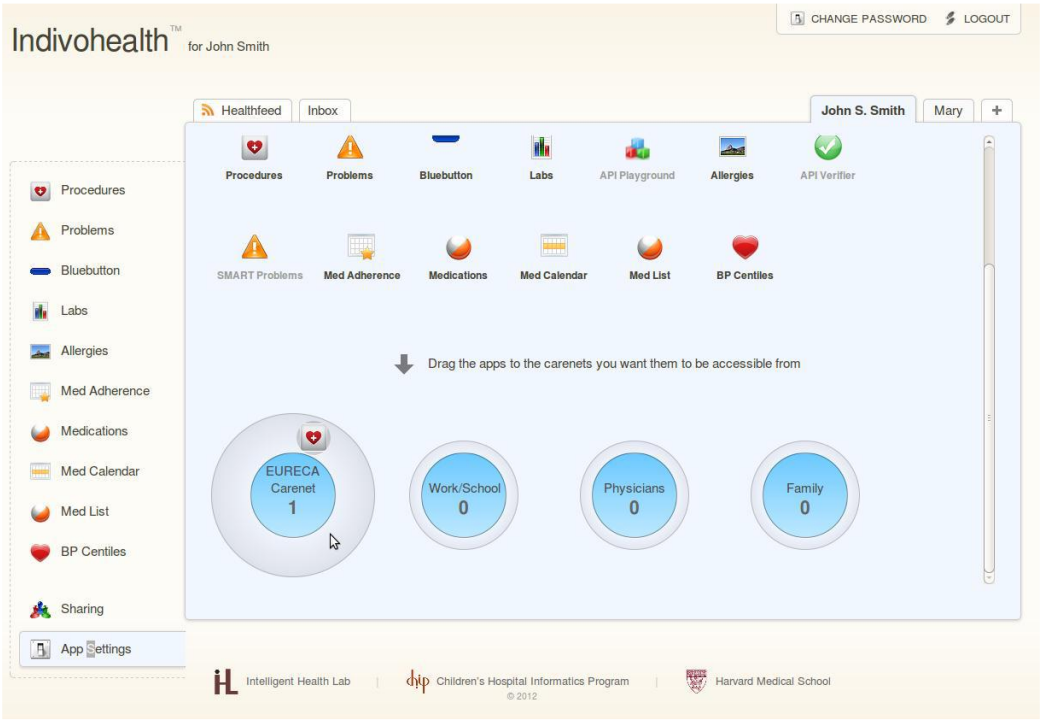


Figure 3-5: Authorising EURECA carenet to access patient

Pull: Besides pushing data to the DWH, Indivo-X allows a patient to share data with a specific carenet. The patient authenticates a Carenet, and then the specific Carenet can use the API calls to retrieve the XML, JSON or RDF data. A Carenet can be a doctor, a group of doctors, a research project or a team of people. A screenshot of this functionality is shown in the figure below. A patient can drag-and-drop the different apps to the individual carenet to allow data sharing with that carenet. Then the people in that carenet are authorized to query the patient data when using Indivo-X API calls.

3.1.4 Linking to Social Networks

A Social Network expresses a network structured of social relationships usually between individuals. A social relationship represents the social interaction in a pair of individuals which includes information sharing and communication. In recent years, numerous web-based applications have been developed to support the construction and management of social

networks, which are defined as Social Networking Services (SNS)[71]. In Social Networking Services, individuals are represented by internet user profiles.

Social Networking Services can be websites, such as Facebook[72] and Twitter[73] (that support social networks of people with common interests such as friends, family, colleagues and classmates), compatible with a variety of desktop and mobile web browsers. SNS can also be native mobile applications, such as WhatsApp[74] and Snapchat[75] (that support social messaging with mobile devices), compatible with a variety of mobile operating systems. In this report, for simplicity reasons, the term Social Networks refers to all types of Social Networking Services.

Secondary use of Social Networks' information has the potential to provide significant benefits to medical research, and improve the quality of healthcare. In the following sections we will describe the advantages of linkage to social networks through numerous case studies identified in the literature.

3.1.4.1 Social Networks for patients

Patient Social Networks have been developed to support patients with common symptoms, diseases, and treatments, such as the websites of PatientsLikeMe[76], CureTogether[77] and MedHelp[78]. The aim of these patient social networks is the enhancement of the communication between patients in order for them to gain psychological support from each other and become more informed on their conditions and available treatments. Secondary use of data in patient social networks has emerged as a new innovative area for medical research (research on Amyotrophic Lateral Sclerosis based on PatientsLikeMe patient data)[79]. Patient Social Networks can serve as high volume information databases that provide cross-regional patient data. Linkage of research medical systems to patient social networks can

[71] Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. Journal of Computer-Mediated Communication, 13(1), 210-230.
[72] https://www.facebook.com/
[73] https://twitter.com/
[74] WhatsApp, http://www.whatsapp.com/
[75] https://www.snapchat.com/
[76] https://www.patientslikeme.com/
[77] http://curetogether.com/
[78] http://www.medhelp.org/
[79] Swan, M. (2009). Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. International journal of environmental research and public health, 6(2), 492-525.

provide new raw material for analysis[80]. Some patient social networks such as PatientsLikeMe utilize research services that enable the identification of people with similar conditions based not only on the disease but also on the symptoms and treatments of the patients, in order to support their communication. Patient Social networks have a twofold contribution to medical research:

- The provision of publicly available patient data.
- The ability to interact with patient groups to collect new data.

A few patient social networks such as PatientsLikeMe support the longitudinal management of the patients' health state by providing services of disease self-tracking and visual representations of disease activity, behaving thus much like Personal Health Records[81]. PHRs that are managed by patients and EMRs that are managed by clinicians would benefit from linking to Patient Social Networks since this linkage would result in a more comprehensive collection of patient data regarding their health state.

Some patient social networks also support patient enrolment in drug trial studies conducted by clinicians and pharmaceutical companies. Clinical trial management systems can be linked to patient social networks in order to facilitate drug trials. This linkage will enable the recruiting of people for the evaluation of drug effectiveness[82,83].

[80] Pearson, J. F., Brownstein, C. A., & Brownstein, J. S. (2011). Potential for electronic health records and online social networking to redefine medical research. Clinical chemistry, 57(2), 196-204.
[81] Wicks, P., Massagli, M., Frost, J., Brownstein, C., Okun, S., Vaughan, T., … Heywood, J. (2010). Sharing Health Data for Better Outcomes on PatientsLikeMe. Journal of Medical Internet Research, 12(2), e19. http://doi.org/10.2196/jmir.1549
[82] Fenner, Y., Garland, S. M., Moore, E. E., Jayasinghe, Y., Fletcher, A., Tabrizi, S. N., ... & Wark, J. D. (2012). Web-based recruiting for health research using a social networking site: an exploratory study. Journal of Medical Internet Research, 14(1), e20.
[83] Swan, M. (2012). Crowdsourced Health Research Studies: An Important Emerging Complement to Clinical Trials in the Public Health Research Ecosystem. Journal of Medical Internet Research, 14(2), e46. http://doi.org/10.2196/jmir.1988

Physician-profile and Physician-rating websites[84,85,86,87] can also be linked to patient social networks to provide additional benefits to both physicians and patients by supporting their interaction for better outcomes. Patients will be able to find physicians of high reputation to support their healthcare, while physicians will be able to interact with patient communities in order to advertise their medical services.

3.1.4.2 General purpose Social Networks

While patient social networks connect individuals that are diagnosed with diseases, there are numerous other social networks that connect healthy individuals who share health-related information. Runkeeper[88,89] and fitbit[90] are applications with social networking support that are focused on exercise tracking. They provide information sharing and motivating fitness competitions with friends and family. Other applications such as WeightLossBuddy[91], FatSecret[92] and EatingWell[93] provide social networking support focused on diet information tracking and sharing between individuals. Linkage to lifestyle social networks that manage health-related information can facilitate future medical research and enhance quality of healthcare.

[84] Lagu, T., Hannon, N. S., Rothberg, M. B., & Lindenauer, P. K. (2010). Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. Journal of general internal medicine, 25(9), 942-946.
[85] Emmert, M., Meier, F., Pisch, F., & Sander, U. (2013). Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. Journal of medical Internet research, 15(8).
[86] Kadry, B., Chu, L. F., Kadry, B., Gammas, D., & Macario, A. (2011). Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. Journal of medical Internet research, 13(4).
[87] Hanauer, D. A., Zheng, K., Singer, D. C., Gebremariam, A., & Davis, M. M. (2014). Public awareness, perception, and use of online physician rating sites.JAMA, 311(7), 734-735.
[88] https://runkeeper.com/
[89] Stragier, J., and Mechant, P. 2013. Mobile fitness apps: profiling RunKeeper users. In 2013 Annual meeting of the International Society for Behavioral Nutrition and Physical Activity, ISBNPA 2013.
[90] https://www.fitbit.com
[91] http://www.weightlossbuddy.com/
[92] http://www.fatsecret.com/
[93] http://www.eatingwell.com/

PHR and EMR systems that are linked to the lifestyle social networks of the patients, will be able to support the communications with clinicians[94,95] and also access information about the patients' everyday activities to provide a more comprehensive view of the patient's life and health state. This linkage supports a holistic approach to healthcare, since the lifestyle information that is provided from the social networks will be complementary to the development of more accurate patient health profiles that will guide the clinical care process.

Medical systems can also benefit significantly from linking to lifestyle social networks by utilizing the individuals' everyday activities and social interactions information in support of various research purposes such as research on diabetes[96], cardiovascular disease[97], dementia[98], infectious diseases[99] and depression analysis[100]. Thus, lifestyle social networks can be proven an invaluable source of health information to conduct medical research.

3.2 Linking health and health related data: Purposes and Sources

The previous review provided an insight on the advantages of secondary use of information in various health data sources such as EMRs, Biomedical repositories, PHRs and (Patient or Lifestyle) Social Networks. An efficient linkage of a recipient system to external health data sources is able to support secondary use of their integrated information. Below we present a summarization of the literature findings regarding the different types of data sources, the types of the exchanged information and the purposes of the linkage to the sources.

A. Electronic Medical Records used by Healthcare Professionals

Data Types

- Patient Demographics
- Patient Characteristics

[94] Shachak, A., & Jadad, A. R. (2010). Electronic health records in the age of social networks and global telecommunications. JAMA, 303(5), 452-453.
[95] Hawn, C. (2009). Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care. Health affairs,28(2), 361-368.
[96] Wing, R. R., Goldstein, M. G., Acton, K. J., Birch, L. L., Jakicic, J. M., Sallis, J. F., ... & Surwit, R. S. (2001). Behavioral science research in diabetes lifestyle changes related to obesity, eating behavior, and physical activity.Diabetes care, 24(1), 117-123.
[97] Kawachi, I., Colditz, G. A., Ascherio, A., Rimm, E. B., Giovannucci, E., Stampfer, M. J., & Willett, W. C. (1996). A prospective study of social networks in relation to total mortality and cardiovascular disease in men in the USA. Journal of epidemiology and community health, 50(3), 245-251.
[98] Fratiglioni, L., Paillard-Borg, S., & Winblad, B. (2004). An active and socially integrated lifestyle in late life might protect against dementia. The Lancet Neurology, 3(6), 343-353.
[99] Klovdahl, A. S., Potterat, J. J., Woodhouse, D. E., Muth, J. B., Muth, S. Q., & Darrow, W. W. (1994). Social networks and infectious disease: The Colorado Springs study. Social science & medicine, 38(1), 79-88.
[100] Mueller, D. P. (1980). Social networks: a promising direction for research on the relationship of the social environment to psychiatric disorder. Social Science & Medicine. Part A: Medical Psychology & Medical Sociology, 14(2), 147-161.

- Patient Phenotypic and Genetic Information
- Clinical Examinations (e.g. allergies, immunizations, blood pressure, radiographies etc.)
- Patient Diseases Diagnosis
- Patient Drug Treatments
- Patient Adverse Drug Reactions
- Patient Vital Signs
- Patient Treatment Outcomes
- Clinician Follow-Up Notes Clinician Tests Orders

Linkage Purposes

- Genetic Research on phenotypes for Dementia, Cataracts, Peripheral Arterial Disease, Type 2 Diabetes and Cardiac Conduction on Diagnosis through secondary use of EMRs data[28]
- Research on Drug Safety through secondary use of ADRs data in EMRs[24,25]
- Clinical Research on heart failure through secondary use of EMR information[26]
- Clinical research on Rheumatoid arthritis utilizing the EMR information[27]

B. Biomedical Data Repositories used by Researchers

Data Types

- Patient Disease Diagnosis
- Patient Bio-samples Information
- Genes information
- DNA and RNA sequences information
- Protein tests

Linkage Purposes

- Pharmaco-genomic research through linkage of biomedical data repositories to EMRs[53]
- Translational Research through linkage of biomedical data repositories to clinical administrative and laboratory data systems[50,52]
- Cancer Clinical Trial Management through linkage of clinical and biological databases[54, 55]

C. Personal Health Records used by Individuals

Data Types

- Patient Demographics
- Patient Characteristics
- Patient Health Notes
  - Nutrition Notes
  - Exercise Notes
  - Observations
- Unhealthy Habits
  - Smoking
  - Alcohol Consumption
- Diseases Diagnosis
- Clinical Examinations
  - e.g. allergies, immunizations, blood pressure, radiographies etc.
- Treatments Information
- Adverse Drug Reactions
- Vital Signs Self-Monitoring MeasurementsTreatment Outcomes

Linkage Purposes

- Clinician-Patient Communications enhancement through linkage of EMRs to PHRs[60,63]
    - Hospital Laboratory Exams Automatic Access
    - Hospital Appointment Scheduling[65]
    - Clinical Drug Prescription Renewals[65]
    - Type 2 Diabetes Mellitus Management[66]
- Clinical Decision Support enhancement through provision of PHR information to EMRs
  Medical Research based on secondary use of PHRs information[68]

D. Social Networks used by Individuals

a) Social Networks for patients

Data Types

- Patient Diseases
- Patient Symptoms
- Patient Treatments
- Patient Outcomes
- Patient Disease Activity Visual Representations
- Clinical Trials
- Similar Profile Patients
- Physician Information
- Patient Social Interactions

Linkage Purposes

- Medical research based on secondary use of information in Patient Social Networks
    - Research on Amyotrophic Lateral Sclerosis[79]
- Provision of additional patient health information to the PHR or EMR through linkage to Patient Social Networks
- Drug Trial Patient Recruitment through linkage of Clinical Trial Management Systems to Patient Social Networks [82, 83]
- Physician-Patient Communications Enhancement through linking of Physician-profile and
- Physician-rating websites to Patient Social Networks

b) General purpose Social Networks

Data Types

- Exercise Information
- Diet Information
- Social Interactions
- Everyday life activities

Linkage Purposes

- Medical Research
- Diabetes research
- Cardiovascular disease research
- Dementia research

- Infectious Diseases research
- Depression research

The previous analysis attempted to highlight the benefits of linkage to external health data sources through numerous case studies identified in the literature. It also tries to show the heterogeneity and diversity of data that MHA had to be able to support in order to link external data sources to the platform.

3.3 Core Requirements of an Efficient Linkage to Health and Health-related data of Sources

3.3.1 Data sources for accessing data from a legal point of view

The explanations above show that external data sources such as social networks, external data warehouses and hospital records can be important to feed the avatar with correct and updated data. However, from the legal point of view the different data sources should be addressed separately.

3.3.1.1 Personal health records

With regard to personal health records as data source a distinction should be made between two possible approaches to allow the patient to benefit from external data sources for his avatar.

The first approach is that the patient makes use of his right (under data protection law) to access his personal health record, and then he enters the data on his own accord in the platform. This two-step approach[101] bears the risk, though, that data that are stored in his personal avatar are not correct, complete and up-to-date because the patient could forget or misunderstand the accessed data and/or mis-enter it during the upload process. The latter risk is especially relevant if the patient receives the data in a non-digital format that requires the user to enter the data manually.

Furthermore, the patient may be put to significant effort if he is a current patient of the hospital or physician, as he will need to repeat the process on a frequent basis in order to keep the data in the platform up to date. Also with respect to data security and data integrity the two-step-approach is not recommendable: the MHA user does usually not have the technical expertise to enter the data in a secure way and could contaminate or corrupt the data inadvertently.

[101] Please see D11.3, p. 22 f. for details.

3.3.1.2 Hospital information systems and other external warehouses

For the above-mentioned reasons, a second approach appears advantageous. This would be to link directly to hospital information systems (HIS) and other external data warehouses, thus ensuring that data are complete, accurate and up-to-date.

However, the consortium is aware that creating such linkage to hospital information systems and other external warehouses also raises challenges, and that some hospitals may be unwilling to transfer their data to MHA. The reasons for this is that hospitals (and other external warehouses) are responsible for controlling their stored data, and decisions as to data processing, including the sharing of data (even where the patient requests this) remain matters at their discretion. It is likely, as a minimum, that MHA would need to satisfy the addressee hospital not only of its legitimacy and credentials, but also that it will employ security safeguards to protect the data equivalent to those used by the hospital. MHA should therefore present a set of choices to which the hospitals can respond according to the dictates of their specific policy. Here, data transmission, control, strict access control mechanisms to ensure that data access can only be by authorised persons are crucial.

Having all these challenges in mind LUH drafted a patient transfer request to hospital that can be used by the MHA user and a data transfer agreement between hospital and MHA.[102]

Another challenge is that the hospital must be sure that the MHA user requesting transfer of HIS data to his personal avatar is really the relevant hospital patient. Interoperability issues can arise, too.

Please see D11.3, p. 18 ff. and D11.4, p. 26 for further details.

3.3.1.3 Social networks

Social networks could contain especially lifestyle data that the user wants to upload to MHA.

Users of Facebook and Twitter often post daily information about nutrition (restaurant visits, self-cooked meals) their sporting activities, as well as information concerning their mood. These pieces of information could be relevant in the context of MyHealthAvatar, not only for the patient/citizen himself, but for other users, such as his physician, with whom he chooses to share it. Another benefit is that users of social networks are already used to posting such information on a frequent basis, in an automatic way with little perceived effort on their part. MHA could benefit from this by a direct access link to such social network platforms, meaning the user to enter the same data a second time in the MHA platform.

However, from the legal point of view linking data with social networks also raises legal issues.

[102] Please see Annexes 4 and 5 of D11.3, pp. 74 ff.

The most important point is that the MHA user should be informed, where he inputs data in the social network platform, with a view to it (also) being sourced by MyHealthAvatar, that MyHealthAvatar has no influence on the privacy policies and security framework of those third parties (please see D11.3, p. 29 f. for further details). Another threat to privacy comes up if users do not only wish to access the data from social networks, but also would like to link information from MHA to social networks such as Twitter and Facebook. Here the risk is that such third parties might make use of the data without the knowledge of the MHA user or even against his will. Also this point has to be highlighted in the consent form if MHA would have such a functionality. Clause VIII. Demo App Developers and Third Party of the current Privacy Policy points out that products from providers external to the project are subject to their own privacy rules and that MHA has no control over these.

Finally, it should be noted that, insofar as MHA were to espouse third party apps to users as a means of collecting and accessing their data, this could potentially trigger product liability issues in relation to apps from outside the EU. This may be so under Article 3 (1) Product Liability Directive if, in presenting the app and encouraging its take-up, MHA is deemed to have 'imported' the relevant app into the EU (please see D11.4, p. 37 ff.).

3.3.2 The Representation of Data: Health Information standards

One of the core processes of an efficient linkage to external health data sources is the acquisition of information from the sources in a specified representation format. This process is performed after establishing secure and legitimate access to the data that addresses privacy and ownership concerns. However, the data acquisition from heterogeneous sources will most likely involve systems that utilize different representation methodologies which will impede the integrative analysis of the information. In this direction, there have been numerous efforts, described in this section, that aim on standardizing the representation format of the data for various health information domains, in order to facilitate health information exchange and correct data interpretation.

Two data standards organizations that have achieved wide acceptance are the Regulated Clinical Research Technical Committee of Health Level Seven (HL7)[103] and the Clinical Data Interchange Standards Consortium (CDISC)[30], presented in section 3.1.1.1.1. HL7 is a not-for-profit organization that provides, among others, standards for knowledge representation and sharing regarding clinical and administrative data domains and standards of EHR functional requirements for clinical research. One of the core HL7 standards is the CDA (HL7 Version 3

[103]http://www.hl7.org/

Clinical Document Architecture) which is used for the representation of clinical reports in order to facilitate their exchange between care professionals and patients. A CDA is able to represent clinical documents such as an admission report, pathology report, and imaging report and discharge summary.

The ONC provided the comparative report "Interoperability Standards Advisory" (2016) on the best available standards to meet various health information interoperability needs[104]. The study compared standards with regard to the following criteria, the maturity of the standard, its adoption level in USA, its cost and availability. A relevant part of the study's findings on the best available health data terminology and syntax standards along with their adoption level is provided in the following table.

Health Data Representation Domain		Standard	Adoption level
Allergies	Allergic Reactions	SNOMED-CT	●●●●○
	Allergies Medications	RxNorm	●●●●○
Encounter Diagnosis	Medical Encounter Diagnosis	SNOMED-CT	●●●●○
		ICD-10-CM	●●●●○
	Dental Encounter Diagnosis	SNOMED-CT	●●●●○
Family Health History		SNOMED-CT	●●●○○
Functional Status/Disability		-	-
Gender Identity, Sex, and Sexual Orientation	Patient Gender Identity	SNOMED-CT	Unknown
	Patient Sex (at birth)	For Male and Female, HL7 Version 3 Value Set for Administrative Gender; For Unknown, HL7 Version 3 Null Flavor	●●●●○
	Patient-identified sexual orientation	SNOMED-CT	Unknown
Immunizations		HL7 Standard Code Set CVX—Clinical Vaccines Administered	●●●●●
Industry and Occupation		-	-
Lab tests		LOINC	●●●○○
Medications		RxNorm	●●●●●
Numerical References & Values		The Unified Code for Units of Measure	●●○○○
Patient Clinical "Problems" (i.e., conditions)		SNOMED-CT	●●●●●
Procedures	Medical Procedures	SNOMED-CT	●●●●●

[104] https://www.healthit.gov/sites/default/files/2016-interoperability-standards-advisory-final-508.pdf

	Dental Procedures		
Imaging (Diagnostics, interventions and procedures)		LOINC	●●○○○
Tobacco Use (Smoking Status)		SNOMED-CT	●●●●●
Unique Device Identification		HL7 Harmonization Pattern for Unique Device Identifiers	●○○○○
Vital Signs		LOINC	●●●●●
Admission, Discharge, and Transfer		HL7 2.5.1 (or later) ADT message	●●●●●
Care Plan		HL7 Clinical Document Architecture (CDA), Release 2.0, Final Edition	●●●●●
Clinical Decision Support		HL7 Implementation Guide: Clinical Decision Support Knowledge Artifact Implementation Guide, Release 1.3, Draft Standard for Trial Use.	Unknown
Family health history (clinical genomics)		HL7 Version 3 Standard: Clinical Genomics; Pedigree	●○○○○
Images		Digital Imaging and Communications in Medicine (DICOM)	●●●●●
Orders for electronic laboratory tests and their results		HL7 2.5.1	●●●●●
Patient Education Materials		HL7 Version 3 Standard: Context Aware Knowledge Retrieval Application. ("Infobutton"), Knowledge Request, Release 2.	●●●●○
Patient Preference/Consent		IHE Basic Patient Privacy Consents (BPPC)	●●○○○
Segmentation of sensitive information	HL7 Clinical Document Architecture (CDA), Release 2.0, Final Edition		●●●●●
Summary care record	HL7 Clinical Document Architecture (CDA), Release 2.0, Final Edition		●●●●●

There is thus, a significant standardization effort for the formal representation of health information. However there is still a lack of standards for representing knowledge in certain medical domains (gaps) while there are multiple alternative standards for others (overlaps)[105]. An efficient linkage to external health data sources should include policies to resolve issues related to gaps and overlaps in the data representation standards utilized by the sources.

3.3.3 The Quality of the Data

The *Data Quality* is defined as the ability of the information, which is provided from heterogeneous health data sources to a linked recipient system, to support a particular integrative information analysis. The data quality depends highly on the level of pre-processing that is required, in order for the information to be integrated and analyzed uniformly. Low data quality may constitute the integrative analysis impossible.

Quality assessment of information has been the subject of a large amount of research studies identified in the literature[106,107,108,109]. However, there is a lack of studies that address this problem in the domain of integrated health information from heterogeneous sources. In this report, we distinguish the following low quality health information categories, (a) semantically inconsistent data, (b) contextually inconsistent data, (c) missing data, (d) conflicting data and (e) erroneous data. In this section, we will describe these categories in detail, and present related studies that attempt to address the quality issues of information provided from heterogeneous health data sources.

3.3.3.1 Semantically Inconsistent Data

Semantically Inconsistent Data is defined as low quality health information that is provided to a recipient system but cannot be interpreted correctly due to discrepancies in concept definitions or syntactical structure. These data can be divided into two subcategories, terminologically and syntactically inconsistent data.

3.3.3.1.1 Terminologically Inconsistent Data

Terminologically Inconsistent Data is defined as low quality information which cannot be utilized due to inconsistent conceptual terminology. *Terminologically* inconsistent data

[105] Richesson, R. L., & Krischer, J. (2007). Data standards in clinical research: gaps, overlaps, challenges and future directions. *Journal of the American Medical Informatics Association*, *14*(6), 687-696.

[106] Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. Communications of the ACM, 40(5), 103-110.

[107] Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. Communications of the ACM, 45(4), 211-218.

[108] Eckerson, W. W. (2002). Data quality and the bottom line. TDWI Report, The Data Warehouse Institute.

[109] Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. Journal of management information systems, 5-33.

provided from health data sources to a linked recipient system, cannot be utilized by the recipient system due to the following causes:

- *Unknown (Health) Data*: Data are expressed in terms that cannot be identified. Unknown data that are provided from health data sources to a linked recipient system are expressed in terms that are unrecognizable to the vocabulary of the recipient system.
- *Misinterpreted (Health) Data*: Data are expressed in terms that are interpreted differently in relation to the source's initial terminology. Misinterpreted data that are provided from health data sources to a linked recipient system are expressed in recognizable terms to the vocabulary of the recipient system, but the source's terminology does not match the vocabulary definitions of the recipient system, leading thus to misinterpretations.

3.3.3.1.2 Syntactically Inconsistent Data

Syntactically Inconsistent (Health) Data is defined as low quality health information which cannot be interpreted due to inconsistent syntactical structure. Syntactically inconsistent data provided to a recipient system do not match the syntactical rules of the information representation language (IRL) utilized by the system.

Linkage to external health data sources should include an evaluation process of the terminological and syntactical consistency of the information that is received from the sources. HIMSS (Healthcare Information and Management Systems Society)[110,111] provided definitions of Syntactic and Semantic Interoperability to specify the need for semantic consistency in health information exchange performed in a systems network. These definitions are described below.

- *Syntactic (or Structural) Interoperability* is defined as the ability of two or more health information systems to exchange information in a recognizable and commonly acceptable syntactical structure.
- *Semantic Interoperability* is defined as the ability of two or more health information systems to exchange information that can be interpreted correctly in order to be used for analysis.

The definition of syntactic interoperability specifies that the exchanged information in a system network should not fall into the category of syntactically inconsistent data while the

[110] http://www.himss.org/
[111] HIMSS Definition on Interoperability April 5, 2013 [online]:
http://www.himss.org/library/interoperability-standards/what-is-interoperability

definition of semantic interoperability specifies that the exchanged information should not fall into the semantically (*terminologically* or syntactically) inconsistent data category. Essentially, semantic interoperability contains the requirement for syntactic interoperability in the systems network. Semantic interoperability in a HIS network can be achieved by systems' conformance to the same health data representation standards.

The study of Bicer et. al.[112] presents a framework that is able to resolve semantic inconsistency in health information that is provided from an institution's HIS to the linked recipient system of another institution. The study described the utilization of a mapping tool for the transformation of HL7 v2 information, provided by the first system, to HL7 v3 information managed by a linked recipient system. The framework provides a tool for mapping a source ontology to a target ontology, in order to be able to transform instances of the source to instances of the target. Essentially, the source ontology is the data representation model used by the first system and the target ontology is the representation model of the second.

Federal administrative policies have been developed in the USA including financial incentives to support Regional Health Information Organizations (RHIOs) that facilitate exchange of health information in a HIS network of a particular region in the country, and also support semantic interoperability in the network[113,114]. The US nation aims further to the development of a national health information network (NHIN) that will connect the RHIOs. The study of Overhage et.al.[115] analyzes the adoption of health data representation standards (such as ICD-9, SNOMED-CT and LOINC) by the RHIOs in US communities.

An efficient linkage of a recipient system to external health data sources should include policies to resolve semantic inconsistency (terminological and syntactical) in health

---

[112] Bicer, V., Laleci, G. B., Dogac, A., & Kabak, Y. (2005). Artemis message exchange framework: semantic interoperability of exchanged messages in the healthcare domain. ACM Sigmod Record, 34(3), 71-76.
[113] Adler-Milstein, J., Bates, D. W., & Jha, A. K. (2009). US Regional health information organizations: progress and challenges. Health Affairs, 28(2), 483-492.
[114] Adler-Milstein, J., McAfee, A. P., Bates, D. W., & Jha, A. K. (2008). The state of regional health information organizations: current activities and financing.Health Affairs, 27(1), w60-w69.
[115] Overhage, J. M., Evans, L., & Marchibroda, J. (2005). Communities' readiness for health information exchange: the National Landscape in 2004. Journal of the American Medical Informatics Association, 12(2), 107-112.

information provided from the external sources in order to be able to interpret the information correctly.

3.3.3.2 Contextually Inconsistent Data

Contextually Inconsistent Data is defined as low quality health information that does not conform to contextual data constraints defined for a specific analysis. Cases of contextual data constraints are the specification of inclusion and exclusion patient criteria and the requirement for unique patient identification in various research studies. The difference of contextual consistency with semantic consistency is that the second is concerned with the correct interpretation of the information while the first focuses on the ability of the correctly interpreted information to match the context of a specific analysis.

The study of Tuszynski et. al[116] presented inclusion and exclusion patient criteria as contextual data constraints in clinical drug trials to treat spinal cord injury. The analysis of Hontscha et. al.[117] described inclusion and exclusion patient criteria in the context of clinical trials for cancer treatment with CIK cells (natural killer T-lymphocytes called cytokine-induced killer cells) immunotherapy. The analysis combined the results of 11 studies on clinical trials on cancer treatment with CIK cells. Inclusion and exclusion patient criteria described in these studies are contextual data constraints that support information relevancy regarding specific analytical purposes.

Policies to specify contextual constraints such as inclusion and exclusion patient criteria should be supported by a recipient system linked to HDS to filter irrelevant data provided from the sources. More sophisticated policies can also be developed to measure information relevancy based on information retrieval (IR) quality metrics such as recall and precision. In this direction, the study of Savova et. al.[118] describes an approach for automatic information extraction from clinicians' free text (represented in natural language) in EMRs based on NLP

[116] Tuszynski, M. H., Steeves, J. D., Fawcett, J. W., Lammertse, D., Kalichman, M., Rask, C., ... & Ellaway, P. H. (2007). Guidelines for the conduct of clinical trials for spinal cord injury as developed by the ICCP Panel: clinical trial inclusion/exclusion criteria and ethics. Spinal Cord, 45(3), 222-231.
[117] Hontscha, C., Borck, Y., Zhou, H. W., Messmer, D., & Schmidt-Wolf, I. G. H. (2011). Clinical trials on CIK cells: first report of the international registry on CIK cells (IRCC). Journal of cancer research and clinical oncology, 137(2), 305-310.
[118] Savova, G. K., Masanz, J. J., Ogren, P. V., Zheng, J., Sohn, S., Kipper-Schuler, K. C., & Chute, C. G. (2010). Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. Journal of the American Medical Informatics Association,17(5), 507-513.

(natural language processing) methodologies. In this study, the evaluation of data relevancy utilizes recall, precision, F-score and accuracy metrics.

Another case of contextual data constraints, is the requirement for unique patient identifiers (which represent the patients' unique identities) that enable the mapping of the patients' information that is provided from multiple health data sources to the correct patient identities. This constraint ensures that a patient's health information, which is distributed in multiple sources, will be mapped only to his unique patient identifier. The study of Quantin et. al.[119] describes national efforts in several countries (Germany, Australia, Canada, Denmark, Finland, Luxembourg, the United States, New Zealand, The Netherlands and the United Kingdom) worldwide to develop national unique patient identifiers in order to facilitate integration of patient information from hospital EMRs. The study also emphasizes on the heterogeneity of patient identifiers utilized in the different countries, which impedes the cross-border EMR data integration.

A recipient system linked to external HDS should develop policies that support the specification contextual data constraints such as unique patient identifiers (given that they are required to the information analysis) that will enable mapping of patient information provided from the sources to the correct patient identities. In case that unique patient identifiers cannot be established in the systems network (e.g. due to de-identified patient information[120]), there are alternative more error-prone patient identification policies based on statistical matching of patient characteristics such as name, day of birth, address, and Social Security number[121, 122].

An efficient linkage of a recipient system to external health data sources should include policies that are able to specify contextual constraints such as patient inclusion/exclusion patient criteria and unique patient identifiers (given that they are required to the information analysis) and utilize them to the resolution of contextually inconsistent information derived from the sources.

---

[119] Quantin, C., Cohen, O., Riandey, B., & Allaert, F. A. (2007). Unique patient concept: a key choice for European epidemiology. International Journal of Medical Informatics, 76(5), 419-426.
[120] Uzuner, Ö., Luo, Y., & Szolovits, P. (2007). Evaluating the state-of-the-art in automatic de-identification. Journal of the American Medical Informatics Association, 14(5), 550-563.
[121] Castro, D. (2009). Explaining international IT application leadership: Health IT.Available at SSRN 1477486.
[122] Identity crisis: An examination of the costs and benefits of a unique patient identifier for the US health care system. RAND Corporation, 2008.

3.3.3.3 Missing Data

Missing Data is defined as low quality health information that is absent upon request. High amount of missing data from the sources may prohibit the analysis that is required by the recipient system.

The study of Smith et. al.[123] analyzes the issue of missing clinical information during primary care visits. Missing information was identified, in laboratory and radiology results, physical examinations and medications. Possible causes of missing information are related in this study to privacy legislations and regulations, decentralized medical systems, inadequate inter-physician communication[124], the transfer of patients' within and across care settings and changes in patients' insurance plans. The analysis identified that the use of EMRs in healthcare facilities did not resolve the problem of missing clinical information. The study reported that missing information can lead to medical errors, such as unnecessary medical examinations and procedures, drug interactions and delayed diagnoses. In addition, missing information in hospital systems may lead to information bias in epidemiological research studies[125,126].

In addition, the wider problem of missing patient information in healthcare facilities when citizens transfer across national and international borders, has motivated initiatives such as the Continuity of Care Document Data Standard[127]. The standard is a joint effort of HL7[128] International and ASTM[129] using the HL7 Version 3 Clinical Document Architecture to define a minimal set of critical patient information, in order to facilitate continuity of care anytime and anywhere efficiently. This information will essentially be a clinical patient summary available to the patient or the patient's physician which will be printable and accessible using native mobile and web applications.

---

[123] Smith, P. C., Araya-Guerra, R., Bublitz, C., Parnes, B., Dickinson, L. M., Van Vorst, R., & Pace, W. D. (2005). Missing clinical information during primary care visits. Jama, 293(5), 565-571.
[124] Kripalani, S., LeFevre, F., Phillips, C. O., Williams, M. V., Basaviah, P., & Baker, D. W. (2007). Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care. Jama, 297(8), 831-841.
[125] SØRENSEN, H. T., Sabroe, S., & OLSEN, J. (1996). A framework for evaluation of secondary data sources for epidemiological research.International Journal of Epidemiology, 25(2), 435-442.
[126] Foley, S. M., et. al. (1992). Comorbidities, complications, and coding bias: does the number of diagnosis codes matter in predicting in-hospital mortality?. Jama, 267(16), 2197-2203.
[127] Ferranti, J. M., Musser, R. C., Kawamoto, K., & Hammond, W. E. (2006). The clinical document architecture and the continuity of care record. Journal of the American Medical Informatics Association, 13(3), 245-252.
[128] http://www.hl7.org/
[129] http://www.astm.org/

Simple policies to manage missing data in the information provided from external sources, include the filtering of information based on specified optional and required data types, or based on a specified threshold to the acceptable amount of missing data. The study of Lau et. al.[130] described another approach to manage missing data in EMRs in order to support cancer research. The approach applied data imputation procedures to missing information in an effort to improve the completeness of EMR data.

An efficient linkage of a recipient system to external health data sources such as EMRs, PHRs, Biorepositories and Social Networks (Patient and Lifestyle SNS) should include an evaluation of the information that is provided from the sources with regard to missing data, in order to avoid bias in the information analysis. Various policies can be supported by the recipient system with respect to the management of missing information from the data sources. Finally, the RS should ideally perform an evaluation of uncertainty in the results of the integrative data analysis due to missing information.

3.3.3.4 Conflicting Data

Conflicting data is information that contains opposite facts[131]. Information provided by external health data sources to a linked recipient system may contain conflicting data. The identification process of conflicting information should be implemented after the initial processing of information has yield semantically consistent data.

The study of Peabody et. al.[132] used standardized patient visits to three healthcare facilities to assess the accuracy of clinical diagnoses. The study identified conflicting data from the care settings with regard to patient (early and final) diagnoses. This study provides evidence that linkage of a recipient system to EMRs of healthcare facilities may lead to conflicting data that will impede the integrative analysis of the information.

An efficient linkage to external health data sources should include policies to resolve data conflicts[133] that may be identified in the information provided from the sources. However, there is a lack of studies to address the information conflicts that may arise in the health data

[130] Lau, E. C., Mowat, F. S., Kelsh, M. A., Legg, J. C., Engel-Nitz, N. M., Watson, H. N., ... & Whyte, J. L. (2011). Use of electronic medical records (EMR) for oncology outcomes research: assessing the comparability of EMR information to patient registry and health claims data. Clinical epidemiology, 3, 259.

[131] Dong, X. L., Berti-Equille, L., & Srivastava, D. (2009). Integrating conflicting data: the role of source dependence. Proceedings of the VLDB Endowment,2(1), 550-561.

[132] Peabody, J. W., Luck, J., Jain, S., Bertenthal, D., & Glassman, P. (2004). Assessing the accuracy of administrative data in health information systems.Medical care, 42(11), 1066-1072.

[133] Dong, X. L., & Naumann, F. (2009). Data fusion: resolving data conflicts for integration. Proceedings of the VLDB Endowment, 2(2), 1654-1655.

integration domain that involves heterogeneous sources such as EMRs, PHRs, Biomedical Repositories and Social Networks.

Strategies for information conflict resolution, which can be applied in the health data integration domain, are provided below:

- Conflict resolution in a per-conflict basis. Each conflict is managed individually as a separate case, when it is identified.
- Conflict resolution based on the timeliness of the conflicting information. For example, a resolution policy of this type may specify that given a data conflict, the most recent information is selected. The *temporal dimension* of the data is analyzed in this strategy.
- Conflict resolution based on the location of the data sources. For example, a resolution policy of this type may specify that information recorded in a healthcare facility should be selected to patient information recorded in non-hospital environment. The *spatial dimension* of the data is analyzed in this strategy.
- Conflict resolution based on the reliability of the data sources. For example, a resolution policy of this type may specify that given a data conflict, information provided from the most reliable sources is selected. Simple approaches, can specify pre-defined reliability levels to the health data sources while more sophisticated models can be developed to analyze the reliability of the heath data sources in real-time. The *source quality dimension* is analyzed in this strategy.
- Conflict resolution based on the quantity of the conflicting facts. For example, two healthcare facilities may have concluded to the same patient diagnosis recorded in their EMRs, in contrast to another hospital that specified a different diagnosis in its EMR. A recipient system linked to the three EMRs will receive conflicting information. A conflict resolution policy of this type can select the first diagnosis which in greater quantity. The *data quantity dimension* is analyzed in this strategy.
- Conflict Resolution based on the actual data types and values that are received. For example, a resolution policy of this type may conclude that given conflicting information about a patient's hematocrit value from different laboratories, the lowest value should be selected. The *data composition dimension* is analyzed in this strategy.
- Conflict resolution based on the results of each possible case of data selection. This policy considers all possibilities in the resolution of a data conflict. Eventually, the data conflict is resolved by selecting the information that leads to specific results. The *analysis result dimension* is considered in this strategy.
- Conflict resolution based on a hybrid approach using multiple of the above strategies.

An efficient linkage to external health data sources should include a strategy for the resolution of data conflicts that may arise in the integrated information provided from the sources.

3.3.3.5 Erroneous Data

Erroneous Data is defined as low quality health information which expresses an incorrect belief.

EMR data sources may contain erroneous information which can be caused by patient misreported medications, inter-clinician miscommunications, clinician misdiagnosis, or just human error during data insertion[134,135,136]. In addition, PHR and Social Network data sources may contain inaccurate data since a lot of information in these sources expresses the patients' personal beliefs and views which are not clinically validated. Linkage to external health data sources should include a process of data validation (as feasible as this can be) and error management policies in case of error identification.

In general, an efficient linkage to external health information systems should include a data quality assessment process along with appropriate policies to manage all the aforementioned categories of low quality information that can be provided from the sources. Data quality assessment is a composite process of information evaluation that includes an analysis for semantically inconsistent, contextually inconsistent, missing, conflicting and erroneous data. The evaluation of data quality is an essential process to minimize bias in the integrative analysis of heterogeneous information that is provided from external health data sources.

3.3.4 The Integration of Data: Architectural Infrastructures

According to Wikipedia, Semantic Integration[137] is the process of interrelating information from diverse sources and has to resolve several heterogeneity problems. These problems are usually divided in syntactic heterogeneities and semantic heterogeneities. The former refer to those due to differences in the access interface, query language and database models. The latter are caused by different data representations for schemas or instances. During the last 15 years, numerous systems have been developed, often targeting specific problems or areas. The main approaches, are either centralized – e.g. data warehouses, where data is stored locally – or federated – where data is left at the sources and accessed on demand. The selection of either approach depends on the type of solution to be deployed. Data warehouses might deal with data privacy issues and with outdated data. However, they provide better efficiency and allow tighter control to data managers over what data will be available. Federated approaches always access updated data, allow partial and non-managed data connections, but suffer from efficiency issues. Federated approaches, also known as query translation rely on a virtual schema that represents the space of queries that the user can

[134] Wagner, M. M., & Hogan, W. R. (1996). The accuracy of medication data in an outpatient electronic medical record. Journal of the American Medical Informatics Association, 3(3), 234.
[135] Staroselsky, M., Volk, L. A., Tsurikova, R., Newmark, L. P., Lippincott, M., Litvak, I., ... & Bates, D. W. (2008). An effort to improve electronic health record medication list accuracy between visits: patients' and physicians' response. International journal of medical informatics, 77(3), 153-160.
[136] Goldberg, S. I., Shubina, M., Niemierko, A., & Turchin, A. (2010). A weighty problem: identification, characteristics and risk factors for errors in EMR data. In AMIA Annual Symposium Proceedings (Vol. 2010, p. 251). American Medical Informatics Association.
[137] http://en.wikipedia.org/wiki/Semantic_integration

submit to the system. It is called 'virtual' because no data is stored centrally. Instead, each query is dynamically translated into a set of sub-queries for the databases to integrate, and their single results are merged into a global result, which is presented to the end-users as answer to his initial query.

Either in virtual integration or data warehousing, during the last years, ontologies have been used in order to integrate structured and semi-structured data, obtaining promising results, for example in the fields of biomedicine and bioinformatics[138]. However, there is not a single correct way to model a domain and several ontologies exist. Example such ontologies include Symptom Ontology[139], was designed around the guiding concept of a symptom, the Disease Ontology[140] (DO) is trying to link disparate datasets through disease concepts, the Foundational Model of Anatomy[141] has to do with the phenotypic structure of the human body, whereas Adverse Event Ontology[142] tries to model adverse events. The Experimental Factor Ontology focuses on experimental variables in Gene Expression Atlas[143], the Clinical Care Classification System[144] tries to code health care settings and the Current Procedural Terminology[145] (CPT) is a medical nomenclature used to report medical procedures and services under public and private health insurance programs. UMLS[146], the Unified Medical Language System, is a unifying framework, which integrates different terminologies, which are relevant to medicine and biomedical information technologies. The Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) is a clinical terminology, which has been promoted as a reference terminology for electronic health record (EHR) systems. SNOMED CT is used by the College of American Pathologists[147], the UMLS Metathesaurus[148], the European project epSOS[149] and the European project SemanticHealthNet[150]. The Medical Subject Headings[151] (MeSH) are a medical thesaurus published and annually updated by the US National Library of Medicine (NLM). It is used for cataloguing of the library holdings and

[138] Tom Mitchell 1997 "Machine Learning", McGraw Hill
[139] http://symptomontologywiki.igs.umaryland.edu/wiki/index.php/Main_Page
[140] http://www.obofoundry.org/cgi-bin/detail.cgi?id=disease_ontology
[141] http://sig.biostr.washington.edu/projects/fm/AboutFM.html
[142] He, Y., Xiang, Z., Sarntivijai, S., Toldo, L., Ceusters W. AEO: A Realism-Based Biomedical Ontology for the Representation of Adverse Events, Int. Conf. on Biomedical Ontology, Representing Adverse Events Workshop, July 26, 2011
[143] http://www.ebi.ac.uk/gxa/
[144] http://en.wikipedia.org/wiki/Clinical_Care_Classification_System
[145] http://www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/cpt.page
[146] http://www.nlm.nih.gov/research/umls/
[147] http://www.cap.org/apps/cap.portal
[148] http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html
[149] http://www.epsos.eu/
[150] http://www.semantichealthnet.eu/
[151] http://www.ncbi.nlm.nih.gov/mesh

for indexing of the databases that are produced by the NLM (e.g. MEDLINE). ACGT MO[152] tries to model medical knowledge in the Cancer domain. The International Classification of Diseases[153] is the world's standard tool to capture mortality and morbidity data. LOINC is a database and a universal standard for identifying medical laboratory and clinical observations and Medical Dictionary for Regulatory Activities (MEDRA) is a clinically validated international medical terminology for diagnoses, symptoms, surgeries and other medical procedures. The Thesaurus of the National Cancer Institute[154] (NCI) covers vocabulary for clinical care, translational and basic research and public information and administrative activities. Moreover, other ontologies try to model multiscale data such as the Systems Biology Ontology[155] and Gene Ontology[156] (GO)**Error! Reference source not found.**, which supports biologically meaningful annotation of genes and their products in different databases.

Besides these ontologies that refer to core medical knowledge mostly other ontologies try to cover the domain of social entities that are related to health care such as Ontology of Medically Related Social Entities[157] and the BioCaster Ontology[158] (BCO), which tries to describe the terms and relations necessary to detect and risk assess public health events. The FHHO (Peace & Brennan[159], 2007) is representing the family health histories of persons related by biological and/or social family relationships (e.g. step, adoptive) who share genetic, behavioural, and/or environmental risk factors for disease.

Obviously, the amount of information available, the heterogeneity of the information and the wide range of proposed ontologies dictate the identification of a solution being able to handle all this information available. This is why the latest years approaches have been developed[160] that use multiple ontologies as target schemata and then they use those ontologies to formulate queries, which are being answered by the underlying data management solution.

4 MHA Link with external sources: technical description and implementation

Prevention, early diagnosis and personalized treatment are the three core processes of health management. The case studies which were previously presented shown that linking to

[152] http://bioportal.bioontology.org/ontologies/1126
[153] http://www.who.int/classifications/icd/en/
[154] http://ncit.nci.nih.gov/
[155] http://www.ebi.ac.uk/sbo/main/
[156] http://www.geneontology.org/GO.consortiumlist.shtml
[157] http://omrse.googlecode.com/svn/trunk/omrse/omrse.owl
[158] Collier, N., et al. An ontology-driven system for detecting global health events, Int. Conf. on Computational Linguistics (COLING), 2010, 215-222
[159] Peace, J, Brennan, P.F. Ontological representation of family and family history, at AMIA Annu Symp Proc. 2007
[160] Tom Mitchell 1997 “Machine Learning”, McGraw Hill

external health and health related data sources can extent and empower medical research eHealth services and clinical practise, and thus enhance the quality of healthcare. Nevertheless, the process of implementing an efficient physical linkage to external health data sources has many issues that should be addressed, and these were also described in the previous sections. Data accessibility, representation, quality evaluation and integration, are the main processes of an efficient linkage to external health data sources that require research and decision making, since there is a lack of standard policies to guide them. Ideally, risk assessment should also be conducted at each process to support patient safety and privacy, since the information that is exchanged is sensitive medical data.

The next section will present the technical approach that is developed in MyHealthAvatar project to overcome all of these barriers and achieve an efficient linkage to external health data sources that will enable health information exchange and the consequent integrative analysis of medical information.

4.1 Social networks

MHA infrastructure supports linking to social networks. MHA provides social web mechanisms and encourage the patients/citizens to adopt those in order to define their digital avatar. This requires integration with the social network accounts that the patients maintain already and the extraction of the social graph and other information. MHA is able to collect data from online patient diary using the utilities provided by T6.1. Volunteers will be organized to participate the research in this task. The popularity of social media allows users to link their account profiles from social networks like Twitter, Facebook or Google+ with MHA platform that become a communication hub for collecting people's personal stories and life experiences. MHA will then be able to contain a large volume of potential personal health information. The use of data mining techniques in the exploration of personal health information from these social networking services is the goal of this specific research task. The key problem we identified and focus on is on how to extract meaningful information from a large volume of data from popular social media services like Facebook, Twitter, etc. Details on the storage of the information have already been reported in D6.1 and implementation/evaluation of the repositories can be found in D6.2.

We have explored and implemented the connection to the main social networks including Facebook, Twitter and Google+, we support connect the social network by their API (through OAuth 1 and 2). The detail of how it is connected and architecture graphic is described in previous deliverables. We have taken into consideration of Identity Federation with social networking, namely Login with Facebook/Twitter/Google, which would ease the user in terms of manage their credentials. With users' explicit authorisation, MHA is able to read user's post, friend list, etc. and also able to post information back to the social network. E.g. Their daily activities from MHA. The post back to social network will increase the MHA exposure to general public, and potentially attracts new users to MHA. Link with social networks implementation

MyHealthAvatar platform utilises the APIs provided by social networks in order to make dual way communication with the social networks. In the flow of connection with social networks, MyHealthAvatar server act as API consumer while social networks are service providers.

During the project development, three social networks connection has been implemented which are Twitter (OAuth 1), Facebook (OAuth 2) and Google+ (OAuth 2). Two sets of Java libraries are utilised for the OAuth protocol, which are Spring Social and ScribeJava. The main reason two sets of libraries are used are mainly because of the API of the social networks obsolete and update frequently, the Spring Social set of libraries are more functionally complete, however updates relatively slow. ScribeJava is relatively simple libraries, which need extra work on top to support required functionalities. These two sets of libraries complement each other during the project development cycle.

As mentioned above that Spring Social set of libraries are more functionally rich, which also means that they have introduces quite a lot dependencies into MyHealthAvatar projects, which are listed in following figure.

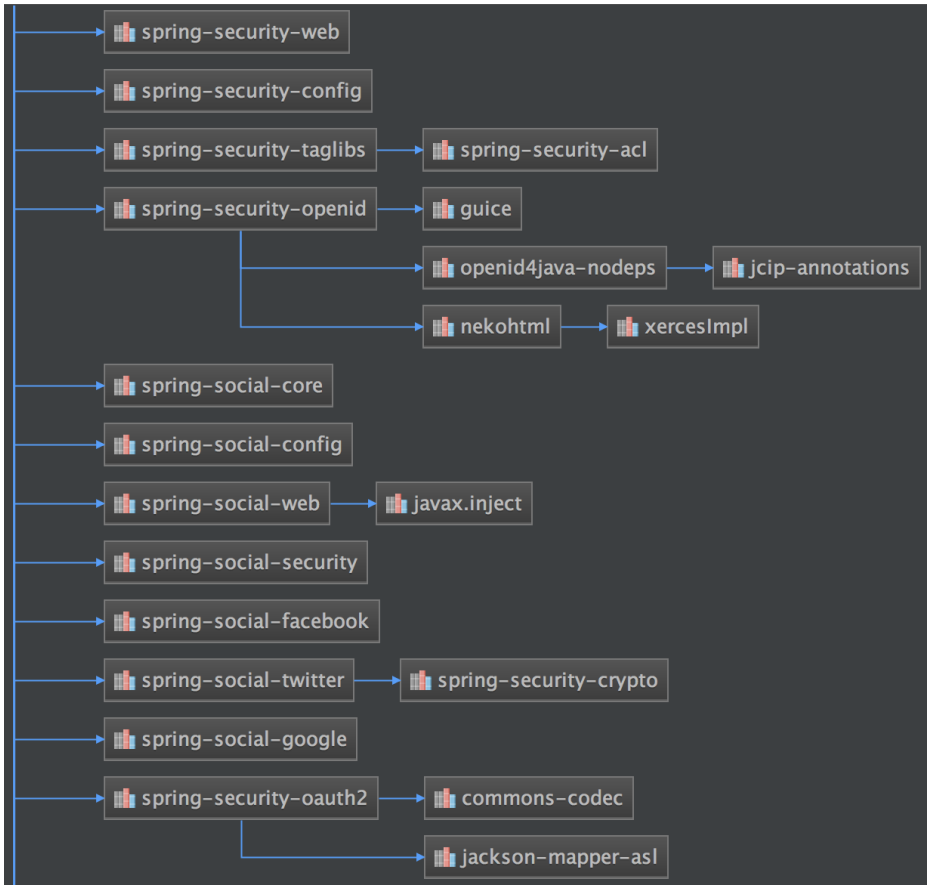


Figure 4-1: *Spring Social Dependencies in MyHealthAvatar*

The design of the data repository could be found in D6.1 and data repository implementation details and evaluation could be found in D6.2.

4.1.1.1 Functionality of the link with social networks

Since the MyHealthAvatar is available from public URL (https://myhealthavatar.org/mha/) and allows registration of general public, the following description and screenshots are directly from the public site (in order to show several operations in one figure, special commands are issues to keep UI state of previous operations). Technically in development environment, MyHealthAvatar connects to Twitter, Facebook and Google+. However, please note at time this document is written, only Twitter is listed in the social networks option. This is related to recent finalisation of MyHealthAvatar legal framework, and other social networks are still reviewing MyHealthAvatar privacy statements, terms and conditions in order to allow MyHealthAvatar's public site to connect with them.

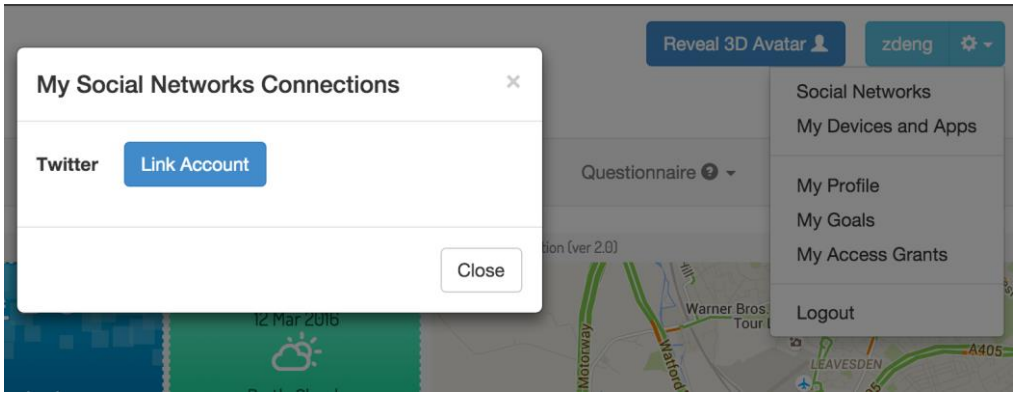


Figure 4-2: MyHealthAvatar Link with Social Network

As you can see from above figure, after user register and login MyHealthAvatar, from the dropdown menu item, click on 'Social Networks' will show the pop up overlay of 'My Social Network Connections'. Click on the 'Link Account' button will need to user authorize page of twitter (assume the user is already logged in Twitter, or user will be prompt to login first).

Figure 4-3: User Authorize Page for MyHealthAvatar

Once user is connected, MyHealthAvatar will keep the token issue by Twitter for connected user and able to communicate with Twitter using the token, like viewing the user's storyline and friend list. Also user is now able to publish the health activity related information to Twitter through MyHealthAvatar by few simple clicks. Following graph shows few tweets read by MyHealthAvatar from Twitter.

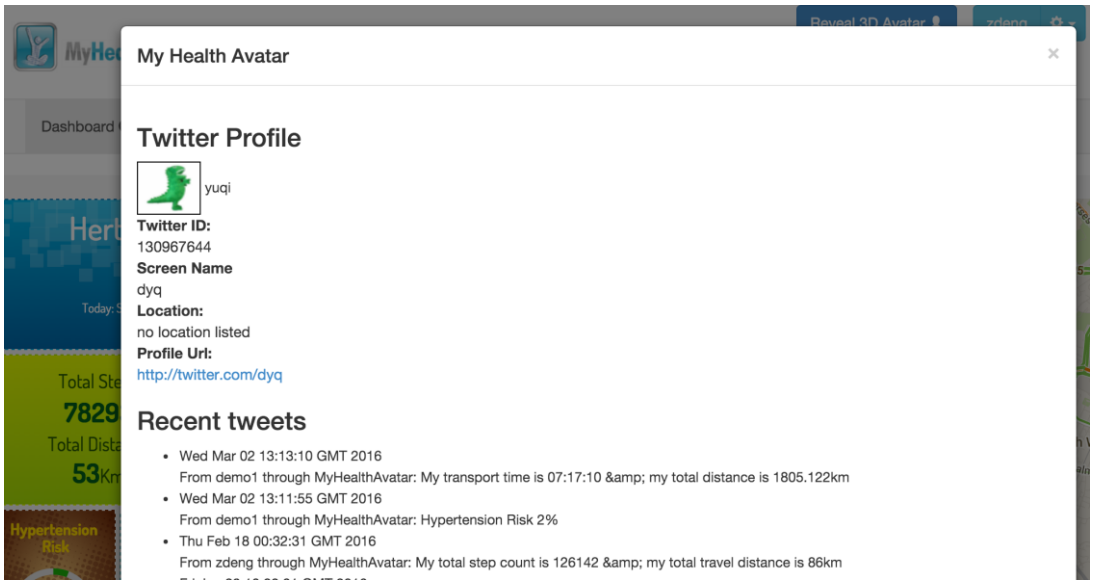


Figure 4-4: MyHealthAvatar retrieved Twitter Information

As you might have noticed that the top three messages are actually published to this twitter account from within MyHealthAvatar itself. The information published by user includes transport time, total active distance and hypertension risk (through the risk models of MyHealthAvatar platform).

All social networks operates in more or less similar way, with users' explicit authorisation, MyHealthAvatar is able to behave as he central hub for read and write users' health related information from and to social networks. The information includes texts, images, location information (if enabled by user), and friend list etc.

4.2 External data warehouses

4.2.1 Linking with external data repositories (CHIC repository)

MyHealthAvatar has created a link to the CHIC project, by utilizing its Clinical Data Repository as an external data source. The CHIC repository hosts all medical data produced or collected by the CHIC project and has interfaces to import or export its contents. The collected data are intended to be reusable by other projects. For each patient all the relevant medical data, including imaging data, clinical data, histological data and genetic data are stored.

Under the signed CHIC-MyHealthAvatar agreement, the CHIC repository will be involved in the following workflow pertaining to the Nephroblastoma Use Case:i) MyhealthAvatar partners will create a set of synthetic data for use as input for the Nephroblastoma Oncosimulator.

ii) The CHIC project will create a section in the CHIC repository where these synthetic data will be stored. In addition it will provide credentials for access and application programming interfaces (access API) for data storage and retrieval, which will be used from MyHealthAvatar to log in and store their synthetic data

iii)For a Nephroblastoma Oncosimulator Execution, MyHealthAvatar will log to the CHIC repository via its API and chose the necessary data. Then the Nephroblastoma Oncosimulator will log to the CHIC repository, retrieve the data and proceed with its execution.

iv) The execution results will be stored back to the MyHealthAvatar Platform by the Nephroblastoma Oncosimulator, by using the platform's API.

The following figure demonstrates the connections taking place in the aforementioned workflow:

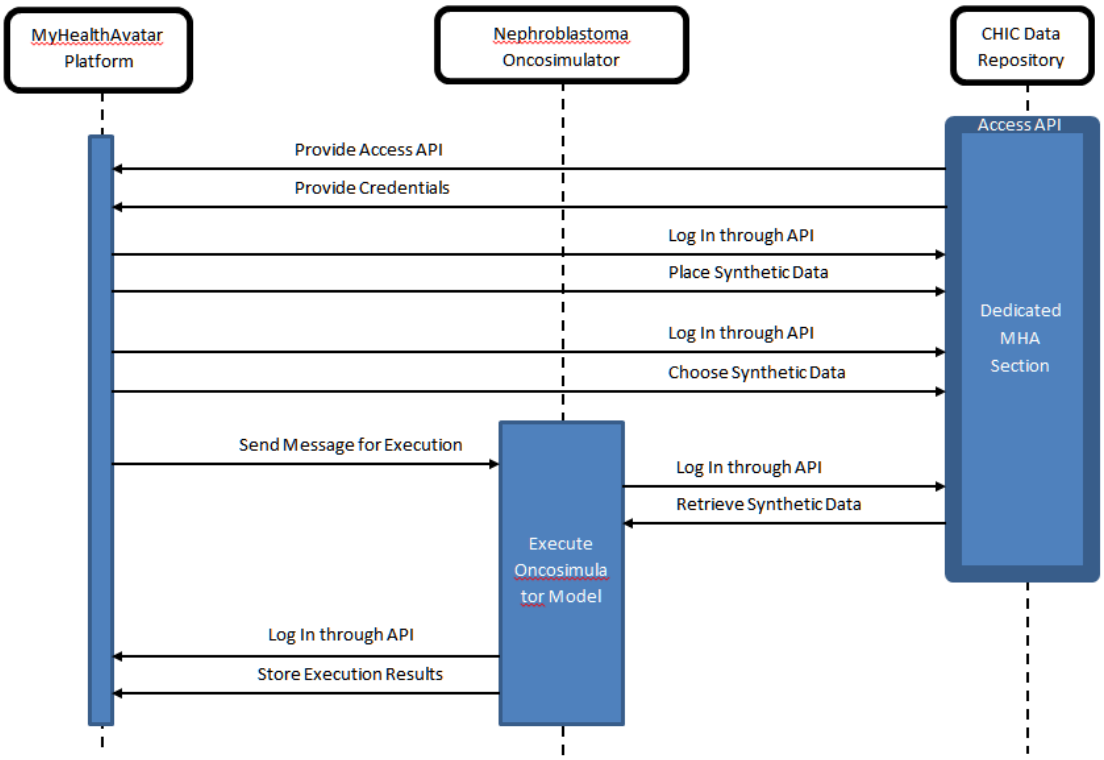


Figure 4-5: Workflow of access and data for the Nephroblastoma Use Case

For the implementation of the described workflow, the Nephroblastoma Oncosimulator Application, which encases the simulation model, is build using python's Django. Therefore, libraries such as urllib2 and/or pycurl (where necessary), are used to make requests to the MHA platform and the CHIC repository API's, using the necessary credentials as headers. Through these requests, files are exchanged via streams (content_type='application/octet-stream) and the descriptive data are exchanged via JSON. The API documentations for the MHA platform and the CHIC repository are given through the corresponding links: https://myhealthavatar.org/api/doc/v2/index.html and https://cdr-chic.ics.forth.gr/api/help (production version). Since the CHIC project is still ongoing, there is also a development version of the repository's API: https://cdr-dev-chic.ics.forth.gr/api/help

4.2.2 Linking with external drug data repositories

Personal health systems try to deal with issues of well-being, prevention and management of diseases and thus enhance patient empowerment and self-care management. To this respect a major challenge for patients –especially for chronic diseases where comorbidities may co-exist –is the optimum prescription of administered drugs in order to avoid any adverse drug reactions (ADRs), drug –drug interactions (DDIs) as well as patient's compliance with physician's instructions for lifestyle (i.e. alcohol consumption and smoking) and optimum drug administration.

DDIs describe the modulation of action of one or more concurrently administered medications due to prescriptions errors or due to self –medication[161]. The problem is becoming complex taking into account patient's lifestyle and the use of alternative therapies (i.e. herbal medicines), alcohol consumption and smoking[162, 163]. The interaction mechanism can be related either with synergistic or antagonistic effects in the site of action (pharmacodynamic, PD) or with alteration of absorption, distribution, metabolism and elimination pathways (pharmacokinetic, PK). The impact of DDIs, as well as DFIs, can be classified based on the severity of the pharmacological outcome (Table 1). To this respect, with the exception of intentionally DDIs, such as combination of treatments for optimum outcome, the resulting DDIs can be of a minimum effect up to cases where combination should be avoided due to high risk for the patient to present adverse drug reactions, treatment failure or toxicity.

DDIs usually are recorded in cases of prescriptions for chronic diseases, such as cardiovascular problems, CNS disorders, diabetes, arthritis, respiratory diseases, viral infections such as HIV/AIDs or hepatitis C, cancer etc.[164]. In all these cases patients are prescribed with more than one medications or they should follow a specific lifestyle in order to optimize treatment's outcome. The significance of the problem can also be understood taking into account the impact of DDIs in hospital admissions and hospital revisits as well as the impact of adverse

[161] P. D. Hansten, and J. R. Horn, Drug Interactions: Analysis and Management: Wolters Kluwer Health, 2006.
[162] I. Vizirianakis, M. Spanakis, A. Termentzi, I. Niopas, and E. Kokkalou, "Clinical and pharmacogenomic assessment of herb-drug interactions to improve drug delivery and pharmacovigilance," Plants in Traditional and Modern Medicine: Chemistry and Activity, pp. 978-81, 2010.
[163] . R. G. Smith, "An appraisal of potential drug interactions in cigarette smokers and alcohol drinkers," J Am Podiatr Med Assoc, vol. 99, no. 1, pp. 81-8, Jan-Feb, 2009.
[164] A. Chatsisvili, I. Sapounidis, G. Pavlidou, E. Zoumpouridou, V. A. Karakousis, M. Spanakis, L. Teperikidis, and I. Niopas, "Potential drug-drug interactions in prescriptions dispensed in community pharmacies in Greece," Pharm World Sci, vol. 32, no. 2, pp. 187-93, Apr, 2010.

drug reactions into mortality for patients in intensive care units[165]. As a result, regulator offices and health organizations worldwide have already adopted the problem of dealing with cases of DDIs and also drug food interactions. The evaluation of the severity for a potential interaction is implemented through clinical studies under specific guidance that are provided from the regulatory offices of European Medicines Agency (EMA) and Food and Drug Administration (FDA)[166,167]. The studies are conducted usually during preclinical and clinical development of novel drugs or they are taking place as clinical research studies whenever use-case scenarios are presented in the literature for special cases of drug administration. As a result, a lot of guidance documents, which are also the basis for software tools and platforms for DDIs, have been developed regarding the research and prediction of interactions but also for the optimum prescription practices that should be followed from medical personnel[168, 169].

Apart of research software, nowadays, the rapid advancements in ICTs has generated a lot of platforms and tools for addressing challenges for information access, collection, sharing and analysis of health data. These options has lead in creation of electronic health records and mobile health applications whereas it was well understood that the vision of creating a virtual representation of health status is within reach[170]. In addition, a lot of online available tools regarding optimum prescription practices focused mainly on drug interaction tools were

[165] S. Dechanont, S. Maphanta, B. Butthum, and C. Kongkaew, “Hospital admissions/visits associated with drug-drug interactions: a systematic review and meta-analysis,” Pharmacoepidemiol Drug Saf, vol. 23, no. 5, pp. 489-97, May, 2014.
[166] EMA, "Guideline on the investigation of drug interactions," Committee for Human Medicinal Products (CHMP), 2012, p. 59.
[167] FDA, "Drug Interaction Studies —Study Design, Data Analysis, Implications for Dosing, and Labeling Recommendations," F. a. D. A. U.S. Department of Health and Human Services, Center for Drug Evaluation and Research (CDER) ed., 2012, p. 79.
[168] N. Ai, X. Fan, and S. Ekins, “In silico methods for predicting drug-drug interactions with cytochrome P-450s, transporters and beyond,” Adv Drug Deliv Rev, vol. 86, pp. 46-60, Jun 23, 2015
[169] T. Roblek, T. Vaupotic, A. Mrhar, and M. Lainscak, “Drug-drug interaction software in clinical practice: a systematic review,” Eur J Clin Pharmacol, vol. 71, no. 2, pp. 131-42, Feb, 2015.
[170] H. Kondylakis, E. G. Spanakis, S. Sfakianakis, V. Sakkalis, M. Tsiknakis, K. Marias, Z. Xia, Y. Hong Qing, and D. Feng, “Digital patient: Personalized and translational data management through the MyHealthAvatar EU project,” Conf Proc IEEE Eng Med Biol Soc, vol. 2015, pp. 1397-400, Aug, 2015.

developed (i.e. Medscape, Drugs.com, AARP, Drugbank, University of Meryland, etc.)[171,172,173,174,175] [16-20]. The main goal was the as much as possible personalization of the providing treatments that a physician intends to prescribe but also to provide as much as possible information in cases that are needed such as combination of therapies when comorbidities exist. However, most of these tools were developed under the prism of being medical-oriented, decision support systems providing data for treating physicians regarding patient's information status and not as patient –oriented tools with the necessary user–friendliness for a patient that may lack the medical knowledge of the medical personnel.

In this section, a DDIs tool that was generated in the context of MyHealthAvatar (MHA) is described. The DDI tool in MHA platform is developed using the following programming languages and frameworks: RESTWeb services, Spring MVC 3 framework, Twitter Bootstrap, HTML5, Java server pages, JQuery, CSS, Apache Tomcat server. The information source in the case of MHA was through the Drugbank[176] bioinformatics and cheminformatics database which was provided via an XML file online. , it is parsed and stored in query optimized relational schema (in Postgresql)

The complete functionality of the application is provided by two autonomous web services**.** The first service gives feedback regarding information for a specific drug when a full text search" in order to support non-exact string searching is typed. The feedback includes the name and the synonyms for the compound, a brief description of the pharmacological action along with the therapeutic category, route of administration and the DDIs that are known for

[171] AARP, "Drug interaction checker http://healthtools.aarp.org/drug-interactions."
[172] U. o. M. M. Center. Drug interaction tool http://umm.edu/health/medical/drug-interaction-tool
[173] Drugs.com. Drugs interaction checker http://www.drugs.com/drug_interactions.php
[174] Medscape. Medscape Drug interaction checker http://reference.medscape.com/drug-interactionchecker,
[175] D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, and J. Woolsey, "DrugBank: a comprehensive resource for in silico drug discovery and exploration," Nucleic Acids Res, vol. 34, no. Database issue, pp. D668-72, Jan 1, 2006.
[176] http://www.drugbank.ca/

this compound. Through this approach MHA manage to provide information for a patient-user regarding drugs that he/she may be receives. Using the MHA Semantic Core ontology[177], a modular ontology developed within MHA, we can describe prescribed drug information, and how these drugs can, harmfully or not, interact among each other. This integration is achieved by reusing terms from the sub-ontologies of the MHA Semantic Core Ontology. There are 34 sub-ontologies within MHA Sematic core ontology linked between to the eTMO ontology and via relations of equivalence (using owl:equivalentClass) and subsumption (rdfs:subClassof). One of these sub-ontologies is the DrugBank Ontology[178] with an equivalent link between the eTMO:Drug with the DrugBank:Drug. As such the entire ontology and information available in the DrugBank can directly be exploited and used. Some specific classes and used to represent patient medication information, DDIs and DFIs are shown in (figure 4-6). The second service that is provided through MHA is a DDIs checker. Through this service information for potential interactions between two drugs can be retrieved and provided to the user. The table below (Table 1) summarizes some characteristic paradigms of drug interactions that were retrieved from the implementation of the tool.

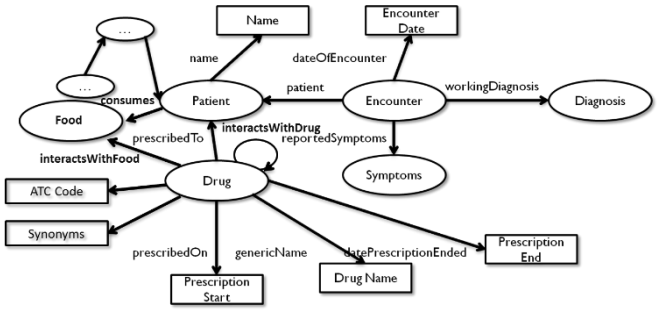


Figure 4-6: Classes from the MHA Semantic Core Ontology related medication DFIs and DDIs.

Table 1: Paradigms of identified interactions as they were retrieved through MHA DDI tool.

Drug	Co-administered Drug	PD/PK Mechanism	Outcome
Cisapride (Gastroprokinetic agent)	Citalopram (antidepressant)	PD	Citalopram may increase the QTc-prolonging activities of Cisapride.

[177] D4.2 Extension of the semantic core ontology, D4.3 Technical evaluation report of ontology including ontology evolution and summarization
[178] https://datahub.io/dataset/bio2rdf-drugbank

Metformin (type 2 diabetes)	Cimetidine (antibiotic)	PK	The serum concentration of Metformin can be increased when it is combined with Cimetidine.
Digoxin (congestive heart failure with atrial fibrillation)	Amiodarone (antianginal and antiarrhythmic drug)	PK	The serum concentration of Digoxin can be increased when it is combined with Amiodarone.
Enalapril (ACE inhibitor)	Acetylsalicylic acid (Aspirin -NSAIDs)	PD	Acetylsalicylic acid may decrease the antihypertensive activities of Enalapril
Losartan (angiotensin II receptor antagonist)	Sulfamethoxazole (antibiotic)	PK	The metabolism of Losartan can be decreased when combined with Sulfisoxazole
Repaglinide (antihyperglycemic)	Gemfibrozil (antilipdemic)	PK	The serum concentration of Repaglinide can be increased when it is combined with Gemfibrozil
Tamoxifen (selective estrogen receptor modulator used in breast cancer)	Paroxetine (SSRI, antidepressant)	PK	The serum concentration of the active metabolites of Tamoxifen can be reduced when Tamoxifen is used in combination with Paroxetine resulting in a loss in efficacy
Warfarin (anti-coagulant)	Azathioprine (Disease-modifying antirheumatic Drug)	PD	Azathioprine may decrease the anticoagulant activities of Warfarin.
	Alternative medicines		
Saquinavir (HIV drug)	St John's Wort (herbal medicine for depression)	PK	The metabolism of Saquinavir can be increased when combined with St. John's wort
Acenocoumarol, warfarin (anticoagulants)	Gingko (herbal medicine adaptogen)	PD	The risk or severity of adverse effects can be increased when Ginkgo biloba is combined with Acenocoumarol
Cyclosporin (immunosuppressa nt)	St John's wort	PK	The metabolism of Cyclosporin can be increased when combined with St. John's wort
	Food		
Selegiline (MAO inhibitor for Parkinson and dementia)	Tyramine containing foods (i.e. cheese, red winefermented	PD	Hypertensive crisis that can be fatal

	meats, fava beans, yeast extracts etc.)		
Warfarin (anti-coagulant)	Vegetables rich in Vitamin K	PD	Antagonism –Warfarin acts by decreasing activity of Vitamin-K in clotting mechanism. Consult your doctor before ingesting large amounts of dietary Vitamin K (e.g. from green leafy vegetables)
Simvastatin, atovarstatin, lovastatin , (hypolipidemic agents)	Grapefruit juice	PK	Inhibition of metabolism – Avoid taking with grapefruit juice
	Alcohol/Smoking		
Lorazepam (anti-anxiety)	Alcohol	PD	Synergistic adverse effects up to fatal respiratory depression. Avoid alcohol
Anesthetic agents	Smoking	PK	Increased metabolism of anesthetic and peri-operative problems during anesthesia

4.3 Hospital records

Linking with Hospital Information Systems (HIS) for the exchange of patient related clinical data requires a bilateral agreement on the data formats, communication protocols (in the case of machine to machine interactions), operational and message exchange patterns (e.g. pull versus push, periodic uploads, etc), etc. and of course the establishment of a close trust network to guarantee the privacy and the integrity of the data transmitted. Due to all these requirements and because of the dependence on the hospitals technical and management teams for setting up their end of the communication channel, we have opted for the adoption of standards based and well defined solutions. In more details the proposed approach is based on the following requirements:

- The general architecture for the HIS-MHA bridge is based on the results of the epSOS project[179]. The epSOS project is an EU co-funded so-called Large Scale Pilot aimed to contribute to the interoperability of eHealth services for an integrated healthcare system at the European level[180]. From a technical point of view the objective of epSOS

[179] http://www.epsos.eu/

[180] Lindén, F. (2009). epsos, smart open services for European patients from strategies to services health as the enabler for cross-border healthcare. Infrastructures for Health Care, 23.

was to provide concrete cross border services like the e-Prescription for the secure and efficient medical treatment of patients when travelling across the Europe. Being cross border (at least at the European level) it presents a perfect match for MyHealthAvatar's "software as a service" online platform.

- The ICT architecture of epSOS caters for many "cross-cutting"/orthogonal requirements of HIS-MHA bridge, such as patient identification and security, especially through liaison activities with specialized projects such as STORK and STORK2.0[181]. Additionally, it builds upon existing National eHealth projects and infrastructures and imposes a generic and HIS independent platform through the introduction of the National Contact Points. The National Contact Point, or NCP for short, is an organization at the national level (e.g. in Germany) to act as a bidirectional technical, organisational and legal interface between the existing different national health providers and infrastructures in order to fulfil the epSOS use cases, e.g. for e-Prescription.
- In terms of the actual content of clinical information that is exchanged in the epSOS domain, the Patient Summary (PS) consists of all the needed information in the context of a resident of one Member State (country A) visiting another MS (country B) and seeking for Health Care (Figure 4-7:). The PS made available to the Health Care professional of country B should contain updated and reliable information. The primary application of electronic Patient Summary is therefore to provide the Health Care professional with a dataset of essential and understandable health information at the point of care to deliver safe patient care during unscheduled care and planned care with its maximal impact in the unscheduled care. In the case of MyHealthAvatar's linkage with hospital records, the patient summaries can be expanded to include rich clinical data that are optional epSOS summaries. In fact the Health Care Encounter Report (HCER), which was added later to the palette of epSOS document types, is closer to the cross-border discharge summary concept and can be used to provide more accurate description of the content of a patient's clinical records.

In conclusion the epSOS architecture seems to provide the whole "bits and pieces" for the implementation of the linking infrastructure. It is important to note that epSOS Patient Summary is compatible with the HL7 Clinical Document Architecture (CDA) and in fact is defined as a CDA document template. The epSOS architecture is generic enough to accommodate the sharing of different document types, such as the Consolidated CDA and the Continuity of Care Documents[182].

[181] https://www.eid-stork2.eu/

[182] Estelrich, A., Solbrig, H., Cangioli, G., Melgara, M., & Chronaki, C. (2014, September). European Patient Summary Guideline and Continuity of Care Document: A Comparison. In Computing in Cardiology Conference (CinC), 2014 (pp. 481-484). IEEE.

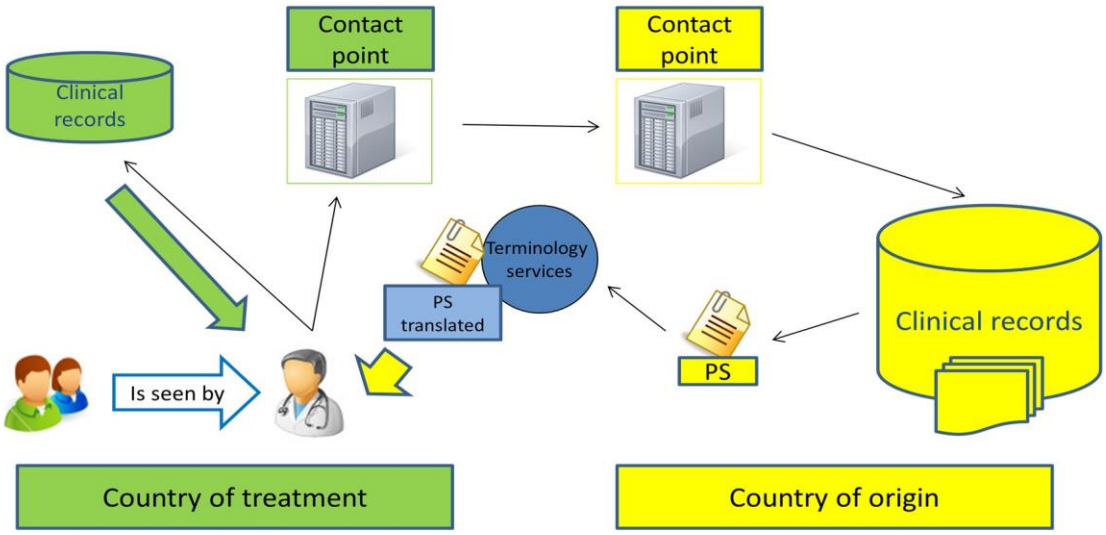


Figure 4-7: A primary use case of epSOS, showing the exchange of Patient Summaries through the National Contact Points

In the following figure we show the flow of the clinical records patient data through the epSOS gateway. According to epSOS guidelines and proposed architecture the MHA epSOS Gateway retrieves "Patient Summaries" from a "National Contact Point (NCP)". The National Contact Point is an organization and IT infrastructure to support internal (in the same country) or cross-border (across countries) communication for the exchange of patient data. The patient summaries retrieved through the NCP contain essential information needed for the continuity of care such as the most important clinical patient data (e.g. allergies, current medical problems, medical implants, or major surgical procedures), a list of current medication including all prescribed medication that the patient is currently taking, etc.

In the current deployment of this use case the implementation of the NCP retrieves the clinical data from a local installation of a production level Hospital Information System[183] using direct SQL queries to the system's database. Nevertheless, this is an internal implementation detail of the epSOS clinical domain. The important thing is that the MHA epSOS Gateway operating in the borders of the clinical domain and the MHA platform uses the epSOS Patient Summary documents to feed the central data repository of the platform. The Gateway filters and keep the relevant patient data, currently the medications and the vital signs measurements.

[183] Integrated Care Solutions (ICS), http://www.ics.forth.gr/ceha/index_main.php?l=e&c=471

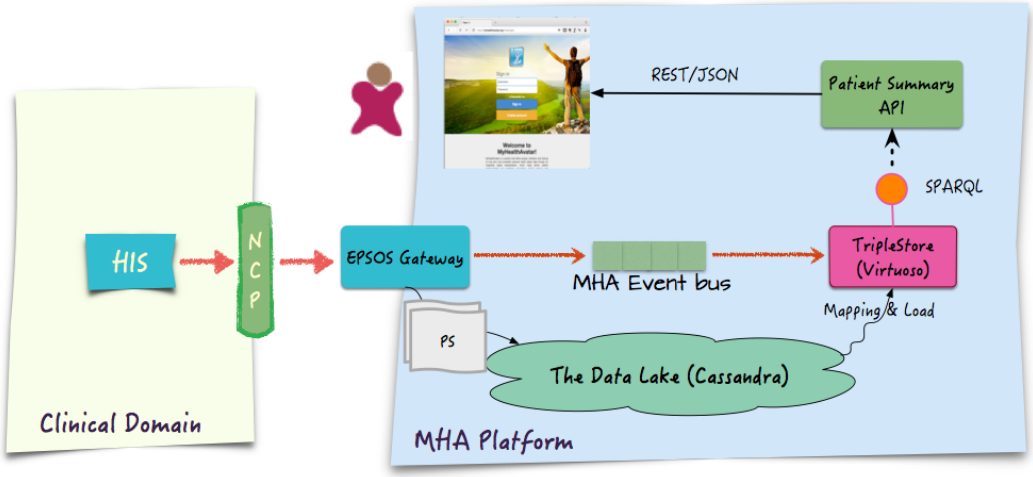


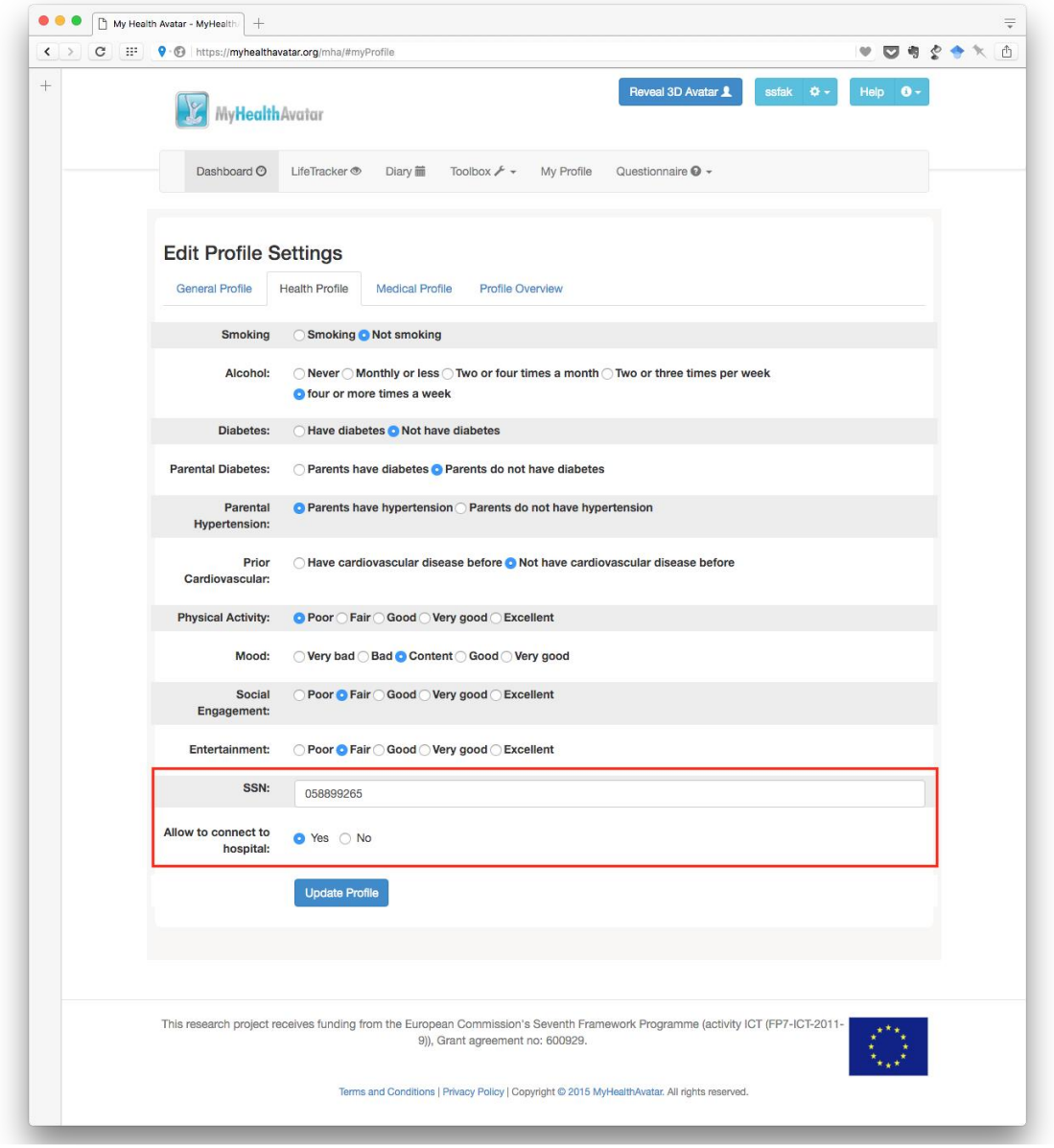
Figure 4-8: Retrieval of clinical records data using the epSOS Gateway and the track they follow inside the MHA platform

The retrieved patient summaries are then stored in the Cassandra repository and a "New Patient Summary" event is published through MyHealthAvatar's "event bus". The event bus component is a message queue that allows the (near) real time notification of the platforms' services when something of interest happens. The Semantic infrastructure with the underlying RDF Triplestore is one of the "subscribers" for the Patient Summary type of events. When a new event for updated patient summary of a user arrives in the Semantic services, the patient summary is retrieved from the Cassandra repository, the proper mappings are applied, and the result semantic content is stored in the RDF Triplestore where is then available to the rest of the platform.

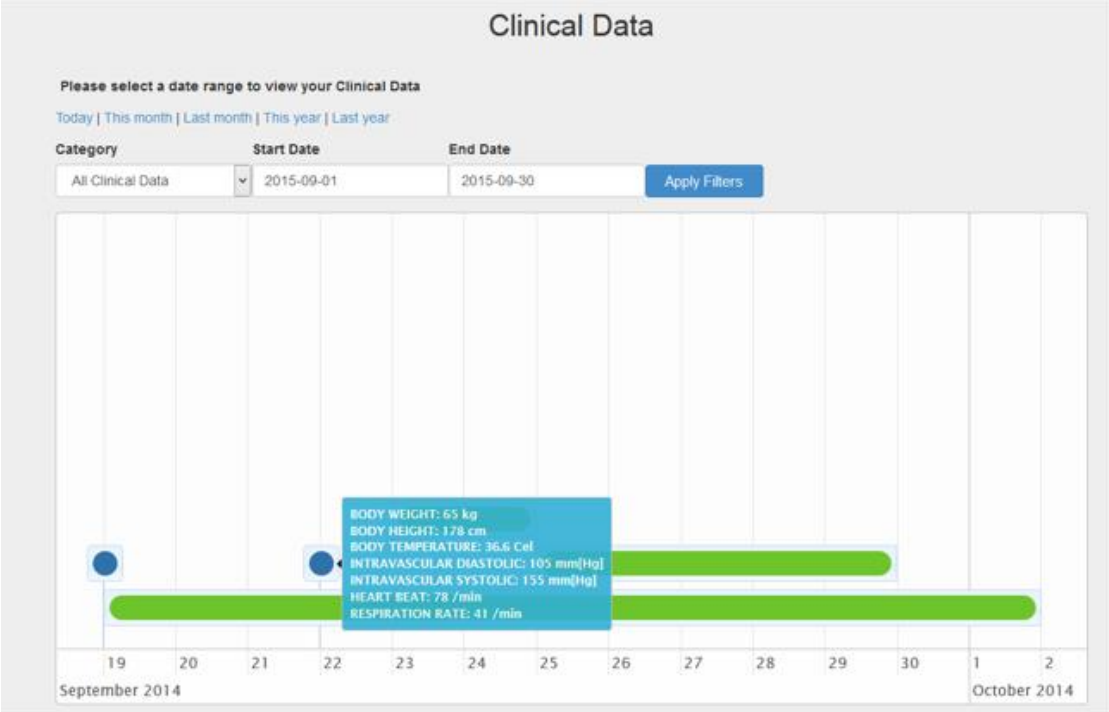
The EPSOS Gateway is configured to periodically "poll" the linked NCPs for updated clinical information on behalf of the MHA users. In order to reduce the unnecessary load in the platform, the Gateway keeps the state of the patient summary per user, so that the events for new information and the storage of the updated patient summaries are performed only when there's really a change in the acquired clinical information. The updated patient summary is stored in full detail in the Cassandra, i.e. the Gateway does not perform any detailed comparisons between the new and the previous version of the summary in order to find and store only the changes (e.g. for a new drug prescription). The identification of the changes is performed by the Semantic services that are also responsible of reconciliating those changes with the existing information according to the business rules and the semantic content of the Triplestore.

4.3.1 MHA Linking with Hospital Records: end user's point of view

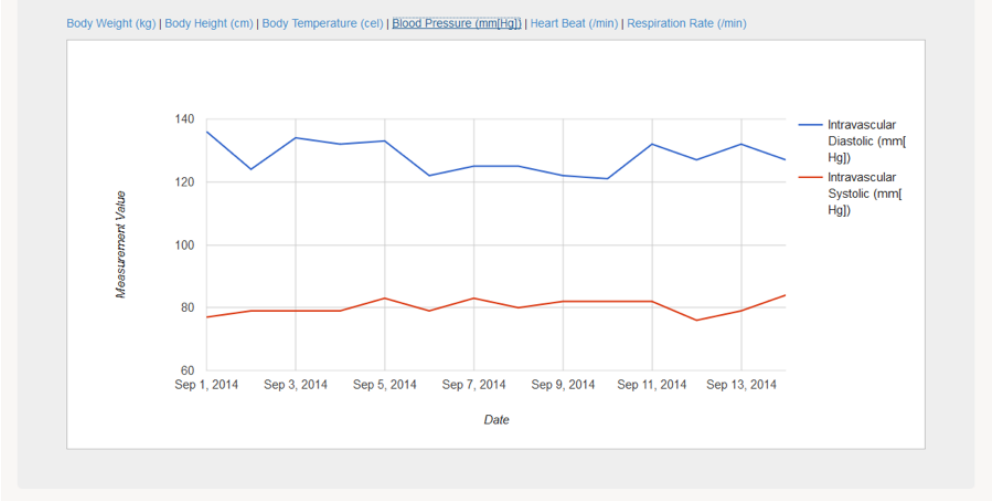
From the user's point of view linkage to the electronic health records is a straightforward process. As shown in the next figure, the User's Profile form allows the user to provide her/his Social Security Number and grant access to the Hospitals' patient records.



When the user has provided this information, the service responsible for the linkage with the electronic patient records, i.e. the EPSOS Gateway, is allowed to retrieve the user's patient summary using the Social Security Number as the patient identifier. When the patient summary has been retrieved and the platform has acquired and index the relevant clinical information as described in Section, the user is able to have a graphical representation of this information in a "timeline view" (Figure below).



Furthermore, the user is able to have a "trends" view of his/her vital signs over the time. The following picture shows the advancement of the user's diastolic and systolic pressures based on the historical data retrieved from his/her patient summary.



4.4 Personal Health Records

MyHealthAvatar store an process personal health record from several sources, it allows user to input personal health record from its UI, upload health record in an XML format which is similar with Continuity of Care Record (CCR) and retrieval data from third party PHR system Indivox-X.

For user direct input, standard web UI component include text box and dropdown selection are used to maintain maximum compatibility with user devices. It looks like following graph:

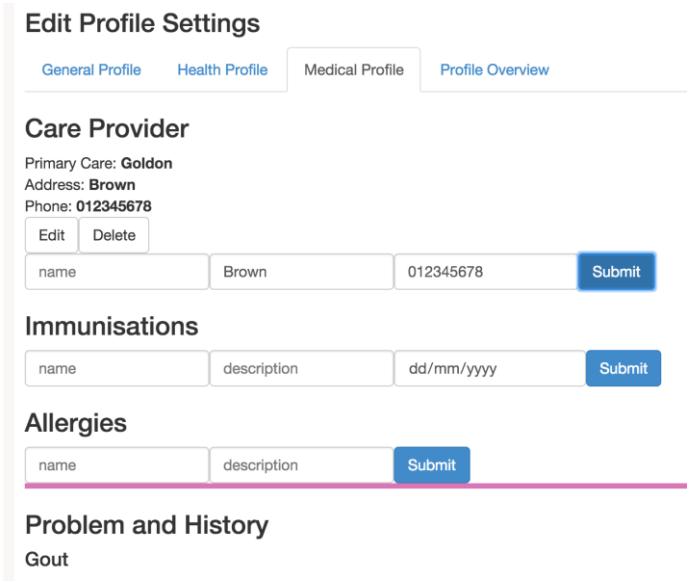


Figure 4-9: User input of health information

User are also allow to upload health record similar to CCR, the data format looks like following XML fragment:

```
<PatientSummary xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xsd="http://www.w3.org/2001/XMLSchema">
   <RecGUID>c42b05b8-5af5-4d90-a9c4-c2151e0ce452</RecGUID>
   <Identification>
      <RecGUID>dabeb48f-0f94-4a5a-8329-d98bf09fe349</RecGUID>
      <NationalPersonalID>12345678ABCDEF</NationalPersonalID>
   </Identification>
   <PersonalInformation LastUpdate="2012-09-14T13:51:37+03:00" User="useridentifier">
      <RecGUID>af48f278-6bee-4210-82ef-583b17ecea42</RecGUID>
      <GivenName>George</GivenName>
      <SurName>Dsouza</SurName>
      <DateofBirth>1990-08-01</DateofBirth>
      <Gender>Male</Gender>
      <MarriedStatus>Married</MarriedStatus>
      <ReligiousAffil>N/A</ReligiousAffil>
      <Ethnicity>White/Caucasian</Ethnicity>
      <Language>India</Language>
      <GenderCode>225134567</GenderCode>
      <Smoke>True</Smoke>
      <ContactInfo>
         <RecGUID>9fdbdb41-0891-42dc-80b4-0a15f33a226b</RecGUID>
         <Address>
            <Street>South Harrow Street</Street>
            <PostCode>HA2 3DE</PostCode>
            <StateProvince>Harrow</StateProvince>
            <Country>Middlesex</Country>
         </Address>
         <Phone>07834234567</Phone>
         <Email>george@georgedsouza.c
```

The upload and display interface for the health record can be found from main MyHealthAvatar web application, user profile page. It is designed with help with visualization expertise to look and feel like PDF file, it further allows export of the information to be a PDF documentation, so user can carry the hard copy to places internet connection is not available.

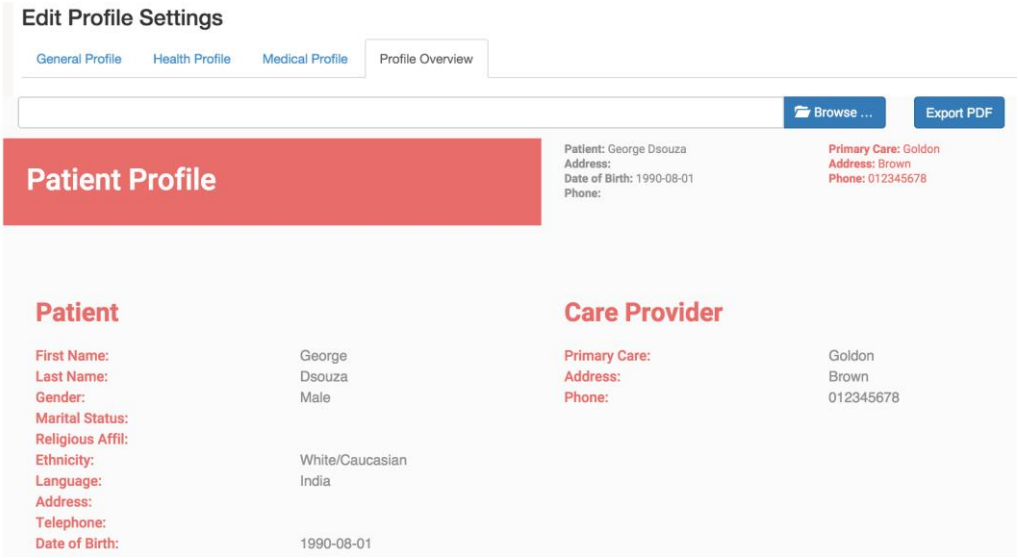


Figure 4-10: health record upload and export view

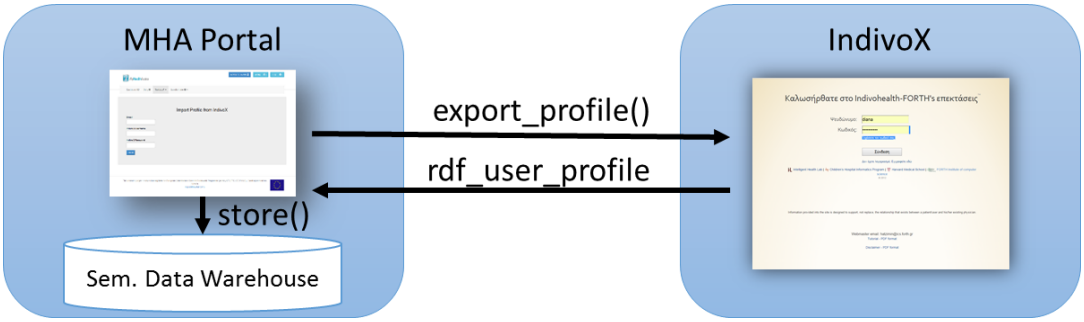


Figure 4-11: Loading Data to the MHA Semantic Data Warehouse

To link with Personal Health Record systems a PHR gateway similar has been implemented following a similar workflow with the mechanisms for connection with third party data source for activity data retrieval. The mechanism has been implemented for the Indivo-X PHR[184] system but it can be reused for any other system that exports data as XML or RDF/S. The implemented workflow is shown in **Error! Reference source not found.**. To start the whole process a MHA user should select from the MHA portal the corresponding service "Import

[184] Kenneth D. Mandl, William W. Simons, William C. R. Crawford, Jonathan M. Abbett: Indivo: a personally controlled health record for health information exchange and communication. BMC Med. Inf. & Decision Making 7: 25 (2007)

Profile from IndivoX" shown in Figure 4-12. Obviously the user requesting his profile to be imported from IndivoX should have already an account to the aforementioned platform. As soon as the MyHealthAvatar user enters his IndivoX and presses the export button, his entire profile from the PHR is exported from IndivoX, as an RDF/S. The IndivoX API allows exporting RDF/S data and an example describing two allergies is shown in figure 4-11.

The exported RDF/S document is then loaded to the Semantic Data Warehouse linked to his existing data and integrated into his entire profile.

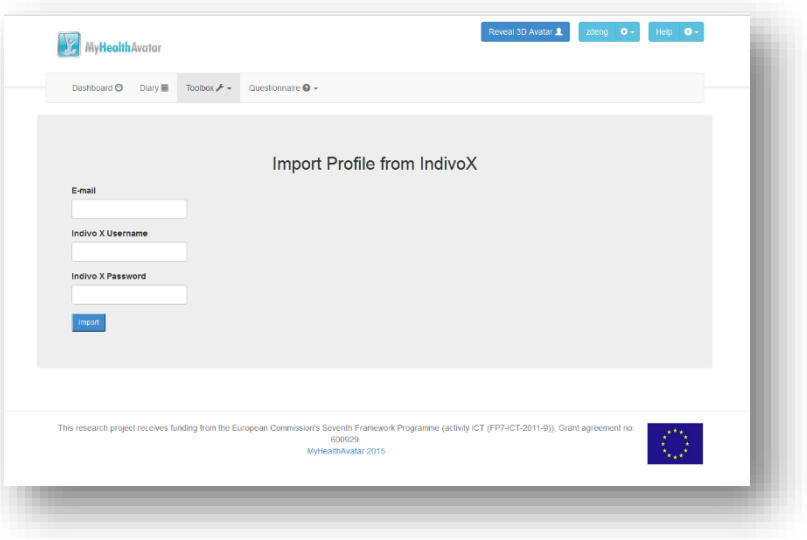


Figure 4-12: The GUI for importing the entire personal health record profile from IndivoX

```
<?xml version="1.0" encoding="UTF-8"?>

<rdf:RDF

  xmlns:dcterms="http://purl.org/dc/terms/"

  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"

  xmlns:sp="http://smartplatforms.org/terms#"

>

 <rdf:Description rdf:about="http://purl.bioontology.org/ontology/SNOMEDCT/271807003">

  <dcterms:title>skin rash</dcterms:title>

  <dcterms:identifier>271807003</dcterms:identifier>

  <sp:system>http://purl.bioontology.org/ontology/SNOMEDCT/</sp:system>

  <rdf:type rdf:resource="http://smartplatforms.org/terms/codes/SNOMED"/>

  <rdf:type rdf:resource="http://smartplatforms.org/terms#Code"/>
```

```
  </rdf:Description>

  <rdf:Description rdf:about="http://indivo.org/records/03059111-af61-4834-8234-
befe5f5a2532/allergies/f8efc96a-7677-4b4f-9879-7fc6d6488d0b">

   <sp:category rdf:nodeID="_865481f6-03ca-4707-8a89-ec468952efa5"/>

   <sp:severity rdf:nodeID="_9f6a6981-1173-4041-8fa3-4462238ab8ae"/>

   <sp:foodAllergen rdf:nodeID="_24be52e0-51a4-4d00-9654-25ae9e0ad2f4"/>

   <rdf:type rdf:resource="http://smartplatforms.org/terms#Allergy"/>

   <sp:belongsTo rdf:nodeID="_f63e49ae-5071-4f99-a62d-329a2e23ce85"/>

   <sp:allergicReaction rdf:nodeID="_7912ae70-da78-443f-a0b6-3f955b9e140a"/>

  </rdf:Description>

  <rdf:Description rdf:nodeID="_865481f6-03ca-4707-8a89-ec468952efa5">

   <dcterms:title>food allergy</dcterms:title>

   <rdf:type rdf:resource="http://smartplatforms.org/terms#CodedValue"/>

   <sp:code rdf:resource="http://purl.bioontology.org/ontology/SNOMEDCT/414285001"/>

  </rdf:Description>

</rdf:RDF>
```

Figure 4-13: Exporting allergies from IndivoX

4.5 Connection with third party data source for activity data retrieval and sensor

MyHealthAvatar connects to third party services to synchronise the wearable sensors and mobile applications' data, which mainly are activity data. The connections are mainly through third parties APIs. The connection process is more or less similar to the way MyHealthAvatar connect to social networks.

In order to make the connection, MyHealthAvatar first register with third party and obtain credentials which allows MyHealthAvatar to communicate with them. Following table list the third party registration page URL with notes:

Name	URL	Notes
Google	https://code.google.com/apis/console	APIs & auth => Credentials => CREATE NEW CLIENT ID
Facebook	https://developers.facebook.com/	Apps => Create a new app
Twitter	https://dev.twitter.com/	Top right => My applications => Create New App
Fitbit	https://www.fitbit.com/dev	Register an app

Moves	https://dev.moves-app.com/apps	Create a New App
Withings	http://oauth.withings.com/partner/dashboard	

OAuth 1 and 2 protocols are mainly used by third party providers for API security and user delegation, there are differences in the way protocols are implemented by different providers. Two main reasons cause the difference in implementation, firstly the protocols has evolved quite few revisions until them reaches final and secondly there are some technique details not specified by protocol themselves and leave to providers to interpret. MyHealthAvatar has built a flexible class hierarchy to dealt with these technique challenges, and following class diagram take Fitbit connection as example:

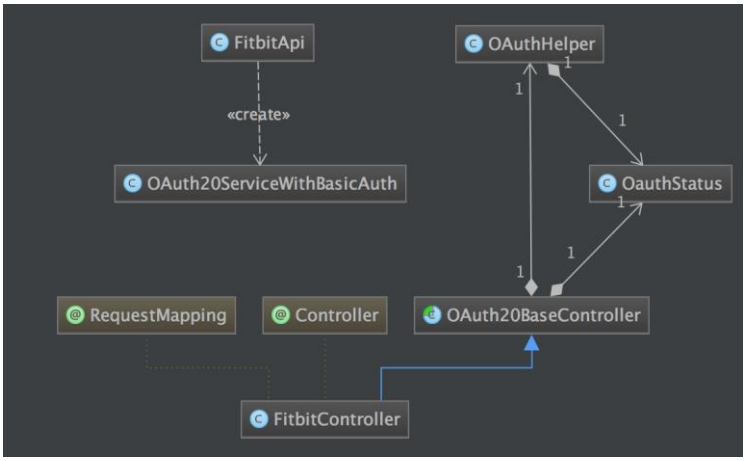


Figure 4-14: MyHealthAvatar Fitbit Connect Class Diagram

The OAuth connection flow and implementation details could be found in related deliverables, after users connect with their devices' services, from MyHealthAvatar interface it would look like following.

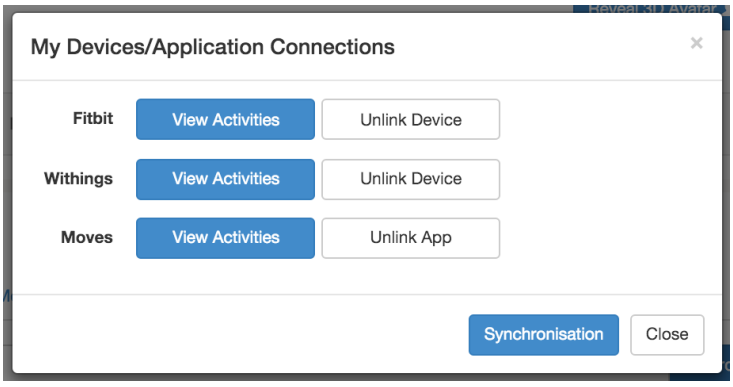


Figure 4-15: MyHealthAvatar connected devices/appsIntellectual property rights and the linking with external sources

4.6 Handling the IPR aspects by linking MHA with external data sources

The IPR issues following from the sharing of data with third party platforms and projects have been extensively considered in D.11.3, Section 4.3.

The legal requirements and rules for handling the IPR issues by linking MHA with external data sources, which are proposed for implementation by MHA at the exploitation stage, are described in Deliverable D.11.4, Section 5.

In particular, connecting MHA to the following platforms and projects has been discussed:

(a) Fitbit – a platform for the users of Fitbit products. Fitbit products allow to track physical activity of the fitbit user, including activity, exercise, food, weight and sleep.[185]

(b) Withings – a platform for the users of Withings devices. Withings devices are smart products and apps that fit into any lifestyle and let the user track his activities in order to improve everyday well-being.[186]

(c) Moves – a platform for the users of Moves. Moves automatically records any walking, cycling, and running of the user and allows to view the distance, duration, steps, and calories burned for each activity.[187]

(d) Twitter – a platform that allows its users to get and provide in-the-moment updates and watch events unfold, in real time, from every angle.[188]

(e) Facebook - a social platform which allows its users to connect to people across the world.[189] The code logic which connects to Facebook - has been developed, but MHA is not supposed to connect to Facebook within the lifetime of the Project.

(f) CHIC – "Computational Horizons In Cancer (CHIC): Developing Meta- and Hyper-Multiscale Models and Repositories for In Silico Oncology", EU FP 7 Project.[190]

The legal aspects, which concern data sharing with third party data warehouses, consider the terms of use of platform APIs, getting the user´s consent on processing the user´s personal data, handling the user generated content and obtaining the copyright license, requirements which third party platforms, like Twitter provide for display of the Twitter content on third party services, security aspects, which need to be observed by the exchange of data.

4.6.1 Sharing and processing IP protected content on MHA services

Sharing IP protected content from third party platforms and processing such content on MHA services has been considered in D.11.3, Section 4.3. The solution for handling the IP rights in

[185] Fitbit, http://www.fitbit.com.

[1] [186] Withings, see http://www2.withings.com/us/en/.

[187] Moves, see https://www.moves-app.com/.

[188] Twitter, see https://twitter.com/.

[189] Facebook, see https://www.facebook.com/.

[190] CHIC, see http://chic-vph.eu/project/.

the user´ generated content and rules for processing IP protected content on MHA services, which are proposed for implementation in MHA, are provided in D.11.4, Section 5 (d). The copyright license on use of the user generated content is included into the MHA Terms of Use, Section VIII.

When connecting to a third party platform via its API, MHA obtains a possibility to share the user´s data, which may include creative content. Some of the content items, such as images, or comments, if produced by intellectual effort and expose certain degree of originality, may be protected by IP rights, most likely copyrights. Processing of content, protected by IP rights on digital media, including software applications and digital services, such as by copying, storage, display, transmission, making the content available to the public constitutes a copyright relevant action and requires authorization of the right holder.

The aspect, that the MHA platform may share some user generated content, protected by IP rights from third party platforms, and how MHA obtains the copyright license on use of such content is addressed in D.11.4, Section 5 (d). The copyright license, which the MHA obtains from the users, who connect to MHA platform, is incorporated into the MHA Terms of Use, Section VIII. User generated content. It reads:

"You understand and agree that some of the content which appears on the MyHealthAvatar Services, such as: text, photographs, audio-visual works, comments, illustrations, sketches and other information, may be protected by intellectual property rights or other proprietary rights ("IP Content"). When you submit, upload post, display or otherwise provide to the Platform any IP Content, you grant MyHealthAvatar a worldwide, non-exclusive, royalty-free license (with the right to sublicense) to use, reproduce, distribute, communicate, make available, transmit, process, display, digitally perform, modify, adapt and otherwise use such IP Content in any and all media, in any format or distribution methods, now existing or later developed, including, but not limited to websites, in audio format, and in any print media as needed and for so long as needed for MyHealthAvatar to provide its Services ("IP License")."

The Extended Terms and Conditions for exploitation stage after the project's end are provided in Annex 4 to D.11.4.

4.6.2 Data sharing with the CHIC project

The legal aspects of collaboration with the CHIC Project have been considered in D.11.4, Section 5 (c), the rules of collaboration and data sharing are laid down in the CHIC MHA Collaboration Agreement, attached to D.11.4 as Annex 8.

The MyHealthAvatar project intends to demonstrate the utility of the MyHealthAvatar platform by allowing its users to perform oncosimulations of Nephroblastoma with the use of clinical data. The MHA decided to complete this endeavor with the use of the CHIC data repository. The MyHealthAvatar project will generate synthetic clinical data for running Nephroblastoma oncosimulations in MyHealthAvatar and will place such data for storing into the CHIC data repository.

As part of its work, the CHIC project has developed a clinical data repository that provides for a secure storage of clinical data, incl. imaging data, histological data, therapy, etc. The data types hosted by the repository include: imaging data (DICOM, etc), descriptive/structural data (age, sex, etc), other files (histological reports), links (to other data repositories), etc.

The workflow and terms of use of the use by the MHA Parties of the CHIC data repository, associated infrastructure and components for the purposes are laid down in the CHIC MHA Collaboration Agreement.

The terms for using the MyHealthAvatar synthetic data are stipulated in Paragraph 4 CHIC MHA Collaboration Agreement. The synthetic data generated by MHA, does not have quality of personal data and does not raise privacy concerns. Access Rights and use of this data in MyHealthAvatar is governed by the rules on Access Rights under Section 9 of MyHealthAvatar Consortium Agreement

The security aspects of data sharing are addressed in the Collaboration Agreement, Section 9. By this, data transfer from CHIC to MHA should occur via a secure interface, provided by UBERN. The MHA Parties shall keep the credentials and identification information for the MyHealthAvatar account in CHIC data repository secure and confidential and take measures to protect the credentials from unauthorized access and use by third parties. UBERN may disconnect access of the MHA parties to the CHIC data repository if there are reasonable grounds to suspect that the MyHealthAvatar account is being used in a fraudulent or negligent manner which may cause liability for CHIC.

Both CHIC and the MHA parties agree to implement adequate Internet security measures, including state of the art data encryption, to ensure secure transfer of data in compliance with Internet governance and applicable laws.

5 Appendix 1 – Abbreviations and acronyms

ADaM Analysis Data Model CDISC standard supporting efficient generation, replication, and review of analysis results

AES Advanced Encryption Standard a specification for the encryption of electronic data based on a design principle known as a Substitution permutation network

AGPL Affero General Public License refers to two free software licenses. Affero General Public License, Version 1 and GNU Affero General Public License, version 3.

API application programming interface is a particular set of rules ('code') and specifications that software programs can follow to communicate with each other

ASCII American Standard Code for Information Interchange a character-encoding scheme based on the ordering of the English alphabet

BMP Bitmap a raster graphics image file format used to store bitmap digital images

BSD Berkeley Software Distribution is a Unix operating system derivative developed and distributed by the Computer Systems Research Group (CSRG) of the University of California, Berkeley

CAS Central Authentication Service a single sign-on web protocol

CA Certification Authority an entity that issues digital certificates

CBC Cipher-Block Chaining a cryptographic mode of operation in which each block of plaintext is XORed with the previous ciphertext block before being encrypted

CCM Counter with CBC-MAC Mode a mode of operation for cryptographic block ciphers

CCZero Creative Commons licenses are several copyright licenses that allow the distribution of copyrighted works

CDASH Clinical Data Acquisition Standards Harmonization CDISC standard describing the basic recommended (minimal) data collection fields for 18 domains, including common header fields, and demographic, adverse events, and other safety domains that are common to all therapeutic areas and phases of clinical research

CDA Clinical Document Architecture is an XML-based markup standard intended to specify the encoding, structure and semantics of clinical documents for exchange

CDE Clinical Document Architecture an XML-based markup standard defined by HL7 intended to specify the encoding, structure and semantics of clinical documents for exchange

CDISC Clinical Data Interchange Standards Consortium - a global, open, multidisciplinary, non-profit organization that has established standards to support the acquisition, exchange, submission and archive of clinical research data and metadata

CDMI Cloud Data Management Interface - defines a functional interface that applications can use to create, retrieve, update and delete data elements from the Cloud

CDS Clinical Decision Support decision support software designed to assist physicians and other health professionals with decision making tasks, as determining diagnosis of patient data

CMWG Cloud Management Work Group focused on standardizing interactions between cloud environments by developing specifications that deliver architectural semantics and implementation details to achieve interoperable cloud management between service providers and their consumers and developers

CRISP-DM Cross Industry Standard Process for Data Mining a data mining process model that describes commonly used approaches that expert data miners use to tackle problems

CRL Certificate Revocation List a list of certificates that have been revoked, and therefore should not be relied upon

CSS Cascading Style Sheets is a style sheet language used to describe the presentation semantics (the look and formatting) of a document written in a markup language

CSV Comma-Separated Values a set of file formats used to store tabular data in which numbers and text are stored in plain-text form that can be easily written and read in a text editor

CWM Common Warehouse Metamodel a specification for modeling metadata for relational, non-relational, multi-dimensional, and most other objects found in a data warehousing environment

CeCILL CEA CNRS INRIA Logiciel Libre is a free software license adapted to both international and French legal matters, in the spirit of and retaining compatibility with the GNU General Public License

CellML Cell Markup Language is an XML based markup language for describing mathematical models

DICOM Digital Imaging and Communications in Medicine - a standard for handling, storing, printing, and transmitting information in medical imaging

DTMF Distributed Management Task Force brings the IT industry together to collaborate on the development, validation and promotion of systems management standards

EHR Electronic health record is an evolving concept defined as a systematic collection of electronic health information about individual patients or populations

EULA End-user licensing agreements An EULA is a legal contract between the manufacturer and/or the author and the end user of an application

EUPL European Union Public License the first European Free/Open Source Software (F/OSS) license

FMA F Foundational Model of Anatomy it is concerned with the representation of classes or types and relationships necessary for the symbolic representation of the phenotypic structure of the human body in a form that is understandable to humans and is also navigable, parseable and interpretable by machine-based systems

FieldMLField Markup Language is an XML based markup language for describing field models

GAS Gridge Authorization Service provides functionality that would be able to fulfill most authorization requirements of grid computing environments

GCM Galois/Counter Mode a mode of operation for symmetric key cryptographic block ciphers that has been widely adopted because of its efficiency and performance

GEM Guideline Elements Model an XML-based guideline document model that can store and organize the heterogeneous information contained in practice guidelines

GNU Gnu's Not Unix is a Unix-like computer operating system developed by the GNU project, ultimately aiming to be a "complete Unix-compatible software system" composed wholly of free software.

GO Gene Ontology is a major bioinformatics initiative with the aim of standardizing the representation of gene and gene product attributes across species and databases

GPL General Public License is the most widely used free software license, originally written by Richard Stallman for the GNU Project

GridFTP GridFTP is an extension of the standard File Transfer Protocol (FTP) for use with Grid computing

HL7 Health Level Seven is an all-volunteer, non-profit organization involved in development of international healthcare informatics interoperability standards

HMAC Hash-based Message Authentication Code a mechanism for message authentication using cryptographic hash functions

HTML Hypertext Markup Language is the predominant markup language for web pages. HTML elements are the basic building-blocks of webpages.

HTTPS Hypertext Transfer Protocol Secure is a combination of the Hypertext Transfer Protocol (HTTP) with SSL/TLS protocol to provide encrypted communication and secure identification of a network web server

IBM International Business Machines

ID-FF Liberty Identity Federation Framework an approach for implementing a single sign-on with federated identities based on commonly deployed technologies

ID-WSF Liberty Identity Web Services Framework a framework for identity-based web services in a federated network identity environment

IEC International Electrotechnical Commission is the world's leading organization that prepares and publishes International Standards for all electrical, electronic and related technologies

IEEE Institute of Electrical and Electronics Engineers is a non-profit professional association headquartered in the United States that is dedicated to advancing technological innovation and excellence

IETF Internet Engineering Task Force a large open international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture and the smooth operation of the Internet

IHE Integrating the Healthcare Enterprise - an initiative by healthcare professionals and industry to improve the way computer systems in healthcare share information

IPSec Internet Protocol Security a protocol suite for securing Internet Protocol (IP) communications by authenticating and encrypting each IP packet of a communication session

ISBN International Standard Book Number is a unique numeric commercial book identifier based upon the 9-digit Standard Book Numbering (SBN) code created by Gordon Foster

ISO International Organization for Standardization is an international standard-setting body composed of representatives from various national standards organizations

InSilicoML InSilico Markup Language is a markup language that can explicitly describe the multi-level hierarchical structures of the physiological functions in mathematical models

JDMP Java Data Mining Package an open source Java library for data analysis and machine learning

JPEG Joint Photographic Experts Group is a commonly used method of lossy compression for digital photography

JSDL Job Submission Description Language is an extensible XML specification from the Global Grid Forum for the description of simple tasks to non-interactive computer execution systems

KNIME Konstanz Information Miner a user-friendly and comprehensive open source data integration, process, analysis and exploration platform

LGPL Lesser General Public License is a free software license published by the Free Software Foundation

LOINC Logical Observation Identifiers Names and Codes is a database and universal standard for identifying medical laboratory observations

MAGE-ML Microarray and Gene Expression - Markup Language markup language format for the representation of gene expression data from microarrays to facilitate the exchange of information between different data systems

MAGE-OM Microarray and Gene Expression - Object Model data exchange model for the representation of gene expression data from microarrays to facilitate the exchange of information between different data systems

MAGE-TAB Microarray and Gene Expression - Tabular tabular format for the representation of gene expression data from microarrays to facilitate the exchange of information between different data systems

MIAME Minimum Information About a Microarray Experiment needed to enable the interpretation of the results of the experiment unambiguously and potentially to reproduce the experiment

MIASE Minimal Information About a Simulation Experiment common set of information a modeller needs to provide in order to enable the execution and reproduction of a numerical simulation experiment, derived from a given set of quantitative models

MIASE Minimum Information About a Simulation Experiment is an effort to list the common set of information a modeller needs to provide in order to enable the execution and reproduction of a numerical simulation experiment, derived from a given set of quantitative models.

MIBBI Minimum Information for Biological and Biomedical Investigations maintains a web-based, freely accessible resource for ”Minimum Information” checklist projects, providing straightforward access to extant checklists (and to complementary data formats, controlled vocabularies, tools and databases), thereby enhancing both transparency and accessibility

MIT MIT License is a free software license originating at the Massachusetts Institute of Technology

ML Markup Language is a modern system for annotating a text in a way that is syntactically distinguishable from that text

MOF MetaObject Facility the foundation of OMG's industry-standard environment where models can be exported from one application, imported into another, transported across a network, stored in a repository and then retrieved, rendered into different formats

MPL Mozilla Public License is a free and open source software license

MS Microsoft is an American public multinational corporation headquartered in Redmond, Washington

MTOM Message Transmission Optimization Mechanism is the W3C Message Transmission Optimization Mechanism, a method of efficiently sending binary data to and from Web services

MedLEE Medical Language Extraction and Encoding system System to extract, structure, and encode clinical information in textual patient reports so that the data can be used by subsequent automated processes

NeuroML Neuro Markup Language is an XML (Extensible Markup Language) based model description language that aims to provide a common data format for defining and exchanging models in computational neuroscience

OASIS Organization for the Advancement of Structured Information Standards a not-for-profit consortium that drives the development, convergence and adoption of open standards for the global information society

OBO Open Biomedical Ontologies is an effort to create controlled vocabularies for shared use across different biological and medical domains

OGSA-BES Open Grid Services Architecture - Basic Execution Services defines Web Services interfaces for creating, monitoring, and controlling computational entities such as UNIX or Windows processes, Web Services, or parallel programswhat we call activities within a defined environment

OGSA-DAI Open Grid Service Architecture-Data Access and Integration allows data resources (e.g. relational or XML databases, files or web services) to be federated and accessed via web services on the web or within grids or clouds. Via these web services, data can be queried, updated, transformed and combined in various ways.

OSI Open Source Initiative is an organization dedicated to promoting open source software

OS Operating System is a set of programs that manages computer hardware resources, and provides common services for application software

OWL-S Ontology Web Language for web Services an ontology of services to discover, invoke, compose, and monitor Web resources offering particular services and having particular properties

OWL Web Ontology Language is a family of knowledge representation languages for authoring ontologies.

OpenID Open Identity provider of web-based SSO services

PAOS Reverse HTTP Binding for SOAP a binding that enables HTTP clients to expose services using the SOAP protocol, where a SOAP request is bound to a HTTP response and vice versa

PATO PATO an ontology of phenotypic qualities, intended for use in a number of applications, primarily defining composite phenotypes and phenotype annotation.

PHP PHP: Hypertext Preprocessor is a general-purpose server-side scripting language originally designed for web development to produce dynamic web pages

PKIX Public-Key Infrastructure Working Group was established in the fall of 1995 with the goal of developing Internet standards to support X.509-based Public Key Infrastructures

PMML Predictive Model Markup Language an XML-based language which provides a way for applications to define statistical and data mining models and to share models between PMML compliant applications

PNG Portable Network Graphics is a bitmapped image format that employs lossless data compression

POST POST is one of many request methods supported by the HTTP protocol used by the World Wide Web

RAD Rapid application development is a software development methodology that uses minimal planning in favor of rapid prototyping

RDF Resource Description Framework is a family of World Wide Web Consortium (W3C) specifications originally designed as a metadata data model

REST Representational state transfer is a style of software architecture for distributed hypermedia systems such as the World Wide Web

RFC Request for Comments is a memorandum published by the Internet Engineering Task Force (IETF) describing methods, behaviors, research, or innovations applicable to the working of the Internet and Internet-connected systems

RICORDO RICORDO is focused on the study and design of a multiscale ontological framework in support of the Virtual Physiological Human community to improve the interoperability amongst its Data and Modelling resources

RIM Reference Information Model is the cornerstone of the HL7 Version 3 development process and an essential part of the HL7 V3 development methodology

SAML Security Assertion Markup Language a standard, XML-based framework for creating and exchanging security information between online partners

SAS Business analytics software and service developer, and independent vendor in the business intelligence market

SAWSDL Semantic Annotations for WSDL defines mechanisms using which semantic annotations can be added to WSDL components

SBML System Biology Markup Language is a representation format, based on XML, for communicating and storing computational models of biological processes

SDTM Study Data Tabulation Model CDISC defining a standard structure for human clinical trial (study) data tabulations that are to be submitted as part of a product application to a regulatory authority

SED-ML Simulation Experiment Description Markup Language an XML-based format for encoding simulation experiments, following the requirements defined in the MIASE guidelines

SHA Secure Hash Algorithm a number of cryptographic hash functions published by the National Institute of Standards and Technology as a U.S. Federal Information Processing Standard

SLO Single Log-Out termination of a SSO action

SNIA Storage Networking Industry Association not-for-profit trade organization for companies and individuals in various sectors of the storage industry

SNOMED-CT Systematized Nomenclature of Medicine - Clinical Term is a systematically organised computer processable collection of medical terminology covering most areas of clinical information such as diseases, findings, procedures, microorganisms, pharmaceuticals etc

SOAP Simple Object Access Protocol is a protocol specification for exchanging structured information in the implementation of Web Services in computer networks

SOAP Simple Object Access Protocol a lightweight XML-based protocol for exchange of structured information in a decentralized, distributed environment

SOA Service-Oriented Architecture s a set of principles and methodologies for designing and developing software in the form of interoperable services

SPARQL SPARQL Protocol and RDF Query Language - query language for RDF

SQL Structured Query Language a standard language for accessing and manipulating databases

SSH Secure Shell is a network protocol for secure data communication, remote shell services or command execution and other secure network services between two networked computers that it connects via a secure channel over an insecure network

SSL Secure Sockets Layer a cryptographic protocol that provides communication security over the Internet, predecessor of TLS

SSO Single Sign-On a mechanism whereby a single action of user authentication and authorization can permit a user to access all computers and systems where he has access permission, without the need to enter multiple passwords

TCP/IP Transmission Control Protocol/Internet Protocol the first two networking protocols defined in the Internet Protocol Suite standard

TDD Test-driven development is a software development process that relies on the repetition of a very short development cycle.

TLS Transport Layer Security a cryptographic protocol that provides communication security over the Internet, successor of SSL

UML Unified Modelling Language a specification defining a graphical language for visualizing, specifying, constructing, and documenting the artifacts of distributed object systems

VDM Vienna Development Method is one of the longest-established Formal Methods for the development of computer-based systems

VPH-NoE Virtual Physiological Human - Network of Excellence is a project which aims to help support and progress European research in biomedical modelling and simulation of the human body

VPH Virtual Physiological Human is a methodological and technological framework that, once established, will enable collaborative investigation of the human body as a single complex system

WAV Waveform Audio File Format is a Microsoft and IBM audio file format standard for storing an audio bitstream on PCs

WS-* Web Services-* common prefix for the family of Web Services specifications

WSDL Web Services Description Language a way to describe the abstract functionalities of a service and concretely how and where to invoke it

WSMO Web Service Modelling Ontology ontology for describing Semantic Web Services

XFree86 A freely redistributable open-source implementation of the X Window System

XHTML Extensible HyperText Markup Language is a family of XML markup languages that mirror or extend versions of the widely-used Hypertext Markup Language (HTML), the language in which web pages are written

XML Extensible Markup Language - a format for encoding documents in machine-readable form, similar in syntax to HTML

XTS XEX-based Tweaked Codebook a mode of operation for cryptographic block ciphers

caBIG Cancer Biomedical Informatics Grid a virtual network of interconnected data, individuals, and organizations that work together to redefine how cancer research is conducted