Conceptualization and Evaluation of Interoperable and Modular IT-Framework Components for Exchanging Big Data Information Sets

Por

Philipp Urbauer

Orientador: Stefan Sauermann

Co-orientador: João Agostinho Batista Lacerda Pavão

Tese submetida à
UNIVERSIDADE DE TRÁS-OS-MONTES E ALTO DOURO
para obtenção do grau de
DOUTOR
em Engenharia Eletrotécnica.e de Computadores, de acordo com o disposto no
DR I série-Nº 151, Decereto-Lei Nº 115/2013 de 7 de agosto e no
Regulamento Geral dos Ciclos de Estudo Conducentes ao Grau de Doutor
DR, 2ª série-Nº 133 de 13 de julho de 2016

Conceptualization and Evaluation of Interoperable and Modular IT-Framework Components for Exchanging Big Data Information Sets

Por

Philipp Urbauer

Orientador: Stefan Sauermann

Co-orientador: João Agostinho Batista Lacerda Pavão

Tese submetida à
UNIVERSIDADE DE TRÁS-OS-MONTES E ALTO DOURO
para obtenção do grau de
DOUTOR
em Engenharia Eletrotécnica.e de Computadores, de acordo com o disposto no
DR I série-Nº 151, Decereto-Lei Nº 115/2013 de 7 de agosto e no
Regulamento Geral dos Ciclos de Estudo Conducentes ao Grau de Doutor
DR, 2ª série-Nº 133 de 13 de julho de 2016

Orientação Científica :

Stefan Sauermann

Professor do
Department of Life Science Engineering
University of Applied Sciences Technikum Wien

João Agostinho Batista Lacerda Pavão

Professor Auxiliar do
Departamento Engenharias da Escola de Ciências e Tecnologia
Universidade de Trás-os-Montes e Alto Douro

UNIVERSIDADE DE TRÁS-OS-MONTES E ALTO DOURO

Doutor em Engenharia Electrotécnica e de Computadores

Os membros do Júri recomendam à Universidade de Trás-os-Montes e Alto Douro a aceitação da tese intitulada " **Conceptualization and Evaluation of Interoperable and Modular IT-Framework Components for Exchanging Big Data Information Sets**" realizada por **Philipp Urbauer** para satisfação parcial dos requisitos do grau de **Doutor**.

Dezembro 2018

Presidente: **Professor Doutor Vitor Manuel de Jesus Filipe**
Professor Associado com Agregação da Escola de Ciências e Tecnologia
Universidade de Trás-os-Montes e Alto Douro

Vogais do Júri: **Professora Doutora Paula Christina Ribeiro Coutinho de Oliveira**
Professor Auxiliar da Escola de Ciências e Tecnologia
Universidade de Trás-os-Montes e Alto Douro

FH-Professor Doktor Johannes Martinek
FH-Professor do Department of Life Science Engineering
University of Applied Sciences Technikum Wien

Professor Doutor Nelson Fernando Pacheco da Rocha
Professor Catedrático do Departamento de Ciências Médicas
Universidade de Aveiro

Professorin Doktorin Christina B. Class

Professora Doutora da Informatics Study Group

Ernst-Abbe-Hochschule Jena University of Applied Sciences

Professor Doutor Luís José Calçada Torres Pereira

Professor Auxiliar da Escola de Ciências e Tecnologia

Universidade de Trás-os-Montes e Alto Douro

FH-Professor Doktor Stefan Sauermann

FH-Professor do Department of Life Science Engineering

University of Applied Sciences Technikum Wien

Conceituação e avaliação de componentes de estrutura de IT interoperáveis e modulares para troca de conjuntos de informações de Big Data

Philipp Urbauer

Submetido na Universidade de Trás-os-Montes e Alto Douro
para o preenchimento dos requisitos parciais para obtenção do grau de
Doutor em Engenharia Electrotécnica e de Computadores

Resumo — Sob o termo "digitalização 2.0", smartphones, tablets, relógios inteligentes e sensores de vestuário (wearables) são capazes de gerar enormes quantidades de dados no contexto da Internet das Coisas. As empresas e instituições de investigação estão a usar essas enormes quantidades de dados em diferentes temas de pesquisa. Há expectativas de que a combinação de dados de diferentes domínios, como por exemplo, saúde, meio ambiente ou transporte, possa levar a novas descobertas para melhorar vários aspetos da vida, tais como o melhor tratamento de doenças ou a melhoria da eficiência do atendimento na prestação de serviços. No entanto, um grande desafio é a diversidade de formatos de dados. Requisitos sintáticos e semânticos associados aos dados representam fatores de qualidade muito importantes para os tornar passíveis de troca e comparáveis. Este trabalho investiga a aplicabilidade de padrões e métodos de interoperabilidade do domínio da informática médica a dados de outros domínios, como transporte e meio ambiente, para promover o intercâmbio e melhorar a qualidade dos dados, utilizando padrões internacionais. Para tal foram analisadas plataformas de dados abertas e foram recolhidas e selecionadas de acordo com critérios bem definidos normas de interoperabilidade médica e tecnologias relacionadas. Em consequência, um conceito denominado "Interoperable BDIS Directory" (IBD) -Profile foi desenvolvido para a troca de conjuntos de grandes quantidades de informação - Big Data Information Sets (BDIS). O perfil do IBD é baseado nos serviços web HL7 FHIR e RESTful e inclui descrições de processo e definições de recursos HL7 FHIR. Em três estudos

de viabilidade técnica (transmissão de dados de aplicações de acompanhamento de exercício físico, dados de exposição ao pólen e dados de transportes públicos) foram implementados protótipos e feitos com sucesso testes de conformidade por ferramentas de validação HL7 FHIR. Por fim, a verificação do conceito desenvolvido foi feita através do desempenho de uma revisão de especialistas de acordo com o IEEE 1028. A revisão de especialistas confirma que o conceito desenvolvido é relevante e que o Perfil IBD é um primeiro passo bem-sucedido para introduzir a interoperabilidade com os objetivos desejados. No entanto, é necessário continuar a investigação do conceito no que diz respeito à integração de requisitos de streaming, bem como à melhoria da interligação de fontes e consumidores de dados distribuídos.
Palavras Chave: Interoperabilidade, Normalização, Open Data, Healthcare, Transportes e Ambiente, Health Level 7 (HL7), Integrating the Healthcare Enterprises (IHE)

Conceptualization and Evaluation of Interoperable and Modular IT-Framework Components for Exchanging Big Data Information Sets

Philipp Urbauer

Submitted to the University of Trás-os-Montes and Alto Douro
in partial fulfillment of the requirements for the degree of
Doctor of Electrical and Computer Engineering

Abstract — Under the term "digitization 2.0" smartphones, tablets, smart watches and wearable sensors are generating huge amounts of data in context of the Internet of Things. Companies and research institutions are investigating and using these massive amounts of data in terms of research. There are expectations that combining data from different domains like for example healthcare, environment or transport, might lead to new findings for improving several aspects of life like better treatment of diseases or improving efficiency of the care path. However, a huge challenge is the diversity of data formats. Related syntactic and semantic requirements represent very important quality factors to make data exchangeable and comparable. This work investigates the applicability of interoperability standards and methods from the medical IT domain to data from other domains like transport and environment, to foster the exchange and improve data quality by using international standards. Hence, in this work open data platforms and formats were analyzed, medical interoperability standards and related technologies were collected and selected according to well defined criteria. Based on that, a concept called "Interoperable BDIS Directory" (IBD)-Profile was developed for the exchange of Big Data Information Sets (BDIS). The IBD-Profile is based on HL7 FHIR and RESTful web services and includes process descriptions and HL7 FHIR resource definitions. In three technical feasibility studies (transmitting fitness tracker data, pollen exposure data and public transport data) prototypes were implemented and successfully tested with conformance tests by using HL7 FHIR validation tools.

Finally, verification of the developed concept was done through performance of an experts review according to IEEE 1028. The experts review confirms the developed concept to be meaningful and that the IBD-Profile is a successful first step to introduce interoperability for this purpose. However, further investigations of the concept should be done regarding integration of streaming-requirements as well as improved inter-connection of distributed data sources and sinks.
Key Words: Interoperability, Standardization, Open Data, Healthcare, Transport and Environment, Health Level 7 (HL7), Integrating the Healthcare Enterprises (IHE)

Acknowledgments

My gratitude goes to Professor Stefan Sauermann for his rigidity in hardening my motivation during my study and his support all along my path over the last years.

I would like to thank Professor Joao Pavao for his warmhearted kindness and valuable support in all concerns from computer science over geography to history and cultural aspects.

Furthermore, I would like to express my gratitude to Professor Alexander Mense, who supported my path over all the years from the background with patience and farsightedness.

Additionally, I want to thank Professor Luis Pereira for his wisdoms which are much more far reaching than the Ph.D. context could ever be, but always lead to a solution and Professor Rute Bastardo sharpening my visualization on the really important things.

I would like to take the opportunity to thank the team of office A4.28 at the UAS Technikum Wien, Matthias Frohner, Veronika David, Mathias Forjan, and Birgit Pohn. Thank you for countless times of fruitful discussions and common sharing of pain in times, were the suffer score reached maximum level and could severely not be counted as an outlier.

Last but not least, I want to thank my family in Austria & Hungary. Especially my father and mother for teaching me thoughtful attitudes towards life.
However, no medical device fulfills quality requirements without ISO 13485 and no house is durable without a stable foundation. Thus, my deepest gratitude goes to my wife Nikolett.

"Tövises az út a csillagokig."

To all a heartfelt thank you!
A todos, um sincero obrigado!
Mindenkinek köszönöm!
Ein herzliches Danke an euch!

Vienna/Vila Real, Philipp Urbauer
December 20, 2018

Table of Contents

Index of Tables

Table of Figures

Glossary, Acronyms and Abbreviations

Glossary

Actor — Actors are pieces of software i.e. modules, which provide necessary functionality to exchange information based on IT-standards. Actors have specific purposes, for example building and storing documents or searching for documents etc. and communicate via "Transactions". Actors and their functionality are defined in "Integration Profiles" together with "Transactions".

ATNA — A technical profile from IHE, describing the integration of audits and audit trails in medical information systems. Additionally, authentication on a software component level i.e. module level is described by the profile. It is the basic security profile of IHE, which has to be implemented when developing XD*-based software solutions.

Big Data Information Sets — The term big data refers in general to huge amounts of data. The term Big Data Information Sets (BDIS) was chosen, as for this work only excerpts of data could be taken fulfilling certain quality requirements. These requirements are stated in the respective chapters.

Therefore, BDIS is defined as: "An enclosed set of multi-domain data (e.g. domains of health, environment, transport etc.), related to a specific point in time or time span, including meta-data to describe the context of the data e.g. purpose of data acquisition, size, etc.."

Connect-a-thon — The Connect-a-thon is a testing event carried out by IHE International, but also the affiliate organizations like IHE Europe or IHE Asia-Oceania, to provide a community based approach for interoperability conformance testing. This includes no-peer and peer-to-peer tests under the control of test-specialists called Monitors. These are technical experts having long time experiences in the filed of IT and testing. They visually observer processes and perform conformance tests like exchanging medical documents or authentication and encryption processes and much more.

DEC — An IHE profile describing the integration process of medical device data, like intensive care units or blood pumps data, into medical IT infrastructure. The profile is located in the Patient Care Device (PCD) technical framework and uses HL7 V2 and V3 as a communication standard.

Integration Profile — Integration Profiles are technical specifications based on international IT standards to solve real world scenarios. Each profile includes descriptions of example scenarios, an Actor & Transaction diagram describing the related processes, information formats and requirements through standards, as well as guidance regarding related security & privacy requirements.

IUA — Is a profile described in the IT infrastructure technical framework of IHE and focuses on the management of security tokens used for authorization purposes related to RESTful based web services. It is strongly connected to OAuth.

OID — Object Identifier (OID) are internationally used and world wide unique identifiers. OIDs are issued according to ISO/IEC 9834-1. In Austria OIDs for eHealth are issued by the OID-Portal Austria.

Technical Framework — Technical Frameworks are collections of Integration Profiles, which are defined for specific domains, like IT-Infrastructure, Radiology or Laboratory.

Transaction — Transactions are used for communicating information between "Actors". This includes the specification of technologies to be used, for example SOAP based web-services together with WS-Security etc., as well as requirements of structuring this data and adding semantics to it.

XACML — Is an XML based standard for the technical implementation of policies in XML format. This includes a data model for defining XML-policies, but additionally adds actors and a protocol for policy decision making processes and enforcement.

XDR — Describes information exchange of medical data similar to XDS using SOAP based web services, but with a focus on point-to-point communication in case no high sophisticated medical IT infrastructure is available.

XDS — A profile from IHE, describing the content agnostic exchange of health information between enterprises in healthcare (e.g. hospitals etc.) based on web services using SOAP and related WS-* security technologies. The profile describes and explains the processes, requirements, applicable and recommended technologies. HL7 CDA is strongly connected to this profile as it defines interoperability for medical documents send by IHE XDS based systems.

XUA — Is a profile in the IT infrastructure technical framework of IHE and focuses on enabling SOAP based web services (e.g. IHE XDS or XDR) the use of authorization data in form of e.g. SAML assertions holding relevant authorization criteria.

List of Acronyms

Initials	Expanded
ABAC	*Attribute-Based-Access-Control*
AHD	*Application Hosting Device*
ANSI	*American National Standards Institute*
ATNA	*Audit Trail and Node Authentication*
BAN	*Body Area Network*
BDD	*Big Data Domain*
BDIS	*Big Data Information Set*
BLE	*Bluetooth Low Energy*
BPPC	*Basic Patient Privacy Consent*
BT	*Bluetooth*
CDA	*Clinical Document Architecture*
CEN	*Comité Européen de Normalisation*
CIA-Triad	*Confidentiality, Integrity and Availability - Triad*
CKAN	*Comprehensive Knowledge Archive Network*
CPAPSS	*Chronic Pollen Allergy Patient Support System*
CT	*Consistent Time*
DEC	*Device Enterprise Communication*
DICOM	*Digital Imaging and Communications in Medicine*
ECG	*Electrocardiography*
EHR	*Electronic Health Record*
EIF	*European Interoperability Framework*
EIRA	*European Interoperability Reference Architecture*
epSOS	*European Patients Smart Open Services*
ETSI	*European Telecommunications Standards Institute*
EU	*European Union*

Initials	**Expanded**
EUA	*Enterprise User Authentication*
FHIR	*HL7 Fast Healthcare Interoperability Resources*
GDPR	*General Data Protection Regulation*
HDP	*Health Device Profile*
HIPAA	*Health Insurance Portability and Accountability Act*
HIS	*Healthcare Information System*
HL7	*Health Level Seven*
HL7 V2	*HL7 Version 2*
HL7 V3	*HL7 Version 3*
HTML	*Hyper Text Markup Language*
HTTP(s)	*Hypertext Transfer Protocol (Secure)*
IBD	*Interoperable BDIS Directory*
ICD	*International Classification of Diseases*
ICP	*IHE Certified Professional*
ICT	*Information and Communication Technology*
IEEE	*Institute of Electrical and Electronics Engineers*
IHE	*Integrating the Healthcare Enterprises*
IMRAD	*Introduction, Materials & Methods, Results, Discussion and Conclusion*
IoT	*Internet of Things*
ISA^2	*Interoperability Solutions for European Public Administrations 2*
ISO	*International Organization for Standardization*
IT	*Information Technology*
ITS	*Intelligent Transport Systems*
IUA	*Internet User Authorization*
JSON	*Java Script Object Notation*

Initials	**Expanded**
LAN	*Local Area Network*
LOINC	*Logical Observation Identifiers Names and Codes*
MDER	*Medical Device Encoding Rules*
MLLP	*Minimal Low Layer Protocol*
MS	*Multiple Sclerosis*
MSH	*Message Header*
MUV	*Medical University of Vienna*
NTP	*Network Time Protocol*
OASIS	*Organization for the Advancement of Structured Information Standards*
OBX	*Observation Segment of OBR*
OBR	*Observation Request Group*
OECD	*Organization for Economic and Co-operation and Development*
OSI-Model	*Open Systems Interconnection Model*
PAN	*Personal Area Network*
PCHA	*Personal Connected Health Alliance*
PDR	*Pollen Data Requester*
PHD	*Personal Health Device*
PHDSC	*Public Health Data Standards Consortium*
PHR	*Personal Health Record*
PID	*Patient Identifier*
QRPH	*Quality, Research and Public Health*
RBAC	*Role-Based-Access-Control*
ReEIF	*Refined eHealth European Interoperability Framework*
REST	*Representational State Transfer*
RIM	*Reference Information Model*

Initials	**Expanded**
SAML	*Security Assertion Markup Language*
SDOs	*Standards Developing Organizations*
SNOMED-CT	*Systematized Nomenclature of Medicine - Clinical Terms*
SOA	*Service Oriented Architectures*
STU	*Standard for Trial Use*
TCP/IP	*Transmission Control Protocol/Internet Protocol*
TISA	*Traveller Information Service Association*
TPEG	*Transport Protocol Experts Group*
UCUM	*Unified Codes for Units of Measure*
UML	*Unified Modeling Language*
URL	*Uniform Resource Locator*
VHA	*Vienna Hospital Association*
WHO	*World Health Organization*
WLAN	*Wireless Local Area Network*
X73	*ISO/IEEE 11073 Group of Standards*
X73-OEP	*ISO/IEEE 11073-20601 Optimized Exchange Protocol*
XDR	*Cross-Enterprise Document Reliable-Interchange*
XDS	*Cross-Enterprise Document Exchange*
XD*	*Cross-Enterprise Family of Profiles (XDS, XDR, XDM, XD-Content Profiles, etc.)*
XML	*Extensible Markup Language*
XSLT	*Extensible Stylesheet Language Transformation*
XUA	*Cross-Enterprise User Assertion*

Lisit of Abbreviations

Abbreviation	**Significance**
e.g.	exempli gratia, for example
et al.	et aliae, and the other persons)
etc.	et cetera, and so on
i.e.	id est, that means
vid.	vide, see also
vs.	versus, comparison

Introduction

Digitization i.e. the process of conversion of text, pictures or sound to a digital format which can be processed by a computer (Oxford Dictionaries, 2018), is a process that accompanies society since years. At present our society is on the edge of digitization 2.0 and technologies like the Internet of Things (IoT) or blockchain strongly influence this transformation processes (Helbing, 2017). Systems and devices in the context of mobile computing i.e. each smartphone, tablet, smart watch or any wearable sensor can be seen as a highly productive source of generating data. Examples are crowd-sourced approaches like the usage of smartphone ECG applications for the diagnose of health incidents like syncope (Nyotowidjojo et al., 2016) or the use of the smartphone-based data to overcome the lack of missing infrastructure for floating car systems (i.e. providing individual sensor data from cars for to decrease traffic (Briante et al., 2014). The EU-Project COBWEB (COBWE-Project, 2016) focused on investigating and strengthening the potential of crowd-sourced environmental data focusing on the quality of data, its bias and risks. These examples show the wide range of application of these technologies and each industry and different public services generate massive amounts of data. Hence, there is strong evidence that data will further grow through digitalization 2.0.

The expression big data is frequently used in this context. There are different definitions of the term "big data" as shown in (Press, 2014). Summarizing this, the term refers to extremely large volumes of data generated and stored, which can't be handled with common tools for storage and analysis. As the big data revolution does not simply refer to the quantity of the data, but moreover to the ability to handle and interpret the data, the management of Big Data Information Set(s) (BDIS) in an efficient way is of high interest to improve big data analytic outcomes and gain insights into possibly new hidden values (Shaw, 2014; McKinsey Global Institute, 2011).

Furthermore, it is expected to reveal new findings by combining different data silos from different data categories i.e. domains. Examples could be improving crime investigation (Open Data Bits, 2014) or increasing insights on the understanding of impacts to different diseases like Multiple Sclerosis (MS). There is first evidence that regional environmental influences have an impact on the frequency of MS-relapses (Spelman et al., 2014). To combine collected health and environmental data sets from several geographically different EU regions may bring enormously valuable support for such disease research. Another hypothesis in this context could be that the combination of transportation and environmental data, like customer movement streams, in relation with health data, like flu outbreak climax, may be used to steer and optimize patient flows in hospitals. This may be used to improve quality and efficiency of treatment for patients as well as similarly decreasing costs (Drazen and Rhoads, 2011). However, these are just a few examples for the combinations of large piles of data from different domains, which may lead to several additional values for society. The group around Ahmed et al. (2017) investigated the role of big data in IoT and come to the conclusion that it is of high importance to liberate data from data silos to provide support for cross-domain approaches as research topics of the future will likely focus on the combinatorial approach. Moreover, important research challenges are to face the diversity of data, its semantics and interconnectivity as these are needed for increasing data quality, reliability and usability (Ahmed et al., 2017; Günther et al., 2017).

As The term interoperability is concerned with the above stated requirements and is recognized as a very important factor to improve efficiency and sustainability of data, as indicated by (European Commission, 2018c). This is shown in several eHealth related projects regarding Electronic Health Records (EHR), Personal Health Records (PHR), medical IT systems and Personal Health Device (PHD) communication for interchanging medical information. Examples are the national Electronic Health Record in Austria (ELGA, 2018) as well as the project European Patients Smart Open Services (epSOS) (EpSos-Project, 2016) funded by the EU, interconnecting national EHRs in the EU. Industry, national and international governments as well as scientific societies see benefits of standardized interoperability on the sustainability and cost savings regarding IT systems as outlined by the European Interoperability Framework (EIF) (European Commission, 2018e).

1.1 Motivation and Objectives

The intention of this thesis is to explore the combination of Big Data Information Sets (BDIS), while also developing a structured method to combine these data sources in a standardized way, to introduce interoperability for the combination of different data sources efficiently. Thus, the main research question is:

"Are interoperability standards and methods from the medical information technology domain applicable to other domains to support exchange and interpretation of Big Data Information Sets?"

The thesis is a conclusive study, with five main objectives as described in the following. The first objective is to conduct an investigative analysis to get an overview about the open data platforms and their formats. This is followed by objective two, which is a selection process for applicable medical interoperability standards based on well defined criteria. Subsequently, objective three is to establish a big data business domain overview and the conceptualization of re-usable, modular

and combinable IT framework components for standardized exchange of BDIS. The fourth objective is the execution of feasibility studies in the domains of health, environment and transport as a prove of concept of the framework components and especially the data syntax and semantics. This includes technical validation i.e. conformance testing by well defined state-of-the-art testing procedures using validation tools. Finally, the last objective is to evaluate the concept by an international experts review.

1.2 Limits and Scope of the Thesis

As stated in the first paragraphs of this thesis, big data is ubiquitous and independent of the different domains. Hence, providing a meaningful context in this thesis the following points shall narrow the focus of the study and define the scope and limits:

- Big data analysis is out of scope as the focus lies on communication interfaces for data exchange between software components
- Three scenarios for health, environment and transport are selected to study feasibility
- A special focus lies on the syntax and semantics of the data exchange and the protocol
- The prototypes are tested in lab environments and the access to data is limited by the data providers. Hence, the data is selected according to defined criteria (i.e. availability, reliability etc.).
- The performance of the components i.e. data throughput and load balancing is out of focus in this work.

Based on the previously stated facts that the big data revolution does not simply refer to the quantity of the data, but moreover to the ability to efficiently manage and

interpret BDIS i.e. providing suitable syntactic structures and semantics through a data model, the following definition of BDIS is used in the context of this thesis:

"An enclosed set of multi-domain data (e.g. domains of health, environment, transport etc.), related to a specific point in time or time span, including meta-data to describe the context of the data e.g. purpose of data acquisition, size, etc.."

1.3 Scientific Publications

Partial results and preliminary investigations were published in advance of finalizing this thesis. Hence, the contributions are listed here:

- Urbauer, Philipp; Frohner, Matthias; Forjan, Mathias; Pohn, Birgit; Sauermann, Stefan; Mense, Alexander (2012). A Closer Look on Standards Based Personal Health Device Communication: A Résumé over Four Years Implementing Telemonitoring Solutions. European Journal for Biomedical Informatic (EFMI-Journal-2012), Vol. 8(2012), Issue 3, pp. 65-70, ISSN 1801-5603 (print)

- Urbauer, Philipp; Herzog, Juliane; Pohn, Birgit; Forjan, Mathias; Sauermann, Stefan (2014). Certification Programs for eHealth - Status Quo eHealth. eHealth Summit Austria - Health Informatics meets eHealth (Conference-2014), 22-23 May Vienna, Austria

- Philipp Urbauer, Stefan Sauermann, Matthias Frohner, Mathias Forjan, Birgit Pohn, Alexander Mense (2015). Applicability of IHE/Continua components for PHR systems: Learning from experiences. Computers in Biology and Medicine (Elsevier-Journal-2015), Vol 59, Pages 186-193, ISSN 0010-4825

- Urbauer, Philipp; Maximilian, Kmenta; Frohner, Matthias; Mense, Alexander; Sauermann, Stefan (2017). Propose of Standards based IT Architecture to enrich the Value of Allergy Data by Telemonitoring Data. eHealth Summit

Austria - Health Informatics Meets eHealth (Conference-2017), 23-24 May Vienna, Austria

- Philipp Urbauer, Matthias Frohner, Veronika David, Stefan Sauermann (2018). Wearable Activity Trackers Supporting Elderly Living Independently: A Standards based Approach for Data Integration to Health Information Systems. Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion (DSAI-Conference-2018), 20-22 June Thessaloniki, Greece

1.4 Organization of the Thesis

The basic structure of this thesis is according to IMRAD (Introduction, Materials & Methods, Results, Discussion and Conclusion). The chapter 2 "Background") provides an overview about SDO's, interoperability standards and security basics from the medical domain as well as state-of-the-art work related to this thesis. In chapter 3 "Materials & Methods" the detailed process to reach the five defined objectives as well as uthe used infrastructure and components are described. The chapter 4 "Results" includes the results from the data format analysis and the standards selection process as well as the the developed concept containing a business domain overview and the conceptualization of the IT framework components. Furthermore it describes the results of the executed feasibility studies, its technical validation and the results of the experts review. All results are subsequently discussed in chapter 5 "Discussion" and the final conclusion including future perspectives finalizes this thesis in chapter 6 "Conclusion".

2 Background

Interoperability describes the ability of heterogeneous IT applications to exchange data accurately, effectively and consistently in a reliable process and plays an important role in healthcare (Jardim, 2013). The focus lies on the IT components interfaces. This for example includes the transmission technologies like Wi-Fi, ZigBee or Bluetooth (BL), but reaches far more deeper into the Application layer of the Open Systems Interconnection model (OSI model). That is the case as the data itself is structured and made semantically interpretable at this layer in typical Internet based communication procedures in healthcare. Hence, it is a technical quality measure to increases the possibilities of cooperation between healthcare professionals through reliable exchange of medical information (Iroju et al., 2013). In accordance to Tolk et al. (2013) interoperability has seven levels:

- Level 0: No Interoperability: Closed systems which are not interacting with others.

- Level 1: Technical Interoperability: A common infrastructure along with a communication protocol is defined.

- Level 2: Syntactical Interoperability: Adds a specific format to structure the

data which is communicated between the systems.

- Level 3: Semantic Interoperability: At this level the meaning of the data is included in the dataset exchanged e.g. by usage of codes from code-lists or value-sets.

- Level 4: Pragmatic Interoperability: This level adds the context of usage of the data i.e. the systems communicating know how the data is used.

- Level 5: Dynamic Interoperability: The communicating systems are able to understand and react on certain effects of operations during the communication process.

- Level 6: Conceptual Interoperability: Includes the technical specification of conceptual models and its proper description, independent of the concrete implementations.

Although level 1 is defined as technical interoperability abstracted from the other levels, all levels in this definition are strongly focusing technical characteristics and therefore could be defined as technical interoperability together. However, when taking a closer look to the ISA^2 program from the EU, which fosters the "development of digital solutions that enable public administrations, businesses and citizens in Europe to benefit from interoperable cross-border and cross-sector public services!" (European Commission, 2018a), a much broader view and impact of interoperability is revealed. The new EIF (European Commission, 2018e) is part of the Communication (COM(2017)134), which supersedes the former version from 2010 and provides guidance for integration of interoperability in digital public services. The actual version includes 47 recommendations due to several changes through new EU policies. The EIF defines four levels of interoperability to support an interoperability-by-design pattern. Furthermore, the EIF, the Interoperability Action Plan and the European Interoperability Architecture (EIRA) are parts of the interoperability governance on an European level. The four levels of interoperability from EIF(European Commission, 2018d) are:

- Legal Interoperability: Through the diversity of national legal frameworks, policies and strategies of the EU member states, it has to be ensured that organizations can work together. Interoperability checks should be executed to identify interoperability barriers like geographical restrictions, different data license models and restrictions to use specific technologies etc.. In case of new legislation, requirements and impacts of ICT shall be considered as early as possible and along the whole process. However, data protection legislation need to be fulfilled on an EU as well as on all applied member states levels.

- Organizational Interoperability: Targets the integration or alignment of business processes and the alignment in accordance with responsibilities to achieve commonly agreed goals and benefits. These should be documented with accepted modeling approaches. Furthermore organizational relationships need to be clearly defined and formalized regarding establishment and operation of services.

- Semantic Interoperability: Refers to ensure the format and meaning of data, which is exchanged between the communicating parties. This includes syntactic interoperability, for definition of grammar and format of the data, and semantic interoperability i.e.the meaning of data and its relation through the definition of terminologies and schemata. The latter assures that all communication parties understand the data exchanged in the same way. The proper installation of an information management strategy for management of meta-data, master data and reference data should minimizes the probability of duplicate and fragmented data. The development of specifications for formats, terminologies and processes shall be driven in national or European community to overcome the challenges of linguistic, cultural, legal and administrative requirements of the different member states.

- Technical Interoperability: Targets interface specifications, data presentation and exchange, communication protocols from security perspective, which are used by services and applications to link systems. As of historically driven bottom-up development of legacy services in public administration, these

systems were build to fulfill a specific aim, which resulted in fragmented ICT landscapes. Therefore open specifications shall be used when and wherever possible.

These definitions show the complexity of interoperability. However, in the medical domain the Refined eHealth European Interoperability Framework (ReEIF) (eHealth Network, 2015) was developed in 2015. In this work the EIF definitions are refined as they define six out of four levels of interoperability as shown in 2.1.

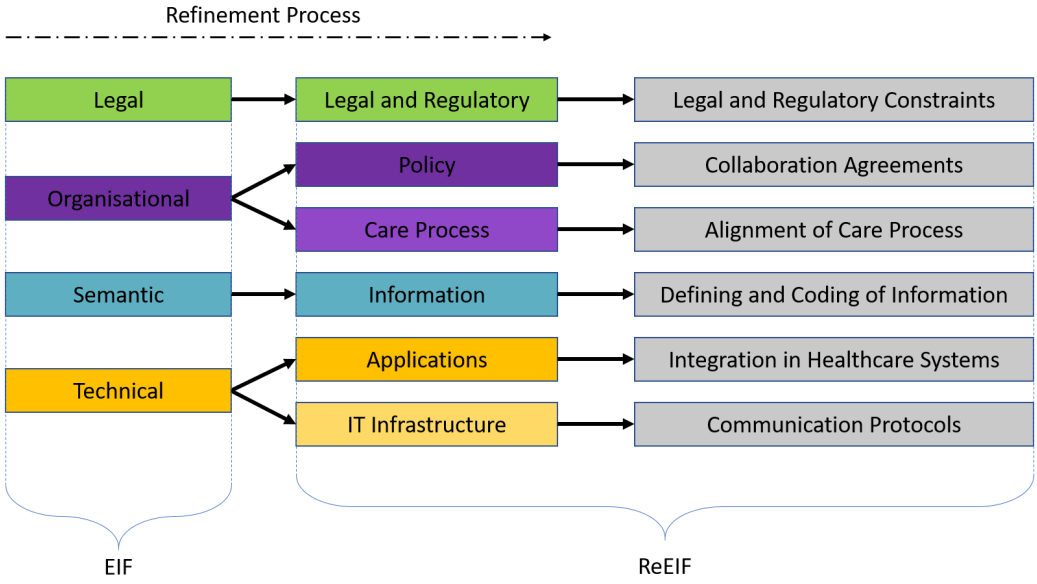


Figure 2.1 – Shows the EIF model, its refinement process and the ReEIF model in accordance to eHealth Network (2015)

The left column shows the four layers of interoperability according the EIF i.e. legal, organizational, semantic and technical. According the refinement process these four levels were extend in six level, where organizational is split into policy and care process levels and technical into applications and IT infrastructure. The processes to split these levels, were justified by the fact that these levels have different actors and responsibilities and from a technical interoperability perspective, different classes of standards.

The six levels of interoperability together with the related gray boxes represents the ReEIF model. The "Legal and Regulatory level", legislation and regulations specify the scope and limits of interoperability across boarders and also within countries of regions. The "Policy level" focuses on the organizational collaboration and its agreements and governance. The "Care Process level" refers to integrate and align processes between collaborating organizations to realize integrated care pathways and shared workflows. The "Information level" represents the semantic level i.e. data models, description of data, terminologies with codes and its linking to data. The "Applications level" targets the communications standards for data exchange and therefore the import and export in Healthcare Information Systems (HIS). Finally "IT Infrastructure level" refers to communication and network protocols, the storage, backup and database engines eHealth Network (2015).

The previously stated definitions of interoperability demonstrate the complexity, which hides behind this concept. Therefore, the main categories of EIF are considered in this thesis and a special focuses is put on ReEIF "Application level"- and "Information level"-Interoperability in the concept to be developed.

2.1 Standard Development Organizations

Research work focusing on the application or development of standards in the domains of health, but also in the domain's of transport & environment, have a clearly specified contexts fulfilling clear aims supported through these standards. Standards Development Organizations (SDOs) play a very important role regarding interoperability in medical informatics and interconnection of medical IT systems (Davis and LaCour, 2014). The Public Health Data Standards Consortium (PHDSC) provides a list of 17 SDOs, reaching from clinical trails, digital images, medical products, emergency data public health over to financial/business transactions and billings (Public Health Data Standards Consortium, 2018). All of the listed SDOs are concerned with developing, maintaining or at least using terminologies for medical devices and IT services used in several medical areas like

pharmacy, radiology, laboratory or clinical trials. Moreover they specify technical communication languages i.e. on an application layer level of the OSI model and interfaces as well as provide tools for knowledge management like terminology databases. They define processes and workflows to align clinical pathways and therefore improve its efficiency, make them comparable and decrease costs. Furthermore, they provide guidance regarding security measures like encryption and authentication, but additionally provide privacy recommendations like policy implementation and integration. The following examples provide a short overview about the manifold areas and kinds of activities of SDOs in the healthcare domain and therefore briefly describe SDOs, which are important for this thesis and its related studies.

2.1.1 Institute of Electrical and Electronics Engineers (IEEE) and IEEE 11073 Group of Standards

IEEE is the worlds largest technical professional organization (Institute of Electrical and Electronics Engineers, 2018) and has activities in nearly all areas of technology. In context of interoperability of medical devices the ISO/IEEE 11073 health informatics - medical/health device communication standards group, develops the 11073 series of standards. Within this family of standards, the focus is the standardized communication between Personal Health Devices (PHD). The ISO/IEEE 11073 (X73) family includes several standards. The most prominent one is the ISO/IEEE 11073-10101 (IEEE Standards Association, 2004), which describes the Nomenclature used to exchange data between PHDs and from PHDs to IT infrastructure. This specification strongly fosters semantic interoperability.

However, also the ISO/IEEE 11073-20601 (Optimized Exchange Protocol) (IEEE Standards Association, 2014) and its device specializations standards are important. The Optimized Exchange Protocol (X73-OEP) defines a basic structure, encodings and processes for PHD communication of data independent of the type of device. Blood pressure monitor, pulse oximeter or weight scale have device specialization

standards, as special requirements through their use and nature of data are common. A complete list of device specializations and a general overview can be found in ISO/IEEE 11073-00103 (IEEE Standards Association, 2012). In this thesis the standards were used within the feasibility studies from a syntactic and semantic perspective.

2.1.2 Integrating the Healthcare Enterprises (IHE)

Healthcare professionals, universities and industry companies unite themselves under the initiative "Integrating the Healthcare Enterprises" short IHE. The main mission of IHE is to improve the way IT systems in healthcare exchange information through promoting the use and application of medical IT standards (IHE International, 2018a). Therefore, IHE is mainly concerned with the topic interoperability from the conceptual, dynamic and pragmatic interoperability levels (in accordance to (Tolk et al., 2013)). IHE additionally provides and collects tools and services supporting development and testing the specifications (IHE International, 2018f,b).

IHE uses a well defined process (IHE International, 2018e) to reach the aim of improving or introducing interoperability in different fields of healthcare based on scenarios. This approach includes all necessary stakeholders, from healthcare professionals like physicians and nurses over technicians and developers, patients, management and external specialists. In the first phase of the process the user requirements are collected and documented in form of scenarios and fitting standards (i.e. HL7, DICOM, IEEE, OASIS and others) are identified. Based on this, technical specifications are developed for this specific scenarios i.e. description of communication processes, data exchanged and terminologies needed. This specifications are called integration profiles and are collected in frameworks regarding their medical domain e.g. radiology, pharmacy or laboratory, but also IT infrastructure for systems focusing on medical documentation like EHR or PHR systems and more.

In phase two manufacturers are implementing these specifications, which are

subsequently tested extensively before and during a testing event called Connect-a-thon. These tests focus severely on interoperability (no-peer and peer-to-peer tests) and are done with the IHE gazelle testing framework (IHE International, 2018b). The tested products receive an integration statement and can then, in phase three, be applied and integrated in existing health IT systems. However, the process is of iterative nature, which leads to a feedback loop and perceptions like problems or missing coverage of requirements are feed-back to the documentation of the requirements. This approach improves the quality of the specifications, supports knowledge transfer between all stakeholders, decreased burdens of implementations and decreases costs on a long range perspective. IHE connects more than 135 organizations in its activities (IHE International, 2018d).

The basics of IHE-terminology are necessary to understand the conceptual approaches in this thesis and therefore the most important once needed are briefly explained:

- IHE Actors: IHE Actors, in short actors, are pieces of software e.g. modules which are using IHE interface specifications for communication. Actors are defined on a scenario level and communicate via IHE transactions. Some actors can be used in several IHE integration profiles, although they are defined once in an concrete integration profile. Furthermore, actors can be grouped together, which therefore allows the combination of integration profiles by a Lego like approach. An example could be to integrate a profile for requesting patient demographic data together with a profile for requesting proper patient identifiers, when communication is done between different IT systems like radiology- and laboratory information systems.
- IHE Transactions: IHE transactions are used between actors to communicate and exchange information. A transaction has a specific purpose and is based on interoperability standards used to define structure and semantics. Nevertheless, transactions can have different manifestations e.g. requesting documents need the ability to formulate different query parameters. Transactions, similar to actors, are sometimes used in more than one profile

e.g. in direct and uni-direct communication of patient information, as this is defined in different profiles.

- IHE Integration Profiles: An integration profile is the technical specification based on requirements collected from the different stakeholders and well defined scenarios (see 4.3.1). Therefore, an integration profile is a collection of actors and transactions to describe the processes which are necessary to promote interoperability in a spec scenarios. This includes narrative description of the purpose of the profile, its application forms, its implications and boundaries. Furthermore, an integration profile includes overviews and links to used standards in the processes and transactions as well as flow charts, sequence diagrams and transaction diagrams for describing on how the actors and transactions work together.

- IHE Frameworks: IHE frameworks are a collection of documents which include the specifications for different scenarios in a medical domain i.e. exchange of documents between organizations or integration of medical devices in enterprise IT systems. These documents commonly underly the schema of "Volumes", which means that Vol.1 includes an overview, describing a specific scenario from a management perspective. That includes actors and its connections via transactions and the different manifestations i.e. pediatric option when demographic information of a woman is transmitted from general hospital to infant clinic. Vol.2 describes the transactions in a detailed level and therefore target at engineers. Vol.3 describes integration profiles interconnections through transactions and content transmitted in the transactions. Vol.4 focuses on semantic interoperability and aim at describing codes and terminologies. However, not all volumes are necessarily defined and used in all existing IHE domain specific frameworks.

Figure 2.2 shows a graphical example of using Actors and Transactions in a profile, which is stored in the IT-Infrastructure Technical Framework of IHE. Therefore the IHE Profile Consistent Time (CT) is shown, which purpose is to show how any kind of Software may synchronize its system time according to a standardized process with

a time server. It has two Actors called Time Client, the software module requesting

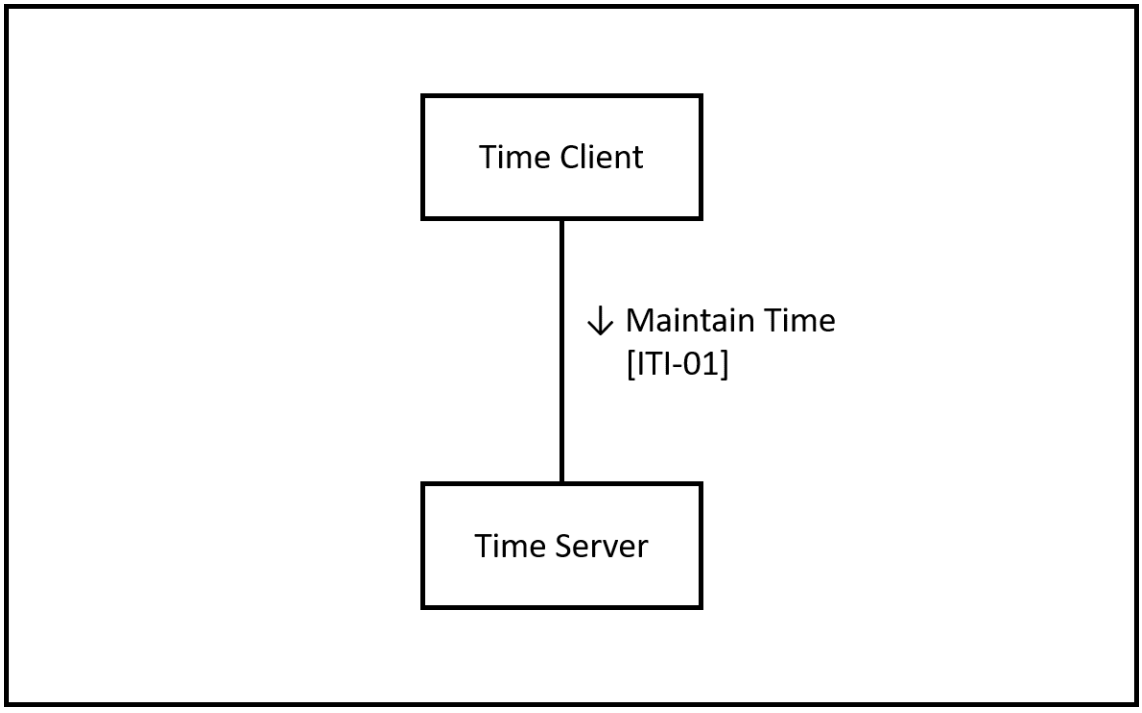


Figure 2.2 – Shows an example IHE Profile, called Consistent Time(CT), which purpose is to show how any kind of Software may synchronize its system time according to a standardized process with a time server.

time synchronization, and the time server, the component centrally providing time for synchronization with several system components. These two Actors use the so called "Maintain Time" Transaction to do the time synchronization via using the Network Time Protocol (NTP).

2.1.3 Personal Connected Health Alliance (PCHA)

An important part of interoperability in healthcare is the integration of medical devices. The PCHA focuses on the seamless integration of a complete interoperable measurement chain from the medical device to the professional medical information system (Personal Connected Health Alliance (PCHA), 2018b). The PCHA focuses on main activity fields like "elderly living independently", "chronic diseases" and

"health & fitness". However, the focus lies on the personal environment of patients or health-conscious persons. The PCHA is concerned with the conceptual, dynamic and pragmatic interoperability levels (in accordance to (Tolk et al., 2013)) and has a similar approach as the IHE (see 2.1.2). Based on their main fields of activity, they provide the Continua Guidelines for implementation guidance (Personal Connected Health Alliance (PCHA), 2016), which are basically the specifications for organizations interested to implement parts of or a complete IT system. Therefore they provide a reference architecture, which is the basis for building an interoperable system and describing the interfaces & communication processes between the PHDs, Apps on mobile platforms like smartphones, tablets or black-box systems and medical IT infrastructure components. The PCHA furthermore provides a test-suite and implementation support for members (Personal Connected Health Alliance (PCHA), 2018c). Additionally a testing event called Plugfest is conducted, where the implementing organizations execute interoperability tests in form of no-peer and peer-to-peer tests.

A huge difference between the approach of the IHE and PCHA is, that the PCHA provides certification of tested devices. This implies that the devices are tested in special test-labs by permitted test organizations. In case of success the devices are allowed to use the PCHA label and are published in the showcase database, which can be found in Personal Connected Health Alliance (PCHA) (2018a). However, IHE and PCHA are working closely together as the PCHA uses IHE specifications, but also standards from IEEE, HL7 and others in their guidelines to integrate health related data in medical information systems.

2.1.4 Health Level Seven (HL7)

Health Level Seven International is one of the core SDOs in healthcare IT, providing standards and frameworks for exchange, integration, sharing and retrieval of electronic health information in several aspects. HL7 is concerned with the topic interoperability from the syntactical, semantical, pragmatic, dynamic and

conceptual interoperability levels in accordance to (Tolk et al., 2013). HL7 has several affiliates in 50 countries with more than 1600 members (Health Level Seven International, 2014). According to (Health Level Seven International, 2018c), the provided standards in the reference categories primary standards, foundational standards, clinical and administrative domains, EHR profiles, implementation guides, rules and references and education & awareness.

The most general and prominent HL7 standards is the Clinical Document Architecture (CDA) focusing on the definition of clinical documents based on XML i.e.laboratory report etc.. Similarly important is HL7 messaging in version 2 (HL7 V2) and version 3 (HL7 V3), which both focus on the exchange of clinical and administrative data in intramural and extramural contexts in form of syntactically and semantically defined messages. The most recent standard is called HL7 Fast Healthcare Interoperability Resources (HL7 FHIR), which is a modern approach to conform to nowadays dynamic and lightweight requirements of the IT sector (Health Level Seven International, 2018c). These core standards are integrated and used in PCHA guidelines as well as in IHEs frameworks and integration profiles to a huge extend. An example for this application is the integration of PHD data do telehealth-service centers. This is done by using HL7 V2/V3 messages based on the IHE profile Device Enterprise Communication (DEC) from the patient care domain i.e. technical framework. This latter example shows that standards are combined. All of these standards were described in more detail in 2.2, as they are fundamental for this thesis.

2.2 Documents, Messages and Resources: Interoperability in eHealth

For the interoperable exchange of data between information systems, syntactical and semantical requirements are of high importance. In medical information technology the standards CDA, HL7 v2 & HL7 v3 messaging and HL7 Fast Healthcare Interoperability Resources (FHIR) are of severe importance when it comes to this

requirements. The general difference between these standards lies in the purpose of using the exchanged data. CDA focuses mainly on definition of medical documents, which includes the aspects of persistence, wholeness, stewardship, context relation and human readability of data. On the other hand, HL7 v2/v3 messaging standards are message based i.e. data is transported regarding specifications, but may or may not be stored in an unspecified way and the message may or may not be terminated. HL7 FHIR takes another approach and is an upcoming alternative to the document and message related approaches in accordance to Brull (2013); Bresnick (2018). In this case, data is structured in resources (i.e. like modules) which hold specific information, but can have relationships i.e. one resource for patient information and one for clinical trial data, and both are (or may be) linked together. HL7 FHIR is a modern approach, where the others are well known and widely applied specifications with their drawbacks e.g. XML based clinical documents can get extremely large through the markups. The modular approach of HL7 FHIR standard supports the integration with CDA documents as well. Nevertheless, the difference of documents, messages and resources should be kept in mind. The following sections will briefly describe the HL7 standards, as some aspects are fundamental to this work.

2.2.1 HL7 Clinical Document Architecture (CDA)

The Clinical Document Architecture (CDA) is a standard specified and developed by HL7 International and is part of the HL7 version 3 standards. CDA is an XML based document-markup standard, which allows to develop highly structured clinical documents for laboratory reports, radiology reports and any other kind of health related document (Dolin et al., 2006). The first version of CDA release 1.0 was proposed in 2000. The current release 2.0 was published in 2005 as an official ANSI standard, later accredited as an ISO standard in 2008 (Rodrigues et al., 2016).

Apart of narrative text, which may or may not be structured, pictures, sound files or other media can additionally be integrated in a CDA document. CDA is a well adopted standard, which fulfills the necessary characteristics for documents in the

healthcare sector. According to (Müller et al., 2005; Ferranti et al., 2006), this includes that a CDA document can exist in an unaltered state (Persistence) and can be maintained by an institution or person (Stewardship). Furthermore a CDA document can provide a complete context to understand its content condensed in one assembly of connected information (Context, Wholeness) and enables authentication through providing the ability to record or attest the signature of a healthcare professional responsible for the content (Potential of authentication). Finally, CDA provides the ability to present its information in human readable form (Human Readability) as the CDA-XML containing the data can be displayed by application of Extensible Stylesheet Language Transformation (XSLT).

As CDA is part of the HL7 version 3 standards family, its specification is based on the Reference Information Model (RIM). Basically a CDA document consists of a header-element and a body-element, were the latter is again structured in sub-elements like sections and entries. Figure 2.3 shows the basic structure of an CDA document in accordance to the RIM.

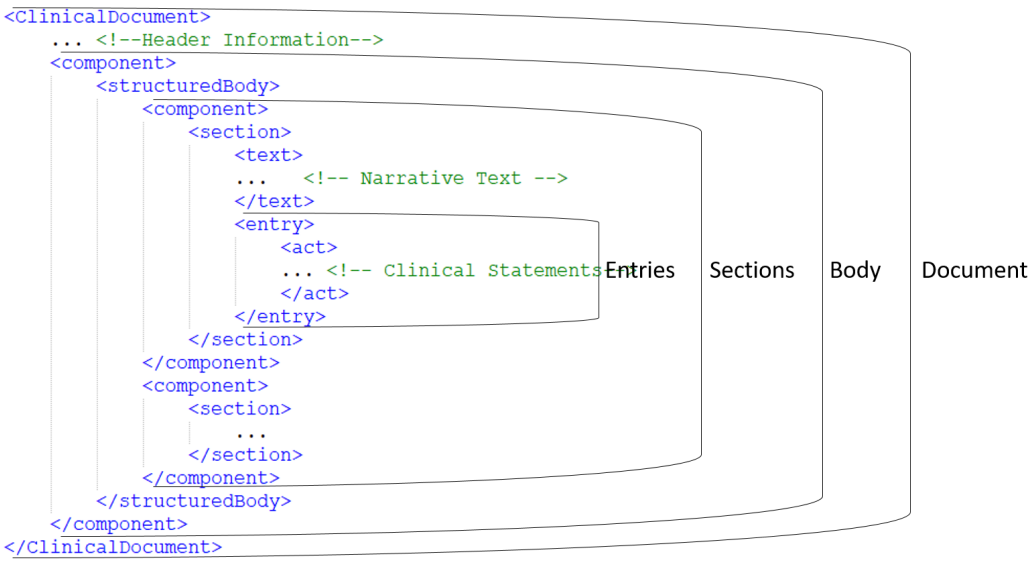


Figure 2.3 – Shows the basic structure of a CDA document in accordance to the RIM.

The CDA-header includes information related to the patient i.e. demographic

information, information related to the involved organisations and persons like physician, nurses etc..Furthermore it incorporates meta-information describing the document itself like identifier of the document, category of the document, time related information and relations to other documents. This information is in a highly structured format connected with fixed semantic definitions, as this are crucial requirements to search, exchange and interpret these documents in a medical information system like an EHR system or HIS (Boone, 2011).

The CDA-body includes the relevant clinical information, whereby at least narrative text describing the desired clinical context is included. This information can furthermore be provided in structured form to improve interoperability and further automated processing of its content. CDA proposes three levels of interoperability, depending on how the data is structured and described (HL7 Austria, 2013):

- CDA-Level 1: Level 1 focuses on human readability and therefore data is integrated in narrative unstructured form or as an embedded document like pdf or text-file. The data is stored in the CDA Body.
- CDA-Level 2: Level 2 provides the possibility to structure this data in the body by using "sections" with defined meaning e.g. diagnose or allergies. These sections are designated with codes to provide identification and improved processing through IT systems.
- CDA-Level 3: Level 3 provides the automated processing of information by IT systems, as detailed information inside "enties" is designated with codes e.g. ICD-10 and LOINC-codes for observations and UCUM codes for physical units. This is done in addition to the narrative text, which is still in an specified XML-element. However, this requires templates to be developed for defining sets of rules.

The specification of the CDA standard provides XML-elements, which can be used to structure the body and are commonly defined as "body structures". Therefore, as shown in 2.3, the body is separated in thematic blocks called sections, which

themselves can be nested by sections. Structures like lists or tables are applicable, which are labeled as ”paragraph”, ”content”, ”caption”, ”table” or ”list”. Sections always include a narrative text for human interaction. The computer readable parts are called CDA-entries (see 2.3). These are used to encode information from the narrative text by usage of classes and attributes from the RIM. Example elements for entries are ”observation” i.e. a measurement like blood pressure measurement, ”procedures” i.e. a surgery, ”observationMedia” like figures, graphs etc. or ”supply” for medication. The standard offers to link these elements recursive or linear. However, the RIM provides a huge set of elements for individual setup and definition of needs, which would go behind the scope of providing an overview in this section of this work. The complete RIM can be found in Health Level Seven International (2018e).

2.2.2 HL7 Messaging (v2/v3)

HL7 version 2 messaging is a standard continuously developed since 1989, which is reflected through several versions starting from 2.1 to 2.8.2. When speaking about all versions, the term HL7 v2.x is used (Health Level Seven International, 2018c). These versions are backwards compatible and the main purpose of HL7 messaging is the communication inside healthcare institutions to support of clinical, administrative, logistic and financial workflows. HL7 v2.x messages are text based messages with a specific syntax for structuring the information to be transmitted (Rodrigues, 2010). Figure 2.4 shows a truncated HL7 Version 2.6 message for transmitting weight data in context of the Continua Design guidelines from the PCHA:

```
MSH|^~&|UASTW^ACDE48234567ABCD^EUI-64||||20120611080245+0200|RU^R01^ORU_R01|1234|P|2.6
PID|||1234^^^UASTW-Hospital^PI||Urbauer^Philipp||19840126|M
OBR|1|UASTW-Hospital|UASTW-HIO|182777000^monitoring of patient^SNOMED-CT|||20180231145510+0000
...
OBX|12|NM|188736^MDC MASS BODY ACTUAL^MDC|1.0.0.8|83|263875^MDC DIM KILO G^MDC|||||R|||20180301
```

Figure 2.4 – This shows a truncated example HL7 version 2.6 for transmission of weight data in context of the Continua Design guidelines of the PCHA (Personal Connected Health Alliance (PCHA), 2016).

HL7 v2.x messages are separated in segments, which is demonstrated by MSH, PID or OBX. Each of this segment contains related information in form of fields and is terminated with a carriage return. The first segment of a message is the message header starting with its abbreviation "MSH". PID depicts the patient identifier segment, OBX stands for an observation segment and is part of an observation request group (OBR). Each of these segments include, as already stated, specific fields which include relevant information and is specified by the v2 standard. Each field is separated by a pipe-character. The header abbreviation is always at position 0 e.g. PID-0 incorporates the abbreviation "PID" and then counting from left to right followed by the next PID fields 1..n. These fields may be optional, conditional or required depending on each segment's specification. Information in each field is separated by circumflex-character. An example like "Urbauer^Philipp^^^^^", shows the separation of family name and given name. In case there is no information between two circumflexes, this indicates there is no need for this information i.e. in this example no second name, suffix, prefix or academic degree. Using the ampersand-character instead of the circumflex-character allows to specify subgroups. The same approach is used to add semantics to these messages. This is shown by the SNOMED-CT code for describing that the purpose of the OBR is the monitoring of a patient. Finally, the first field of the MSH specifies these and other special characters, which are used for separation i.e. "^~&". The message itself is transmitted via TCP/IP, packed into the Minimal Low Layer Protocol (MLLP), which can simply be described as it adds header and trailer characters.

HL7 v2 messaging is very widely applied in medical IT systems from the perspective of the IT systems communication inside healthcare institutions and it has a very pragmatic approach to support quick solutions. However, its drawback is that this approach leads to inconsistencies regarding data, due to different regional modeling approaches. Through the publication of HL7 version 3 (Health Level Seven International, 2018c), the aim was to cover all communication needs derived from the healthcare system and specify format, content, semantics and processes for messaging purposes (Huang et al., 2003, 2005). As HL7 v3 messaging is part of the v3 family of standards it is, similar to CDA, based on the RIM as this is

the basis for all standards in HL7 version 3. Therefore, it uses XML for defining and implementing messages based on the RIMs specifications. However, v3 did not reach the same impact compared to v2 and therefore has a lower application rate as its former standard. This seems justified by the fact that implementation burden is higher due to the RIMs complexity and its related transformation process from the generic model to the platform related model. The authors Bender and Sartipi (2013) describe that the development of supportive tools require approximately 18.000 man hours and mention incompatible applications in Canada through misinterpretations due to the syntactical complexity of HL7 v3. A further problem is, that there are XML schema distributed with the standard, but they are not normative.

2.2.3 HL7 Fast Healthcare Interoperability Resources (FHIR)

Under the abbreviation FHIR, HL7 works on its latest approach for interoperable exchange of healthcare information (Health Level Seven International, 2018b). This started in 2011 and HL7 FHIR targets to improve the drawbacks of HL7 v2 & v3 i.e. data inconsistency in v2 and furthermore the fact of decreased interest of the community to implementing v3 due to its complexity. Therefore HL7 FHIR aims to provide a new alternative in context of simplicity of implementing interoperability standards, without loosing data integrity and overall consistency. Hence, HL7 FHIR is strongly aligned to previously developed HL7 standards and content models i.e. the HL7 RIM. HL7 FHIR includes several advantages derived from HL7 v2/v3 & CDA and additionally extends these. An example is, comparing CDA with FHIR, that CDA supports the exchange of data with clinical relevance, but lacks integration of non-clinical information such as financial aspects. Another fact is that CDA is limited to establish documents targeting only patients (Bender and Sartipi, 2013). On the other hand both, CDA as well as HL7 FHIR, require the data to be human readable including information on how this should be done (Health Level Seven International, 2018a). The previously stated target of ”simplicity”, is not only covered through the HL7 FHIR approach and specification, but furthermore by

support of the latest technologies like RESTful architecture, XML, JSON, HTTP(s) and service-oriented architectures (SOA) to exchange both, messages and documents (Health Level Seven International, 2018f). This provides well known setup for engineers and supports fast application in several contexts like cloud based solutions, EHR, PHR, HIS and smartphone or tablet Apps. HL7 publishes HL7 FHIR as a Standard for Trial Use (STU) and currently HL7 FHIR release 3 (STU 3.0.1) is the active version, which can be found in (Health Level Seven International, 2018b).

HL7 FHIR supports the approach of modularity through definition of "resources". These resources include defined information and can be combined in different ways to fulfill requirements derived from several application areas in healthcare. In healthcare typically a system grows over time, as new requirements and application fields are added continuously. HL7 FHIR tries to contribute to this fact through its modular approach and the possibility of defining and integrating "extensions" according to project requirements. These can be integrated in HL7 FHIR resources in case the resources defined in the standard are not yet sufficient to provide a solution for a defined need (Health Level Seven International, 2018d). All resources share common characteristics as indicated in figure 2.5. This example shows a patient resource as an example.

Each resource has an identifier element and a meta-data section describing the context of the resource. The text section depicts the human readable information using HTML. Subsequent the extension-element supports to include content which is not defined by the actual resource. Finally the last section includes the standard data from the resource specification e.g. in this case the identifier of the patient, name, gender, birth date and further data. HL7 FHIR provides several resources for clinical requirements like patient, provider, medication and diagnostics recourses, but additionally for e.g. financial concerns and other requirements. HL7 FHIR provides a modular setup and the focus on modern lightweight technologies and provides broader application as stated before. However, combinations of standards might make sense in some aspects, especially if it comes to the requirement of reports of results, which by nature are document based.

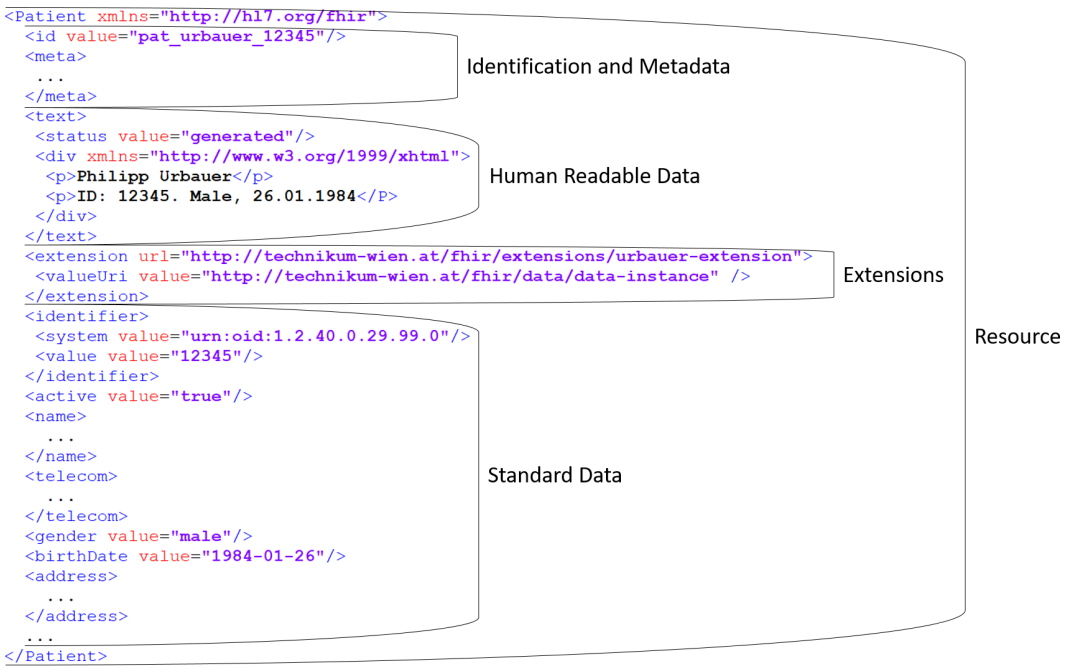


Figure 2.5 – Shows an example patient resource to describe the common characteristics of HL7 FHIR resources in general, according to the specification shown in Health Level Seven International (2018b).

2.3 State-of-the-Art Research and Related Work

Investigating scientific data bases, regarding occurrences of work concerned with HL7 FHIR, shows that until now only a few results could be identified. ScienceDirect (ScienceDirect, 2018) and IEEE Xplore (IEEE Xplore Digital Library, 2018) were used with the keyword "FHIR" individually and in conjunction with "HL7". The timespan was specified from 2011 to 2018, as HL7 FHIR was started in 2011. ScienceDirect results in valid 46 results (research- and review-articles) and IEEE Xplore shows 27 valid results (Conference proceedings). Compared to searching the term "IHE" (Integrating the Healthcare Enterprises), which resulted in 435 (research- and review-articles at ScienceDirect) and 211 (Conference proceedings and journals at IEEE Xplore), this indicates the novelty of HL7 FHIR and associated work.

Actual work in coherence with IHE using HL7 and IEEE standards, are especially concerned with reports on experiences in practical application, but also security research especially focusing on EHR solutions as shown by (Yang et al., 2015; Tseng et al., 2016). The group developed a privacy framework, which can be used in the IHE EHR based communication between hospitals using a privacy matrix representing patients privacy policies. These can then be selected appropriately according to the requirement of the applied scenarios like searching for (registry-stored-query XDS.b transaction) or retrieving documents (retrieve-document-setb XDS.b transaction). In their work, Sloane and Gehlot (2016) described the application of IEEE 11073 standards family and IHE Patient Care Device (PCD) domain profiles, which are also used in the PCHA architecture, to provide a cost efficient and secure way of chronic disease management. Another experiences study form Pahontu et al. (2015) focuses on the integration of medical devices to IHE based IT systems using HL7 messaging and DICOM standards. Clarke et al. (2017) applied an end-to-end approach, following the PCHA guidelines, to transmit data from medical devices to professional IT-systems using interoperability standards like IEEE 11073 family, HL7 messaging, FHIR and IHE profiles (Device Enterprise Communication (DEC)). The area of medical device integration is currently of huge interest through the change of population and the coherence to chronic diseases. Technical feasibility and application was investigated by the author of this thesis and his colleagues earlier. In (Urbauer et al., 2012) experiences in the implementation of IEEE 11073 standards, PCHA guidelines and IHE components are stated as well as problems and solutions are provided. Urbauer et al. (2015) investigated the applicability of IHE components to the area of telemonitoring and described its implications, observed gaps and solutions to overcome these problems. This work is strongly connected to this thesis and is described in more detail in later chapters.

Personal Health Records (PHR) are currently of high interest. In the work described by Hong et al. (2017a), the group proposes an PHR ecosystem using IoT cloud based communication and HL7 FHIR. A FHIR server is included in the HIS providing data storage for FHIR resources. Third party applications or consumers are able to request data only via a special portal granting access to patients data and

therefore acting as an interceptor for privacy reasons. Furthermore, a mobile App was developed to control and administer the accounts and their interconnection. In another work proposed by Aliakbarpoor et al. (2017), a framework for a PHR based on HL7 FHIR was described. In this approach they build a PHR system based on a mobile App, which is communicating with the clinical portal using FHIR. They developed a prototype to prove their concept, which focuses on the "regular home care plan"-process. Results showed that it was feasible to develop a prototype using FHIR specification. However, a lot of research work is done in the area of security, which indicates that these aspects are coming too short in the specification yet. Janki et al. (2017) provide an approach for extending an open source implementation of FHIR, called HAPI FHIR (University Health Network, 2018), with Role-Based-Access-Control (RBAC) and Attribute-Based-Access-Control (ABAC) models apart of its integrated access control model. Another example is the work of Sanchez et al. (2017), where they similarly provide a solution to integrate RBAC security policies with FHIR by using an App, called CT^2, and an OpenEMR instance connected with an HAPI FHIR server. They successfully tested their approach by using different accounts for nursing staff and coaches to prove proper access control. Apart of this main application areas of privacy and PHR, a project was concerned with providing an interactive way to use and understand the HL7 FHIR specification itself (Hong et al., 2017b) and others were mostly concerned with App development and integration of HL7 FHIR as it adds especially support for mobile applications through the RESTful approach (Lamprinakos et al., 2015; Bloomfield et al., 2017).

This state-of-the-art investigation on the application and usage of IHE and HL7 FHIR shows, that actual research focuses strongly on the applicability of these processes and standards on mobile platforms, PHRs and EHRs as well as security and privacy topics related to authentication and authorization in these distributed environments. However, related work focusing on the approach of this work in applying health standards from IHE and HL7 to other domains could not be identified.

2.4 Information Exchange in the Domains of Transport & Environment

In the domain's of transport and environment, ICT solutions are strongly covered by the term of Intelligent Transport Systems (ITS). Interoperability is in terms of ITS equally important as in the health domain as it is shown by (United States Department of Transportation, 2018). The ITS strategic plan 2015-2019 focuses strongly on fostering connected vehicles and collecting, processing, sharing and applying of enterprise data(International Organization for Standardization, 2018). Therefore, interoperability is an explicit aim in this strategy. Equally, also in the EU intensive promotions of ITS can be identified. This is showed by the standardization organizations like Institute of Electrical and Electronics Engineers (IEEE), International Organization for Standardization (ISO), Comite Europeen de Normalisation (CEN) and the European Telecommunications Standards Institute (ETSI) (European Telecommunications Standards Institute, 2018; CEN/TC 278, 2018). These are focusing on autonomous driving via infrastructural integration, urban mobility over to integrated transports in smart cities and much more. However, on a global level the ISO/TC 204 leads the standardization work regarding ITS (International Organization for Standardization, 2018). CEN/TC 278 and ISO/TC 204 have work groups focusing on public transport (WG3(CEN) & WG8(ISO)) as well as human-machine interfacing (WG10 (CEN)).

Regarding information exchange in these domains, two standards have to be stated. The first ISO standard is called TPEG (Transport Protocol Experts Group). The standard was named after the group initially started with its development. Today TISA (Traveller Information Services Association) is responsible for the development of TPEG. First it was published under the term TPEG1 (ISO/TC 204 Intelligent Transport Systems, 2013) and in an improved second version it was named TPEG2 (ISO/TC 204 Intelligent Transport Systems, 2016). TPEG1 focused to support services for information exchange in the context of private transport systems (e.g. navigation systems) and public transport systems. The drawback was, that TPEG1

only supported the data transmission in a binary format via digital radio, mobile communication systems or WiFi. This was changed with the introduction of TPEG2, which supports binary- and XML-format through a common UML model definition. This standard is widely used in today's ITS ICT solutions. Examples are the exchange of traffic information, accident-reports, weather information, fuel prices, public transport data or parking information.

DATEX II is a European technical specification for modelling and exchanging ITS related data in a harmonized format (Doelger and Geissler, 2012). The current version is DATEX II (CEN TS 16157), which was published on the 5 October 2012. DATEX II has a strong focus on travel and traffic data. The technical basis is a model, which is mapped to an XML schema. Hence, this can be used to test conformance of DATEX II XML based data structures. The model additionally allows to apply so called "level B" extensions, which can be understood as a set of rules to enhance the specification and support a certain level of dynamic extensibility. The specification is developed along the project EasyWay, which is supported by the European Commission. The aim is to foster application of the format until 2020. Hence, actual work is done focusing on feasibility studies, like shown by the example of Ruiz-Alarcon-Quintero (2016). The group investigated the application of a proposed data model, along with the DATAX II specification to exchange traffic datasets in the region of Seville and Malaga. As a conclusion, the group stated that the specification efficiently fosters interoperability of traffic datasets and discussed the advantages derived from application, like cost savings and improved re-use of data for analysis.

2.5 Security & Privacy Perspectives

Data storage and transmission between services and/or devices requires to satisfy appropriate security and privacy requirements. The nature of data plays an important role regarding efforts which therefore need to be undertaken. Data related to healthcare is categorized as highly sensitive data (Legal Information System of

the Republic of Austria, 2018a; European Commission, 2018b). Hence, security and privacy requirements for software, devices and IT systems underly several layers of laws, regulations and policies from different institutions.

Taking a top-down approach, the first layer yields laws, regulations and policies from international organizations like World Health Organization (WHO) or Organization for Economic and Co-operation and Development (OECD) and integrates human rights. The second layer concludes the confederation of states and countries like the EU or US. From this point of view law, regulations and policies are covered by the Health Insurance Portability and Accountability Act (HIPAA) in the US (U.S. Department of Health & Human Services, 2018) or the Data Protection Directive 95/46/EC (European Commission, 2018b) in the EU. The latter provides general rules for data processing, but provides a special focus on health related data as this data is highly sensitive. The new regulation on the protection of personal data (GDPR) applies on the 25th of May in 2018 and it is expected to have severe impact to IT industry especially to the healthcare sector (European Patients Forum). This is the case as it fosters the "right to erase" personal data and aims to simplify the language of consents, which therefore amends policies and related processes. In addition, it tightens the requirements for the use of technical IT security measures. Apart of these, the country regulations furthermore add details and restrictions to security requirements. Taking Austria as an example, the basis for data security is defined by the data security law called DSG-2000 (Legal Information System of the Republic of Austria, 2018a). This law adds a clear definition of data categories and therefore the emerging requirements e.g. person related data vs. sensitive data (e.g. health related data). The next layer is the domain related layer, like health or energy domain. Pursuing Austria as an example, the health-telematic law (Legal Information System of the Republic of Austria, 2018b) and the EHR law, called "Elektronische Gesundheitsakte-Gesetz - ELGA-G" (Legal Information System of the Republic of Austria, 2012), describe requirements and restrictions from a legal perspective regarding telemonitoring (e.g. PHR) and EHR applications. In addition, medical societies or engineering societies may influence this layer. Finally, the last layer is the enterprise layer, which covers all "good practice" regulations and rules

as well as technical measures like backup-, recovery-systems and related workflows and processes.

This overview shows the large impact of security and privacy requirements derived from laws, regulations and policies. Therefore, it would go beyond the scope of this thesis to provide an in-depth security analysis and security procedure models as the focus of this work lies on technical interoperability. However, to provide basic considerations and recommendations regarding technical measures, which can be applied together with the developed concept and its applications to fulfill basic IT security and privacy requirements are a necessary secondary objective of this thesis. These technical measures shall fulfill the main aspects described by models like CIA-Triad, Extended CIA-Triad and the Security Star Model (Ragged, 2010; Andress, 2014). A combinatorial approach of this models is shown in figure 2.6:

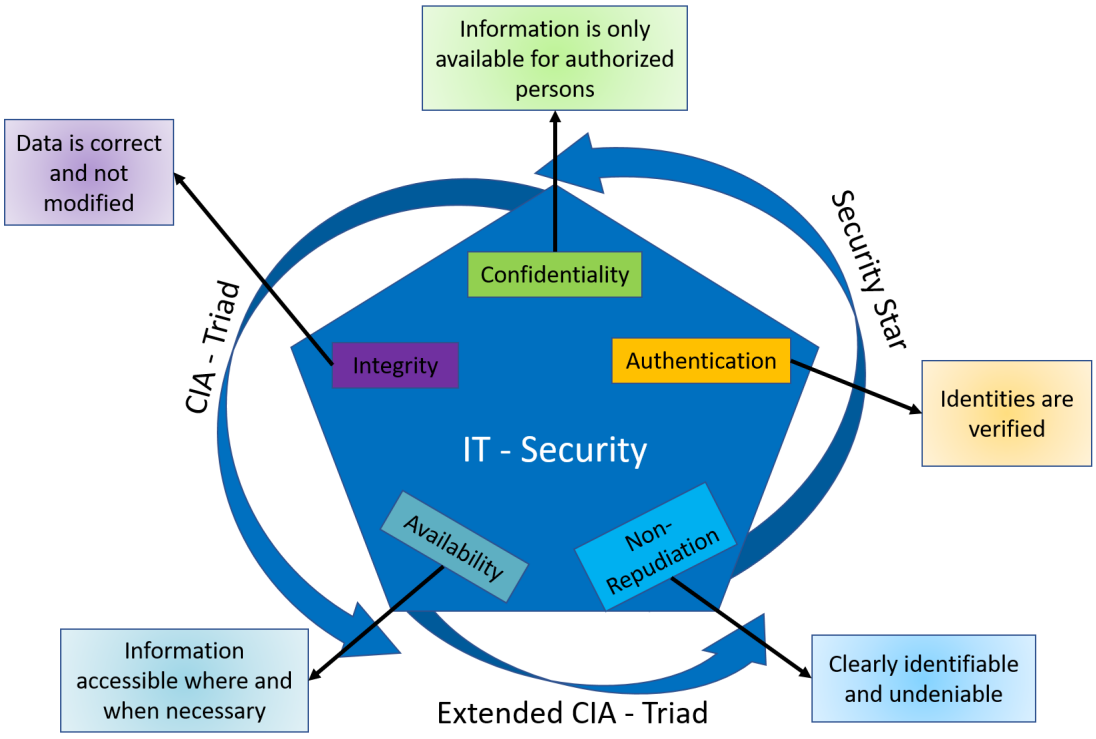


Figure 2.6 – Shows the security principles, described through the CIA-Triad, extended CIA-Triad and Security Star, which shall be considered with technical/non-technical measures.

The interior of the pentagon (see figure 2.6) shows the five important security principles, which need to be guaranteed to fulfill IT security requirements. Hence, "confidentiality" includes measures like e.g. applying Security Assertion Markup Language (SAML) and/or WS-Trust to restrict access to authorized persons or services. "Integrity", for example by usage of digital signatures, shall assure that data is accurate and not manipulated by e.g. man in the middle attacks. Finally, to complete the first model called CIA-Triad, "availability" shall guarantee that data is accessible when and wherever needed e.g. through distributed services and backbone systems. The second model is called extended CIA-Triad, which adds the principle of "non-repudiation" to the first model. This is concerned with the fact that it should be clear who or what is responsible for the data and that this is undeniable i.e. in medical terms a physician is responsible for a medical report and content without any arguing as soon as it is signed by the healthcare professional. The last model, called Security Star, completes the cycle around the pentagon and therefore includes the principle of "authentication". This includes the clear verification of identities, independent if it is a human, machine or service and a technical example could be the usage of certificates. However, this combinatorial approach and its related principles are used for security considerations connected to the developed concept in this thesis.

2.6 Necessity for Consciousness-Raising

The necessity of raising awareness for designing, developing and applying interoperability standards is an important fact and a prerequisite to finally receive the benefits of interoperability. Although the awareness of using interoperability standards e.g. in an interoperability-by-design fashion, is continuously improving, there are still areas in the medical IT domain, were interoperability is recognized and work is done, but application by manufacturers is still missing. This is shown in a feasibility study in section 4 of this thesis, which focused on the integration of Wearable Fitness Tracker devices in medical IT systems. Hence, education is an

important tool to raise consciousness, but more efforts must be made as there are still gaps in education of the medical IT domain workforce. Herzog et al. (2014) showed that on a university level there is already an improved sensibility for eHealth in general and interoperability standards specifically as well as improved integration in teaching plans. However, workforce education in a post-gradual context is mainly in form of certification programs. These were investigated by Urbauer et al. (2014) and figure 2.7 shows that the occurrence of thematic sub-areas in certification programs according to the professions within the EU.

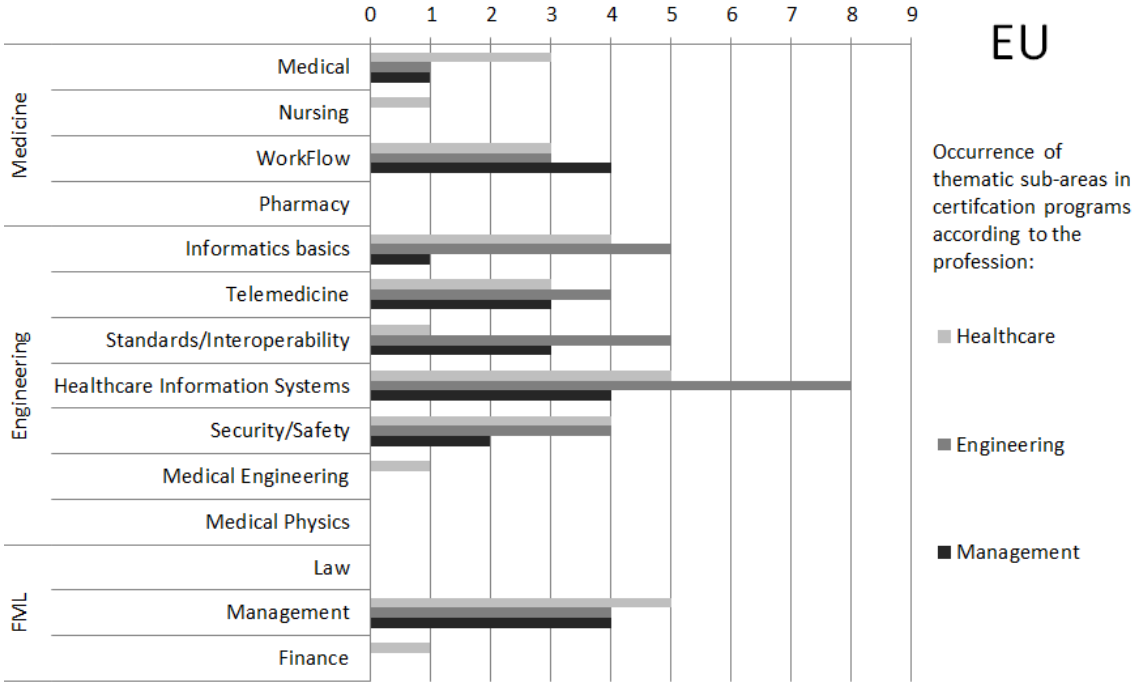


Figure 2.7 – Occurrence of thematic sub-areas in certification programs according to the professions within the EU. Published in Urbauer et al. (2014)

Based on this figure it can be concluded, that the topic of interoperability is strongly underrepresented compared to the needs. However, a positive influence from the view of standardization can be identified as the topic is integrated in programs for the engineers and management profession. Topics like IHE, HL7, LOINC, SNOMED CT, etc. are covered in some programs, which were analyzed during this study. However, official certification programs which were found and

include interoperability topics, didn't seem to be coordinated and harmonized in an international fashion. On that basis, IHE International formed the IHE education work group, focusing on the development of certification program called IHE Certified Professional (ICP) starting with ICP Foundation Level. Additionally to this work-group, the reference category "education and awareness" of HL7 (Health Level Seven International, 2018c) and its related work group, enforce the importance of raising awareness to use interoperability standards and therefore to foster education. Hence, educating all groups of stakeholders in the medical IT domain i.e. managers, medical professionals and engineers would improve the awareness to the topic an therefore hopefully increase the rate of implementation of interoperability standards in products. That might be an important measure to finally receive all advantages through the application of interoperability.

3 Materials and Methods

The work described in this thesis was structured into five stages as shown in figure 3.1. First in stage 1, the user requirements ware collected through an investigative analysis of existing data platforms and its used formats for import and export of data was conducted by applying Internet and literature based research. Based on this results, in stage 2 applicable medical IT-standards were then collected and categorized according to well defined technical-requirements. Subsequently appropriate standards and the basis technologies were selected. In stage 3, a big data business domain overview was worked out, which included a domain description and related conditions. Furthermore, a concept including models, processes and data structures forming the framework components, were developed. S¡ubsequently in stage 4, the application phase started by design and development of prototypic implementations based on the profile derived from the previous stage. This included three feasibility studies in the domains of health, environment and transport and its technical validation through conformance tests. Finally, in stage 5, an experts review was performed to evaluate the overall approach of this thesis. Parts of the used approaches and methods were already published in advance of this thesis (see section 1.3).

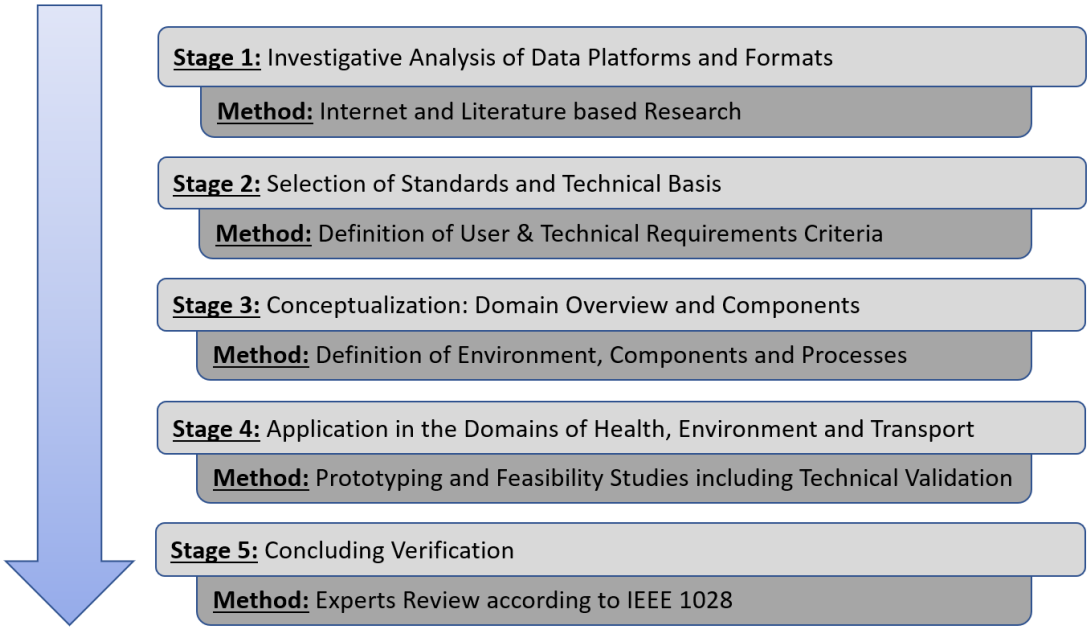


Figure 3.1 – Overview about the methodical procedure, consisting of five stages, applied in this thesis.

3.1 Investigative Analysis of Data Platforms and Formats

In the first step, data platforms were identified through performing an Internet based literature research in accordance to (Shields and Rangarajan, 2013), by using Google as a search engine. The search terms used in this approach were "open data", "big data", "platforms" and "Europe" in singular or by combination of these terms. Based on this research, several open data platforms could be identified. In order to narrow the results, only governmentally supported platforms were further considered in this research. As a result, table 3.1 lists the selected platforms, which were used in the next steps to provide an overview about the domain data and applied formats.

These selected platforms were investigated further on the applied and provided data formats. This was done to get an overview of the most used data formats and

Table 3.1 – Shows the selected open data platforms used for investigation of the applied data formats.

Open Data Platforms			
Country	*URL*	*Country*	*URL*
Austria	https://www.data.gv.at/	Holland	https://data.overheid.nl/
Austria	https://www.opendataportal.at/	Holland	http://opendatanederland.org/
Belgium	http://data.gov.be/	Ireland	http://data.fingal.ie/
Belgium	http://publicdata.belgium.be/	Ireland	https://data.gov.ie/
Cyprus	http://www.data.gov.cy/	Italy	http://www.dati.gov.it/
Denmark	http://www.portal.opendata.dk/	Italy	http://www.datiopen.it/
Finland	https://www.avoindata.fi/	Portugal	http://www.dados.gov.pt/
France	http://www.data.gouv.fr/	Romania	http://data.gov.ro/
Germany	https://offenedaten.de/	Slovenia	http://data.gov.si/
Germany	https://www.govdata.de/	Spain	http://datos.gob.es/
Greece	http://geodata.gov.gr/	Sweden	http://oppnadata.se/
Greece	http://data.gov.gr/	Sweden	http://www.opengov.se/
Holland	http://opendata.cbs.nl/	UK	https://data.gov.uk/

how these look from a syntactical and semantical point of view. After this process applicable formats were selected, which were used in the feasibility studies described later.

3.2 Selection of Standards and Technical Basis

The process for selection of appropriate standards and its technical basis was done through an top-down funnel approach. Hence, the first step was to define general user requirements based on stage 1. Therefore, the following list was defined:

- Selected interoperability standards need to enable the use of non-medical data i.e. no restriction on a specific business domain through its definition
- Interoperability standards shall be applicable on different application levels i.e. device-to-device, device-server and server-server level, but also in connected

information trails like device-device-server

- Modularity and extensibility are crucial requirements for sustainable systems and therefore need to be fulfilled by standards to provide a future proof approach

- From a technical perspective, state-of-the-art light weight technologies are necessary to enable mobile application e.g. data acquisition through crowd-sourced approaches via smartphones

- Security and privacy measures shall be enabled to fulfill national and international requirements derived through laws and regulations

As a subsequent step these user requirements were then classified into the following categories according to the EIF (European Commission, 2018d) to allow a more detailed resolution of applicable SDO-processes, standards and profiles to be applied in a next step:

- Technical requirements

- Security requirements

- Legal requirements

- Organizational requirements

Based on this classification, the requirements were reviewed based on the experiences gained through previous research projects from 2009 to 2014. Especially the technical and security requirements were of importance as they were subsequently used for the selection process and evaluation of interoperability specifications. These experiences were gained through standardization work-group meetings, conference discussions and project meetings with several different stakeholders as listed:

- Patients, without any professional knowledge (whether medical nor technical)

- IT-Architects, Engineers and Software Developers from small and medium-sized industry enterprises and healthcare institutions
- CIOs from medical professional departments like group practices and hospitals
- Decision makers of: national ministries of health, social insurance providers, regional administrations, chambers of commerce, medical chambers, federal IT infrastructure management, national and regional EHR implementation organizations

After categorization and review, the next step was the identification of applicable interoperability standards. A literature research in accordance to Shields and Rangarajan (2013) was performed to first collect interoperability standards used in the field of EHR, PHR and Telemonitoring. The search was applied using the databases Science-Direct (ScienceDirect, 2018), IEEE Xplore (IEEE Xplore Digital Library, 2018) as well as PubMed (PubMed) with the following keywords and their combinations, applied as parameters:

- Electronic Health Records (EHR)
- Personal Health Records (PHR)
- Medical Interoperability Standards
- Telemonitoring
- Telehealth Services
- Health Level Seven (HL7)
- Integrating the Healthcare Enterprises (IHE)
- Personal Connected Health Alliance/Continua Health Alliance

Identified papers were then selected according the quantity of application of the interoperability standards to narrow the results, after a first collection step. The

two tables, 4.1 and 4.2 provide an overview about the specified requirements. As a subsequent step, the interoperability specifications were added to this lists together with the matching requirements. As already stated, the selection of adequate standards was done based on the technical and security requirements, as shown in these two tables. During research a potential list of standards could be identified. These included standards and frameworks published by HL7, IHE, ISO, CEN and PCHA. In order to furthermore narrow the results, criteria such as completeness, widespread use, level of documentation and widely available implementation experience were applied. The databases from IHE for the Connect-a-thon results (IHE International, 2018c) and the PCHA product-showcase database (Personal Connected Health Alliance (PCHA), 2018a) were used as a source for decision criteria for practical applicability, as both exactly describe the degree to which the standards of the previously named SDOs were implemented. Therefore, these data shows not only the organizations which attended the testing events and certification processes, but furthermore shows which interfaces were developed at a specific time. Hence the focus was furthermore narrowed to specifications referenced by IHE profiles and/or within the Continua implementation guidelines from the PCHA. These detailed specifications i.e. the profiles were then selected and added to the respective technical requirements shown in table 4.2 and finally lead to the selection of the standards and connected technical basis for subsequent steps of this work.

3.3 Conceptualization: Domain Overview and Components

As a next step the conceptualization phase for the framework components started. The methodical procedure for the conceptualization, which is based on IHE's process for defining interoperability solutions(IHE International, 2018e), is shown in figure 3.2. The procedure was separated in four steps, initially starting with the "business domain overview". This included the environmental description of the boundary

conditions of the big data domain and explanations of typical example scenarioss related to this domain.

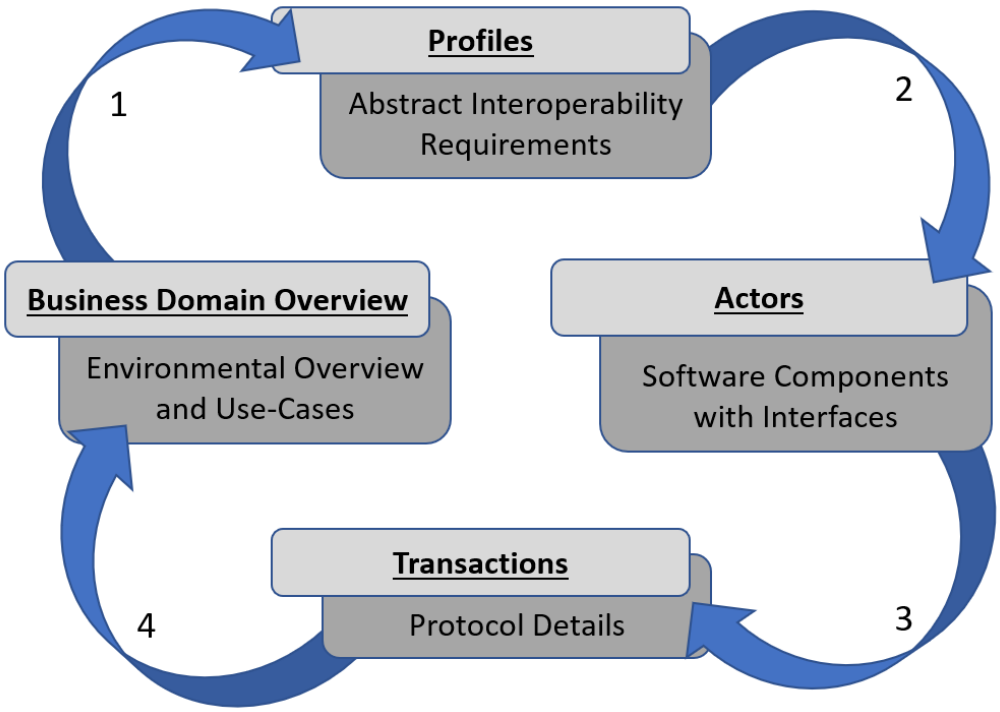


Figure 3.2 – Shows the methodical approach for defining the framework components for the big data domain.

On the basis of this collection of conditions and scenarios, the next step was to define a technical concept. This can be understood as a technical definition and description of related processes, used technologies and resources based on scenarios described in the business domain overview chapter. A clear requirement was to have generic approach, as all parts of the profile needed to be applied to the widest possible scenarios. Therefore the goal was to work out a technical application profile for the Big Data Domain (BDD). Directly related to this was the next step of defining "actors" i.e. pieces of software (e.g. modules) providing the necessary technical interfaces to work with the big data information sets. That included a description of its tasks and participation in work-flows with other actors of connected systems. The final step was to define the communication between these actors which, are called

"transactions". This included the description of the types of transactions, its relation to the actors and details about the data structures and contained data. Finally, step four was performed to add a verification step to the previously conducted steps by comparison to the conceptual definitions in the business domain overview.

3.4 Application in the Domains of Health, Environment and Transport

After verification of the abstract concept through iterative comparison against the scenarios, the application phase was initiated. In this phase three feasibility studies were performed to evaluate the conceptual definitions of the framework components and the defined scenarios. Hence, this chapter is divided into three subsections each of it describing a specific feasibility study. The first was concerned with collecting health domain data. This was done to test applicability of data integration with data strongly related to healthcare i.e. wearable fitness tracker data. The methods of this study are presented in the subsequent subsection. The second feasibility study focused on the integration of environmental domain data i.e. combining pollen data with vital parameter data to provide a better reaction for patients suffering from allergies. Finally, the last feasibility study focused on the integration of public transport data with vital parameter data from the healthcare domain to provide a layperson decision support system. This system provided medical event triggered route guidance for patients.

3.4.1 Scenario: Healthcare Data Integration

Based on the scenario definitions connected to the resulting IBD-Profile the first feasibility study was performed. Hence, the first step was to investigate the most prominent wearable activity trackers, which was done in a pre-study by accomplishing a market analysis (Marton et al., 2017) focusing on the technical properties. A summarized overview about the investigated wrist-wearable trackers

is shown in table 3.2. This includes the devices functionalities and properties. This

Table 3.2 – Shows the information about biometric indicators and interfaces for the investigated activity trackers, worn on the persons wrist. Information taken from (Marton et al., 2017)

Product	Manuf.	Biometric Indicators and Comments	Technical Interface
Basis Peak	*Intel*	heart rate measurement, activity-pattern, sleep-behaviour, temperature, sweat production	Bluetooth is used, but out of specification no detailed information available
A360	Polar	heart rate measurement, activity-pattern, sleep-behaviour	Bluetooth Low Energy (BLE)
Charge HR	*Fitbit*	heart rate measurement, activity-pattern, sleep-behaviour	Bluetooth 4.0
Gear Fit 2	*Samsung*	heart rate measurement, activity-pattern, sleep-behaviour	Bluetooth is used, but out of specification no detailed information available
Mi Band 2	*Xiaomi*	heart rate measurement, activity-pattern, sleep-behaviour	Bluetooth Low Energy (BLE)
vivosmart HR	*Garmin*	heart rate measurement, activity-pattern, sleep-behaviour	Bluetooth Low Energy (BLE)
UP3	*Jawbonde*	heart rate measurement, activity-pattern, sleep-behaviour	Bluetooth is used, but out of specification no detailed information available

study focused on the trackers communication interfaces as well as used protocols and therefore collected the technical conditions derived from the wearable activity trackers. A focus on wrist-wearable trackers emerged out of this study, compared

to other categories like breast strap based-, clothing based- or headphone based-devices. Taking a closer look, it could be identified that wrist-based activity trackers cover the most common biometric indicators, which are generally of interest to track fitness behavior and status. As a result of this research, heart-rate, sleeping-behavior and activity-patterns emphasized to be the common indicators independent of the activity trackers categories stated before. This fact was used as a basic requirement for defining the design and data criteria (biometric indicators and physical values) to be used in this prototype system to study feasibility.

A crucial problem was that, according to the wearable fitness tracker criteria selection processes described before, wearable fitness tracker support different interfaces although the BLE interface is upcoming as indicated by the table 3.2. Hence, the architecture of the mobile application shall provide a generic mapping concept to support a wide variety of wearable fitness trackers. Nevertheless, for the prototype a A360 tracker from the company Polar was used. The mapped data shall then support, through its used standards, the integration in several different medical IT systems e.g. HIS, EHR or PHR.

Based on that criteria and the described scenario, the next step was to define an IT architecture. The aim was to allow integration of non-standardized as well as standardized devices, but however finally to map the data into the HL7 FHIR format and apply the conceptualized framework components defined in chapter 4.3. For this approach the HL7 FHIR observation resource was taken as basis to integrate activity tracking data in form of a prototype to study feasibility. The following list provides the used components in this prototype:

- OnePlus 3T with Android 6.0 (Marshmallow)
- Polar A360 fitness tracker
- Open Source HAPI FHIR
- HL7 FHIR Validator

After implementation, functional tests of the prototype were performed by using the public server of HAPI FHIR open-source software framework (University Health Network, 2018). Subsequently, no-peer interoperability tests using the HL7 FHIR Validator (Health Level Seven International, 2018g) were executed to test conformance according to the HL7 FHIR STU 3.0.1 specification. Hence, testing technical validity of the approach was done with this approach.

3.4.2 Scenario: Environmental Data Integration

This feasibility study was in collaboration with the "Department of Otorhinolaryngology, Research Group Aerobiology and Pollen Information" at the Medical University of Vienna (MUV) who provided pollen exposure data. The data was provided in a proprietary form and format. However, the first step was to collect and analyze the data described by an internal document from the MUV. The data could be received from an internal server of the MUV with provided access rights for a specific time span to conduct this study. Table 3.3 provides examples of the data from the MUV pollen database, used for the prototype. The left column shows the proprietary codes for the types of pollen. Each request to the database delivers pollen exposure values for three days i.e. today, tomorrow and the day after tomorrow. The exposure values indicate the severity of the pollen load in values from 0.0 (no exposure) to 4.0 (very high). Additionally indicated by "AT", the area of interest can be defined to receive pollen particle exposure information on the one hand in ISO-code or on the other hand in geo location format.

A crucial design requirement was to support patients mobility at each point in time. Hence, to get maximum acceptance of patients the use of smartphones, tablets or smartwatches as personalized control centers was indispensable, also for sustainable use of the system. These personalized control center approach should allow the patient to measure, observe and monitor healthcare and environmental information related to this scenario. Based on that, the scenario description and the knowledge of the data and its proprietary formats, a standards based IT architecture in accordance

Table 3.3 – Shows example data, received from the MUV pollen database. This data is integrated in the prototype of this feasibility study.

AT	**Today**	**Tomorrow**	**The Day after Tomorrow**
Particle	2017-03-27	2017-03-28	2017-03-29
ALNU	1.0	1.0	2.0
CORY	1.0	1.0	1.0
BETU	0.0	0.0	0.0
POAC	0.0	0.0	0.0

to the framework components was designed and prototypically implemented as a second proof of the framework components concept from chapter 4.3. This time the prototype system included the following materials:

- Nonin Onyx Vantage 9590 Finger-Puls Oximeter (Continua Certified using IEEE 11073 standards family based)
- A&D Medical Blood Pressure Monitor UA-651ble (Continua Certified & Bluetooth Low Energy)
- OnePlus 3T with Android 6.0 (Marshmallow)
- Open Source HAPI FHIR

The vital parameter data, SpO2, pulse and blood pressure values, were integrated in accordance to the PCHA design- as well as interface-guidelines (Personal Connected Health Alliance (PCHA), 2016) to implement a fully standards based system. After vital parameter measurement the data should be transmitted by application of the proposed framework component. The system was tested with functional tests as well as interoperability tests using the HL7 FHIR Validator to prove compliance to the HL7 FHIR STU 3.0.1 specification and therefore demonstrate technical validity.

3.4.3 Scenario: Transport Data Integration

The final feasibility study focused on developing a layperson decision support system for health event triggered routing support. Hence, the first step was to retrieve criteria through this scenario by analyzing its work-flow.

Continuous vital parameter measurements taken from chronic patients to gain a health status overview and detect health status changes, was taken as a starting point. The system should support these patients in context of smart cities and trigger personalized route guidance in case unusual health characteristics are measured. This should include propose of quickest paths to hospitals or other health service centers by using public transport systems in Vienna.

In order to measure vital parameters, the patients should be equipped with suitable equipment in form of PHDs. In this context, devices measuring blood pressure i.e. systolic-, diastolic-, mean arterial pressure and puls as well as puls oximeter device for measuring oxygen saturation through application on the finger tip, were applied. A requirement for the system was its capability to be extended with further PHDs like blood sugar monitors for patients suffering from diabetes, Holter-ECGs (portable 24hour electrocardiography devices) or weight scales. Similarly, every day life gadgets like wearable fitness tracker devices as used in the former feasibility study or any kind for supportive devices for rehabilitation purposes supporting chronic patients in any way may be integrated in this approach.

Another crucial design requirement was to support patients mobility at each point in time. Hence, to get maximum acceptance of patients, the use of smartphones, tablets or smartwatchs as personalized control centers was indispensable also for sustainable use of the system. These personalized approach should allow the patient to measure, observe and monitor all healthcare, fitness and transport data information related to this scenario.

The transport related data should be shown through the mobile device in case an medical event was triggered. Hence, the data sources for transportation information

should provide tramway, bus and train opportunities for the user. The most important point was the long-term data storage of the collected information in terms of secondary-use. Taking general practitioners into account or the patients themselves, PHR or general practitioners software systems need to be supported as the former systems like EHR, EMR and HIS, may not be accessible in the current context. In order to support additionally secondary use of data, the interfaces need to provide data in a form which is supported by HIS, EMR or EHR, but also non-medical systems as the transport related data might be stored elsewhere for later analytic approaches. Therefore, the aim was to provide a high degree of technical interoperability, independent of the final place of data storage system. However, since healthcare data is highly sensitive data the system needed to support state-of-the-art security measures to fulfill all necessary details regarding legal requirements of different countries. Summarizing this, the following basic requirements for the study were defined, which correlated with the requirements for selecting standards in this thesis:

- Application of light weight technologies to support mobile patients environments at any time
- Highest degree of extensibility and sustainability through use of international standards
- Multilevel application of interoperability standards i.e. PHDs, mobile devices, professional medical IT systems, non-medical IT systems etc.
- Support for state-of-the-art security measures to fulfill data security and privacy required through international and national laws and regulations
- Non-medical data i.e. public transport information need to be integrated with the same interoperability standard

Based on these results and usage of the selected standards and concepts of this thesis, the IT architecture was designed and prototypical implementation according the

technical requirements. The following hard- and software was used in this feasibility study:

- Nonin Onyx Vantage 9590 Finger-Puls Oximeter (Continua Certified using IEEE 11073 standards family based)
- A&D Medical Blood Pressure Monitor UA-651ble (Continua Certified & Bluetooth Low Energy)
- OnePlus 3T with Android 6.0 (Marshmallow)
- Open Source HAPI FHIR

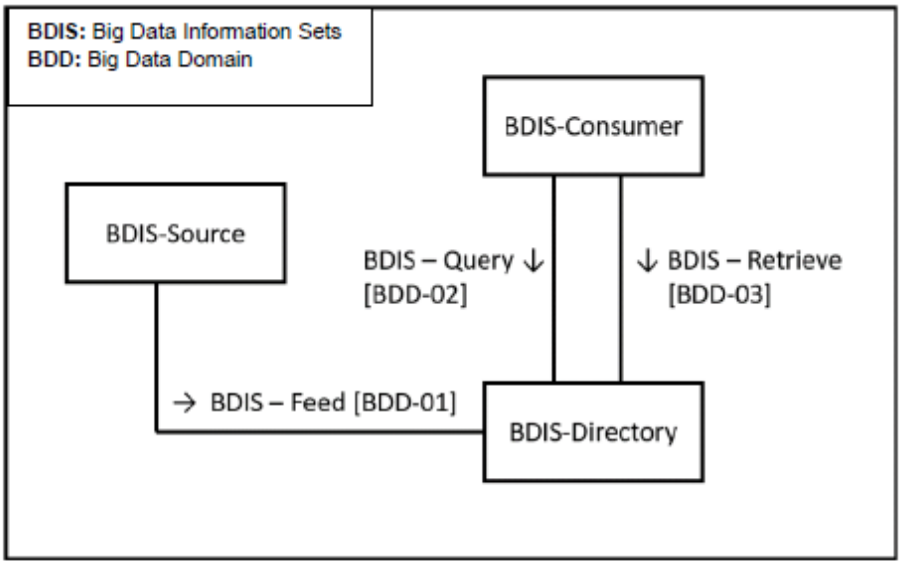
Technical functionality was evaluated by application of unit- and functional tests. Conformance of the prototype in accordance to the interoperability specifications of HL7 FHIR STU 3.0.1 was tested by application of no-peer interoperability tests using the HL7 FHIR Validator.

3.5 Concluding Evaluation

After proving technical feasibility and validity, the final step was to carry out a concluding verification of the thesis procedure, concept and its application. Hence, to verify the approach and outcomes of this study international experts were invited to take part in an experts review process. An inclusion criteria was, that these participating experts had outstanding and longtime experiences in the field of standardization in information technology, eHealth and medical information technology as well as they were willing to shared their views of the applicability and completeness of IHE and HL7 standards for the use of BDIS interoperability.

Table 3.4 – Shows the content of the questionnaire, which was handled to the experts during the experts review.

Are IHE processes applicable to share Big Data Information Sets (BDIS)?



SHARE YOUR EXPERTS OPINION:

Your Profession:	Your Domain:	Experiences in application?
- Engineer ☐	- Health ☐	ITI ☐ RAD ☐
- Management ☐	- Other:	PCC ☐ PCD ☐
- User ☐		LAB ☐ HL7 ☐
		DICOM ☐ Others:

OPTIONAL: I am interested in updates on this topic ☐

NAME:
EMAIL:

Are IHE processes applicable to share Big Data Information Sets (BDIS)?
Do you think that sharing of Big Data Information Sets (BDIS) according to the Actor/Transaction Diagram is feasible?
Do you think that the following standards may be applied in this concept?
What are the main important characteristics a BDIS exchange concept needs to fulfill to become widely accepted?
Which security challenges do you see? Coverable by IHE Profiles?
Which scenarios do you see in your opinion for this application?
Apart from IHE and related standards: Do you see any alternative interoperability specifications for this purpose?

Furthermore, they provided their opinions about the challenges arising from security and legal perspectives. Carrying out this interview process, was done by review of the concept with the questionnaire shown in table 3.4. For implicitly proving reliability and validity of the experts and their input, the questions of the review were separated into two parts. Therefore, some questions required detailed knowledge focusing on IHE and HL7 and can only be answered by experts and others were broader targeted questions.

4 Results

The results of this thesis are structured according the methodical procedure shown by figure 3.1 in chapter 3. As stated previously in section 1.3, results were partially published in scientific journals and on conferences in advance of finalizing this thesis.

4.1 Investigative Analysis of Data Platforms and Formats

Due to legal regulations and policies or commercial concerns of companies, the data of "big data"-platforms and the access to them is limited. Hence, open data platforms and the applied formats were used to provide necessary information about data and its nature in the domains health, environment and transport. As a first step, investigations regarding the amount of available data from the three domains and its accessibility on open data platforms were performed. Therefore, figure 4.1 shows the amount of data sets for the domains of health and transport available on open data platforms in the EU:

This shows that the UK, Germany and Spain were providing the most data sets for

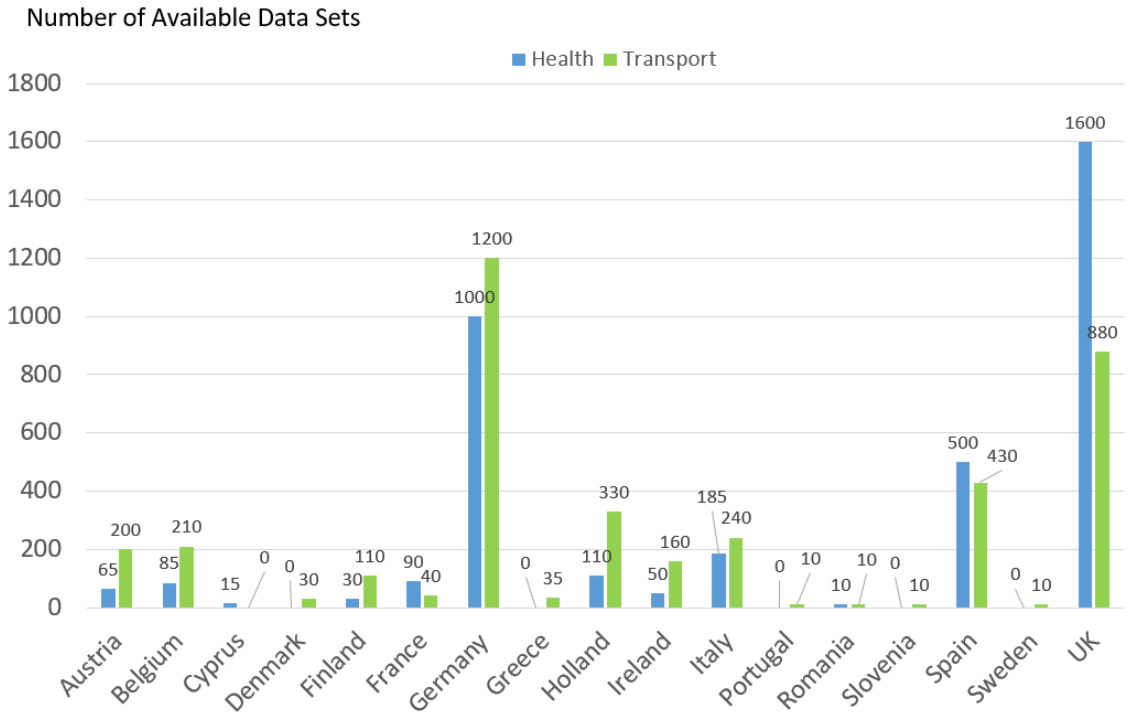


Figure 4.1 – Shows open data platforms in the EU, listed in table 3.1, and the number of health and transport domain related data sets.

these two domains. However, other EU countries didn't provide huge amounts of data at the point of time of this research. Outside of the EU the USA (USA.GOV, 2018), Canada (Government of Canada, 2018) and Japan (Data Go Jp, 2018) were operating open data platforms with considerable amounts of data sets. In case of the latter for example, Japan provided 2068 data sets for environment and 961 data sets of health (Data Go Jp, 2018). However, there was a huge difference in the number of data sets between the three domains. This is underpinned through figure 4.2, which shows the accessible amounts of data sets for the environment domain separated for each EU country investigated. UK and Germany were leading with 8990 and 4821, followed by other countries like Belgium, Netherlands, Italy and Austria providing between 2100 and 1300 data sets.

For this investigation only official governmental platforms were used in this analysis to provide a comparable approach with publicly accessible data. Taking a closer

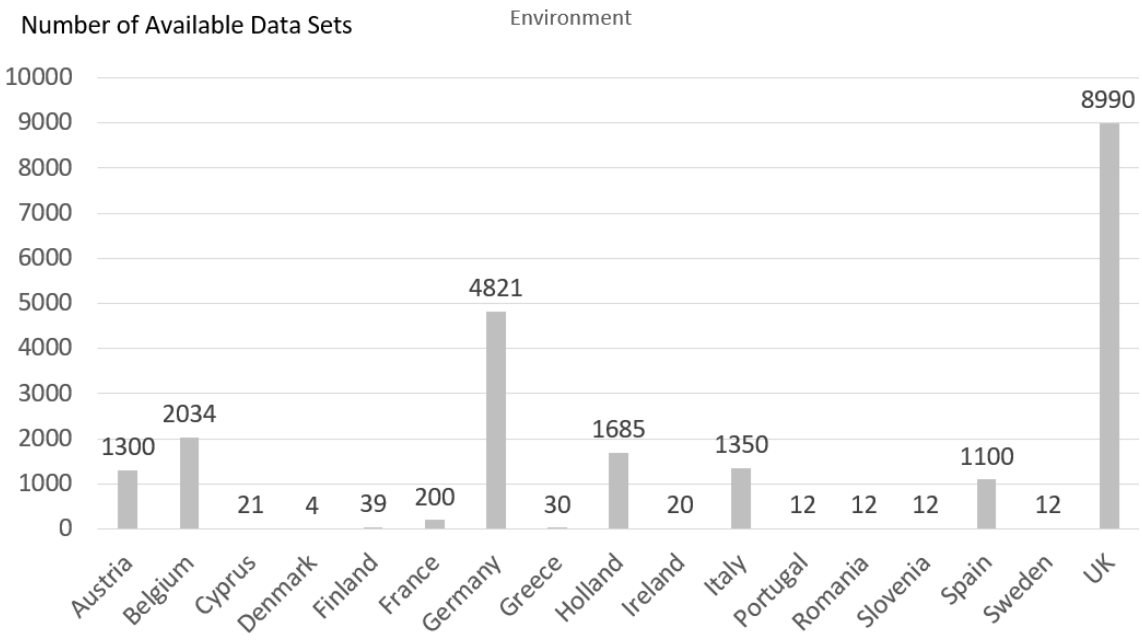


Figure 4.2 – Shows open data platforms in the EU, listed in table 3.1, and the number of environment domain related data sets.

look at the data sets themselves shows that access to the data was furthermore restricted in some cases due to license restrictions. Nevertheless, more interesting were the formats of the data shown in figure 4.3. This figure shows the percentage breakdown, between the most common file formats applied in the respective areas of health, environment and transport.

The European Data Portal (European Data Portal, 2018) was used as a basis in addition to the table 3.1, as it includes all data formats found in the selected platforms during research. The total amount of data sets include 8303 for health, 7826 for transport and 41468 for environment. The top ten applied data formats were selected to narrow the context of the thesis, as only for the health domain more than 100 different formats were used and provided by the platform. Most of these formats only had between 1 to 5 data sets, what made its integration not useful and it furthermore indicated the necessity of reducing to few standardized formats. The sum of datasets in other formats were 1173 for health, 8216 for environment and 927 for transport data. Some examples of file-formats were docx, url, exe, ppt,

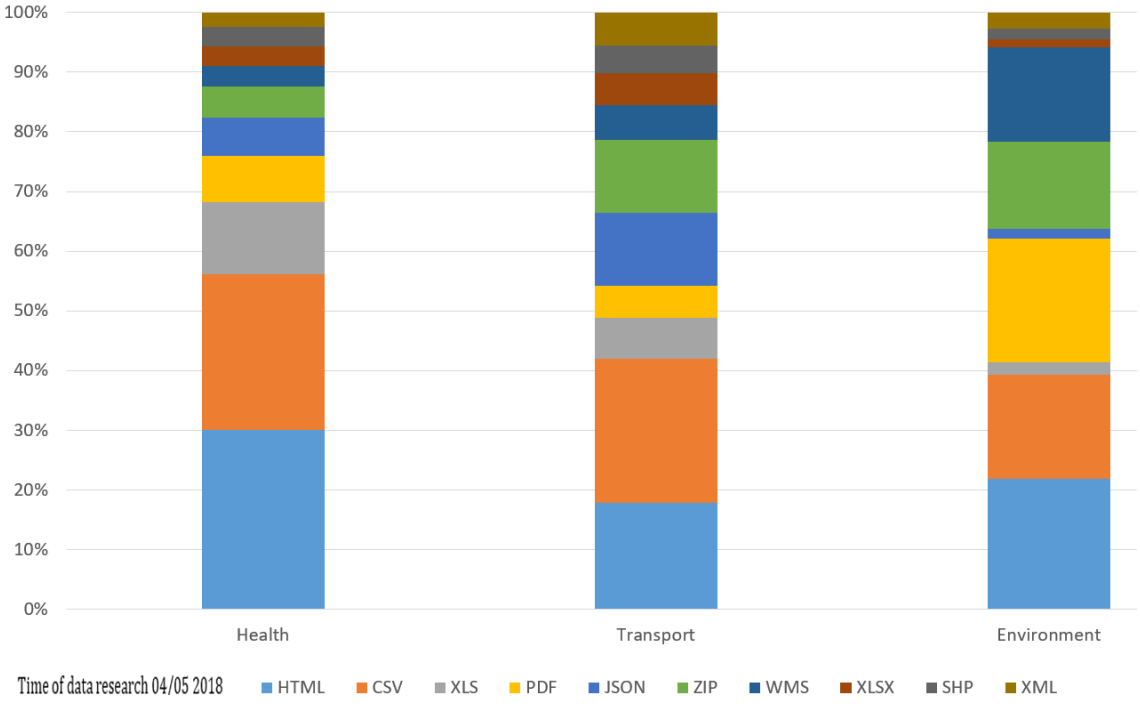


Figure 4.3 – Shows the percentage breakdown between the most common file formats applied in the respective domains of health, environment and transport.

xsd or ASCII. Figure 4.3 shows the amount of data sets available in the specific formats in the respective domains. Hence, most used common formats were HTML, CSV, XLS and PDF. However, using XML or JSON would allow a much better applicability to enable automated processing, but as indicated in the graph, both were used to a low degree of percentage. For health domain XML were 2.51% and JSON 6.27%, for transport domain XML were 5.56% and JSON 12.16% as well as for environment XML were 2.72% and JSON 1.68%. During the investigation a perception was, that most open data platforms were using the "Comprehensive Knowledge Archive Network" (CKAN) web-based management system (CKAN, 2018). This is, independent of the described formats, of advantage as it strongly focuses on the usage of RESTful web services. However, it didn't provide support for syntactical and semantical interoperability for the whole range of the big data information set i.e. from the meta-data down to the raw data.

4.2 Selection of Standards and Technical Basis

Table 4.1 provides an overview about the selected standards and frameworks, which were extracted from literature and meet the defined criteria of the user requirements. Therefore, standards from HL7 are widely spread in the medical IT domain and cover information exchange in terms of text based messages in clinical environments, with HL7 v2 as well as its successor HL7 v3. As a standard for definition and exchange of electronic clinical documents, HL7 CDA is a widely used prominent candidate. The most recent approach for exchange of clinical information, however is HL7 FHIR, which uses modular and well defined data "resources" to do so. However, IHE provides profiles and related processes for medical applications in several medical domains. Most prominent areas are the IT infrastructure, radiology and laboratory domains, where technical frameworks describe these processes, needs and standards to use. Furthermore, the PCHA focuses on a whole communication chain from PHDs up to medical IT infrastructure. The latter is strongly connected to using IHE profiles. However, PCHA uses standards from HL7, IEEE (e.g. X7) Standards family or OASIS (Web-Communication Standards). Finally, OpenEHR is based on archetypes providing concepts to create, manage and exchange health data for EHRs and setup as an open standard.

4.2.1 User Requirements

Table 4.1 shows these standards opposed to the extracted user requirements. The table shows that HL7 FHIR was the most promising standard regarding data structuring on a protocol level to fulfill these user requirements, as it only showed low support regarding security/privacy and for process definitions. However, this might be compensated by IHE and openEHR, which showed support regarding these two requirements. On the other hand openEHR showed lacks in support for integration of non-medical data and applicability to mobile environments. IHE provided a big contrast here, as it was much more supportive in this areas. Along to IHE and HL7

FHIR also PCHA indicated to be a good candidate, as it provided strong support to all requirements except of lower support to integration of non-medical data and process definitions. Hence, taking these findings into account, the three standards of HL7 FHIR, IHE and PCHA were selected for more detailed investigation in the next step focusing on the detailed requirements as these were most important ones.

Table 4.1 – Shows the investigated interoperability standards and frameworks, selected in accordance to the user requirements.

Standards & Frameworks vs. User Requirements	HL7 V2/V3	HL7 CDA	HL7 FHIR	IHE	PCHA	openEHR
Integration of Non-Medical Data	∼	∼	+	∼	∼	∼
Enables Support for Mobile Environments	+	∼	+	+	+	∼
International Standard(s)	+	+	+	+	+	+
Support for Multiple Interface Levels	+	+	+	+	+	∼
State of the Art Technology Support	∼	+	+	+	+	+
Security/Privacy Measures Supported	∼	∼	∼	+	+	+
Support for Process Definitions	-	-	-	+	-	+
Legend: + supported ∼partially supported - not supported						

4.2.2 Detailed Requirements

In order to finalize the decision making process this section provides the results of the analysis, which finally lead to table 4.2. Based on the findings described in the previous chapters, table 4.2 shows the requirement groups (technical, security and

legal) and the resulting related functional requirements as well as their coverage by the previously selected SDOs, frameworks and standards.

The sub-column "Covered by-SDO", indicates the level of support by most promising SDOs, IHE as well as PCHA i.e. their frameworks and processes. Therefore, PCHA shows good support for most technical requirements, but redirects to IHE when its about support for security- or legal requirements. Nevertheless, IHE enables strong support for all requirements except of a lower impact on the legal ones. Taking a look at the second sub-column "Covered by-Profiles", this provides a more detailed view on IHE profiles i.e. parts of specific IHE technical frameworks covering the listed requirements to a certain degree. For the exchange of clinical documents (e.g. CDA) and HL7 v2/v3 message based exchange, the profiles XDS/XDR/DEC are very well applicable. Profiles covering the security requirements to a high degree are ATNA, XUA, IUA, EUA and BPPC. In the last sub-column "Covered by-Standards" the focus lies on the standards, which provide the best coverage of the defined requirements. From a technical perspective HL7 FHIR was the most promising standard, where from a security point of view different base technologies are stated. Most of these technologies are furthermore referenced by IHE and the listed profiles.

Organizational requirements were identified to be criteria like adaptability and flexibility for all kinds of IT systems. This also applies to the interconnection of data sinks in terms of big data information set exchange. It is necessary to adapt new needs, introduce improved technologies to ensure successful and long time operation and systems usage. This is of high importance in this case as data is collected over long periods of time and therefore sustainability is another very important factor.

Furthermore, building components for exchanging big data information sets need to be integrated in existing systems with a low burden as building completely new systems is not possible due to limited resources. Additionally, this adds the need for configurable approaches, as there are several different systems and especially in terms of not centrally stored information the diversity factor is huge. Hence, system administrators need to configure the components specifically to the needs of different user groups, e.g. laypeople, advanced users, technology-savvy people etc..

Therefore customization is a crucial fact to provide a successful approach from an organizational point of view.

Table 4.2 – Provides an overview of SDOs, profiles and standards covering the defined requirements.

Requirement Groups	Requirements	Covered by		
		SDO	Profiles	Standards
Technical Requirements	Future-Proof	+IHE, +PCHA	+XDS/XDR/DEC	+FHIR
	International Standards	+IHE, +PCHA	+XDS/XDR/DEC	+FHIR
	Modular, Flexible, Extensible	+IHE, +PCHA	∼XDS/XDR/DEC	+FHIR
	Non-Medical Data Integration	+IHE, ∼PCHA	∼XDS/XDR/DEC	+FHIR
	Service Oriented	+IHE, +PCHA	+XDS/XDR/DEC	+FHIR
	Mobile Environments	+IHE, +PCHA	∼XDS/XDR/DEC	+FHIR
	Multi-Level Application	+IHE, +PCHA	+XDS/XDR/DEC	+FHIR
Security Requirements	Software Component Authentication	+IHE	+ATNA	+Certificates
	User Authentication	+IHE	+EUA	∼Kerberos
	Logging	+IHE	+ATNA	+Sys-Log
	Authorization and SSO	+IHE	+XUA/IUA	+SAML, WS-Trust, OAuth
	Encryption	+IHE	+ATNA	+Certificates
Legal & Regulatory Requirements	Access Policies	∼IHE	∼XUA	+XACML
	Person/Patient Consent	∼IHE	∼BPPC	∼XML (CDA)
	Regulations	∼IHE	-	-
Legend: + supported ∼partially supported - not supported				

Summarizing the results of this chapter, the strategy of IHE and the standard HL7 FHIR were most promising to fulfill all of the defined requirements through combination. Furthermore, HL7 FHIR enables the support for semantic interoperability through definition of tags (e.g. "codeableConcept") especially made

to include codes from code systems for. This enables machines to automatically interprete the data based on well defined code-systems as well as value-sets. HL7 FHIR is strongly connected to RESTful architecture and enables the use of state-of-the-art security mechanisms (e.g. OAuth), which is also used by IHE (e.g. IUE profile). Hence, the decision was made to use a combinatorial approach for the next steps in this thesis by selecting HL7 FHIR as a standard, RESTful architecture as a technological basis and IHE's strategic approach for defining profiles, actors and transactions (i.e. models and processes).

4.3 Business Domain Overview and Framework Components

Systems and devices in the context of mobile computing like smartphones, tablets, smartwatches or any wearable sensors can be a highly productive source of data. Crowdsourced approaches can use smartphone ECG applications for the diagnose of health incidents like syncope (Nyotowidjojo et al., 2016) or the use of smartphone-based data to overcome the lack of missing infrastructure in the context of floating car systems (Briante et al., 2014). Hence, the expression big data is an important term in context of this data as the term refers generally to large volumes of data generated and stored, which can't be handled with common tools for storage and analysis.

The BDD is a domain related to data from several other domains, which adds a huge complexity to this approach as the nature of data varies and the requirements of data security and privacy is different through national and international regulations. At least the latter is positively influenced when it comes to open data approaches, were data is already open to everyone. Although the "big data revolution" does not simply refer to the quantity of the data, but moreover to the ability to handle and interpret the data, the management of Big Data Information Set(s) (BDIS(s)) in an efficient way is of high interest to improve big data analytic outcomes through prepared data and therefore saving valuable time by reduction of preprocessing.

Taking a closer look on data sinks in the BDD, for example governmentally open data platforms, this shows extremely diverse data formats although progress towards harmonization is visible. There are several tools in context of big data like Hadoop (Apache, 2018), Cloudera (Cloudera, 2018), MongoDB (MongoDB, 2018) or IBM Watson (IBM, 2018), which are used to store, search, prepare and analyze BDIS more efficiently. However, a common problem in the BDD is that most data needs to be preprocessed to conform to the tools proprietary formats for storing, searching and analysis. Therefore the reduction of preprocessing time of the data, through definition of a concept for BDIS sharing and a syntactical and semantical format for import/export, is a valuable contribution. This is shown by the example of IHE, HL7, PCHA and other SDOs who define standards and processes for interoperable data exchange, which decrease implementation effort, improve systems efficiencies and quality as well as reduce costs at the example of the healthcare domain (Maeng et al., 2014; Urbauer et al., 2012). Hence, to take the first step in solving this issue, a concept and related components to improve interoperability in the BDD were proposed. This was done according the process used by IHE and using the medical interoperability standard HL7 FHIR. The resulting IBD-Profile and the results of the studies to prove technical feasibility are described after the scenarios describing examples of application.

4.3.1 Scenarios

The support of sharing of BDIS from different domains like health, environment or transport using internationally applied standards from the healthcare domain is the main intention of this thesis. Hence, the following scenarios describe application examples. Later, these were prototypically implemented and technically tested through functional and conformance tests in section 4.4.

Scenario: Healthcare Data Integration

In this scenario, shown in figure 4.4, elderly people use wearable activity trackers to measure their vitality and mobility in their homes or outside.

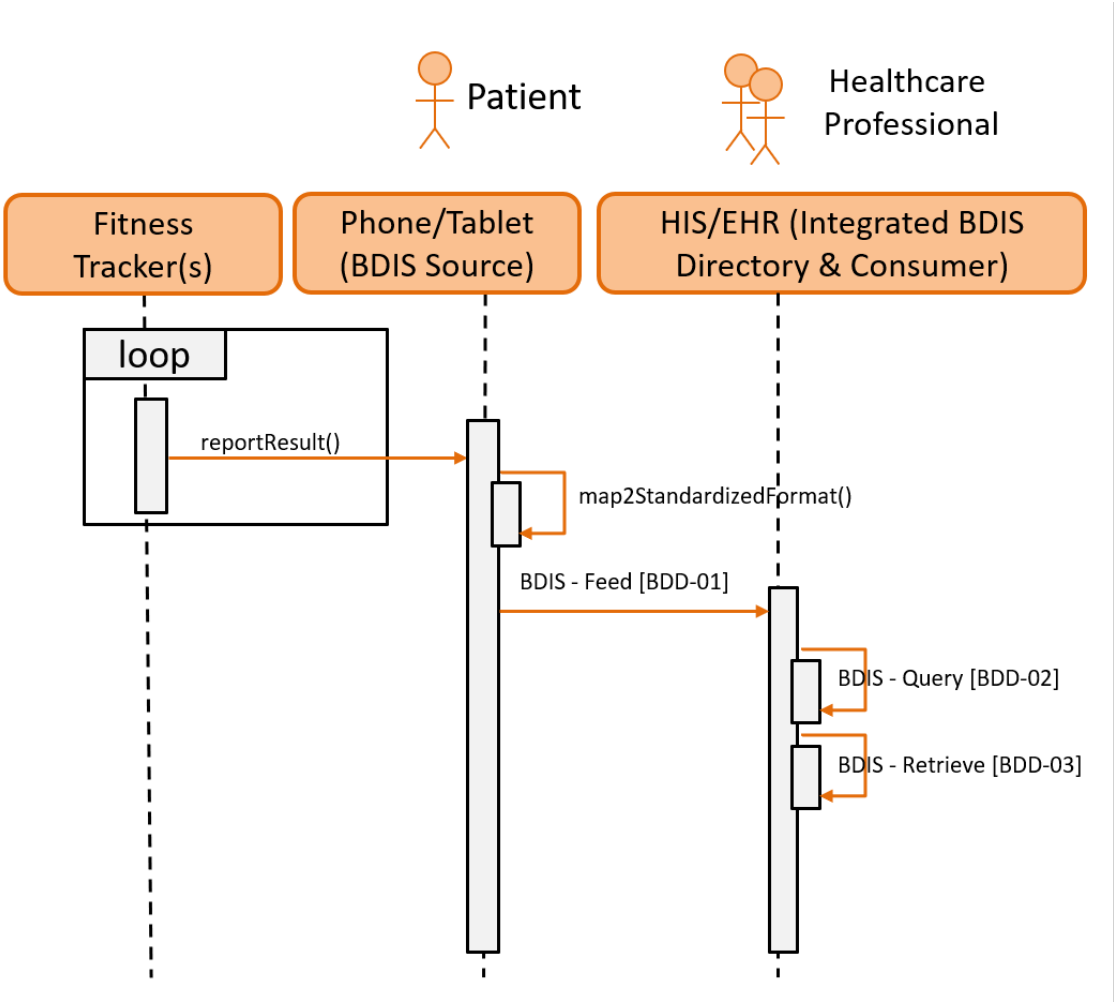


Figure 4.4 – This sequence diagram describes the components and the workflow of the elderly living independently scenario, which focused on the use of wearable fitness tracker data to observer movement and vitality behavior.

Indicated on the left, wearable fitness tracker are used to measure values like movement behavior, heart rate continuously and pulse. The measured data is reported on-demand or at specific points in time in proprietary form or align

to standards (e.g. X73) to any kind of mobile device (smartphone or tablet etc.). Through an internal mechanism, the Software (i.e. Android/iOS-App) shall transform the data to meet the requirements of the BDIS-Source i.e. an HL7 FHIR observation resource with the specified BDIS-Extension shall be generated including the measurement and context related data. In the next step the BDIS-Source shall initiate the BDIS-Feed [BDD-01] transaction, which transmits the resource to an BDIS-Directory i.e. a HL7 FHIR STU 3.0.1 conform server for storage of the data. At this point in time a healthcare professional or scientist may use the BDIS-Directory to investigate on the BDIS data or make any kind of analysis. In this case a BDIS-Consumer is integrated with the BDIS-Directory trough Actor grouping, but it can also be separated. This is done through using BDIS-Query [BDD-02] for searching and subsequently BDIS-Retrieve [BDD-03] to collect several amounts of data sets fulfilling the search criteria. However, the query parameters can be used in accordance to HL7 FHIR STU 3.0.1 for searching HL7 FHIR observation resources or by specifically using the BDIS-Extension's semantic criteria.

Scenario: Environmental Data Integration

Figure 4.5 shows the sequence diagram describing the workflow within a scenario for a chronic pollen allergy patient support system (CPAPSS). In order to measure vital parameters, the patients should be equipped with suitable equipment in form of PHDs. In this context, a device measuring blood pressure i.e. systolic-, diastolic-, mean arterial pressure and puls, as well as puls oximeter device for measuring oxygen saturation through application on the finger tip, may be applied.

The data is either continuously i.e. each hour or sporadically transmitted through standardized medical devices according to PCHA or by proprietary devices and protocols to an mobile device (e.g. smartphone, tablet etc.). On the mobile device the data is mapped to the specified HL7 FHIR STU 3.0.1 observation resource format and semantics connected to the BDIS-Extension, so that the BDIS-Source can subsequently transmit the data through the BDIS-Feed [BDD-01] transaction. As a receiving instance, a HIS or EHR system including a BDIS-Directory actor may

receive and store the BDIS. Hence, this storage process triggers another proprietary service for receiving pollen exposure data to build a context for allergy warning system or allergy research based on BDIS. Therefore, the HIS, EHR or any software module requests the data from the proprietary server (including a proprietary protocol). After the response, the data is connected through an internal process to the vital parameters previously received and stored through BDIS-Feed [BDD-01] transaction.

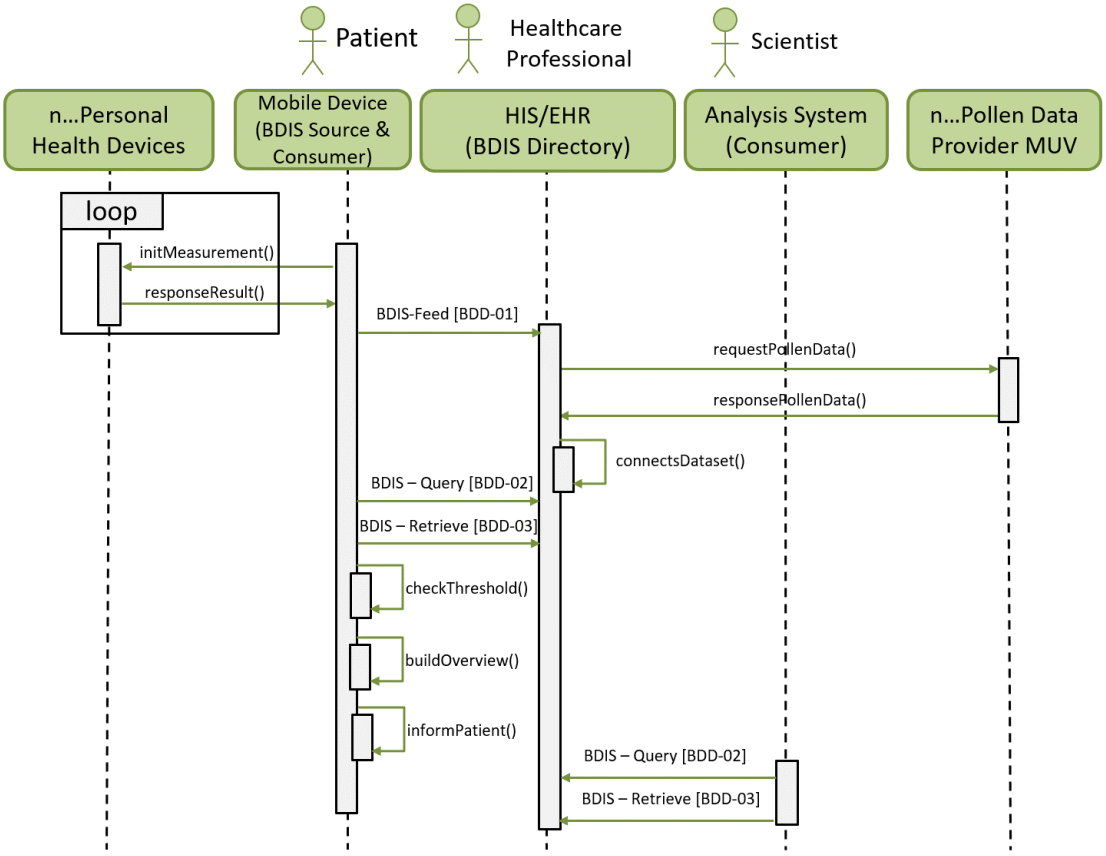


Figure 4.5 – This sequence diagram describes the components and the workflow of the chronic pollen allergy patient support system (CPAPSS), which combines vital parameter data with environmental pollen exposure data.

Once this process is done, on the one hand the mobile device App may use BDIS-Query [BDD-02] and subsequently BDIS-Retrieve [BDD-03] to receive the

BDIS indicating pollen exposure data as well as its relation to health data (vital parameters). This is done through an App internal processes, which checks the thresholds and builds an informative overview for the patient. In parallel a scientist may use similarly BDIS-Query [BDD-02] and BDIS-Retrieve [BDD-03] to receive BDIS with health and environmental data for investigation the pollen exposure for a specific type of pollen i.e. Betula exposure, between the first of March until the fifth of March and the related patient health data.

Scenario: Transport Data Integration

This scenarios describes a "medical event based route guidance system", which is shown in figure 4.6. The basis for this scenario is that continuous vital parameter measurements are taken from chronic patients to gain a health status overview and detect health status changes.

In order to measure vital parameters, the patients should be equipped with suitable equipment in form of PHDs. In this context, a device measuring blood pressure i.e. systolic-, diastolic-, mean arterial pressure and puls, as well as puls oximeter device for measuring oxygen saturation through application on the finger tip, may be applied. A requirement for the system is its extensibility, with further PHDs like blood sugar monitors for patients suffering from diabetes, Holter-ECGs (portable 24hour electrocardiography devices) or weight scales. Similarly, every day's life gadgets, like wearable fitness tracker devices or any kind for supportive devices for rehabilitation purposes supporting chronic patients may be integrated.

The data is either continuously i.e. each hour or sporadically transmitted through standardized medical devices according to PCHA or by proprietary devices and protocols, to a mobile device (e.g. smartphone, tablet etc.). Trough an internal process on the App, the data is checked for thresholds i.e. alarm signals through high blood pressure and the patient is informed independent of reaching a dangerous threshold or not. In case an alarm is indicated, subsequently public transport services IT systems are requested through proprietary processes for data (this could

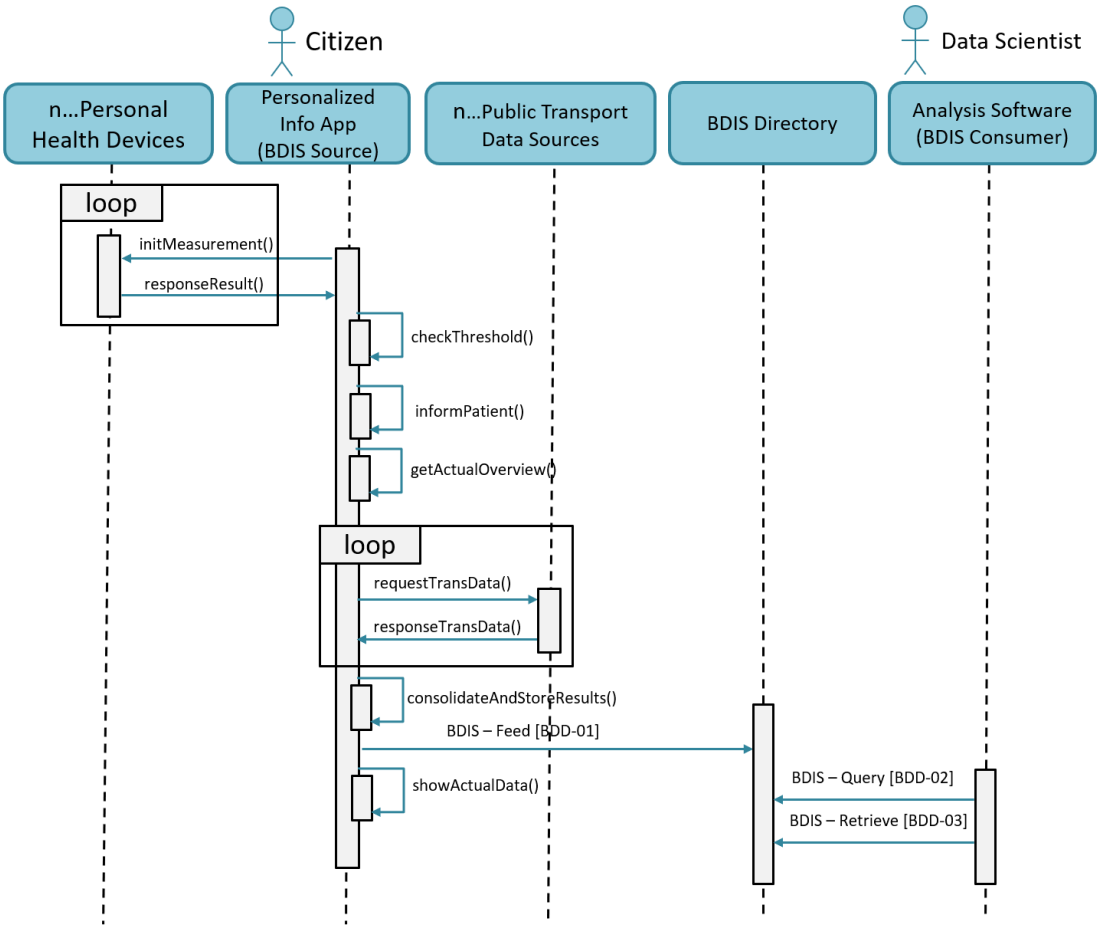


Figure 4.6 – This sequence diagram describes the components and the workflow of the medical event triggered route guidance scenario, which combines data from the domains of transport and health.

be done via using IBD-Actors and transactions too!) according to the next possible departure times of public transport lines near the patients place. Once the data is responded by the public transport data sources, it is consolidated and connected with the health related data. This information, is then presented to the patient who is supported to make a faster decision in choosing a transport to healthcare support.

For later analysis purposes, the App builds a HL7 FHIR STU 3.0.1 observation resource with specified BDIS-Extension and forwards it to the BDIS-Source for sending an BDIS-Feed [BDD-01] transaction to the BDIS-Directory. The latter extracts and stores the data in its database. Once this is done, a scientist may trigger

a BDIS-Query [BDD-02] and subsequently BDIS-Retrieve [BDD-03] to receive the BDIS indicating transport relevant information and relation to health data. Next the the data is investigated on finding ways to improve patient flows to hospitals via public transport systems.

4.3.2 IBD-Profile

Under the term "digitization 2.0" smartphones, tablets, smart watches or any kind wearable sensors are highly productive sources of generating huge amounts of data in context of Internet of Things (IoT). Similarly, companies and research institutions produce and use massive amounts of data for research of new technologies or commercial goals. There are expectations that interconnecting data from different domains like e.g. healthcare, environment and transport, might lead to new findings for improving several aspects of life, like better treatment of diseases or more efficient paths of cares. However, a huge problem is the diversity of data formats although this is a very important quality factor to make data exchangeable and comparable. The "Interoperable Big Data Information Set (BDIS) Directory"-Profile (IBD-Profile) focuses on providing a fundamental basis for syntactical and semantical interoperability of BDIS to be exchange between IT systems for data analysis. A critical requirement derived through the Big Data Domain (BDD) is that sources of data could be huge IT systems like open data platforms or any kind of mobile device within the IoT aspect: Mobile devices in a crowdsourced approach may generate huge amounts of data, usable for data analytics in case the data is structured and semantical interpretable. The use of state-of-the-art lightweight technologies are an important requirement in this context. Therefore, the IBD-Profile is based on using RESTful architecture together with HL7 FHIR STU 3.0.1 as a base standard to support syntactical and semantical requirements fostering interoperability. The most important aim of the IBD-Profile is to use the HL7 FHIR STU 3.0.1 specification to a maximum degree, without changing the specification, but using its provided model of FHIR extensions to enable support for BDIS sharing. The IBD-Profile includes three actors and three transactions as

shown in figure 4.7, providing the functionality of storing, querying and retrieving BDIS in the specified standards based format described in the following chapters.

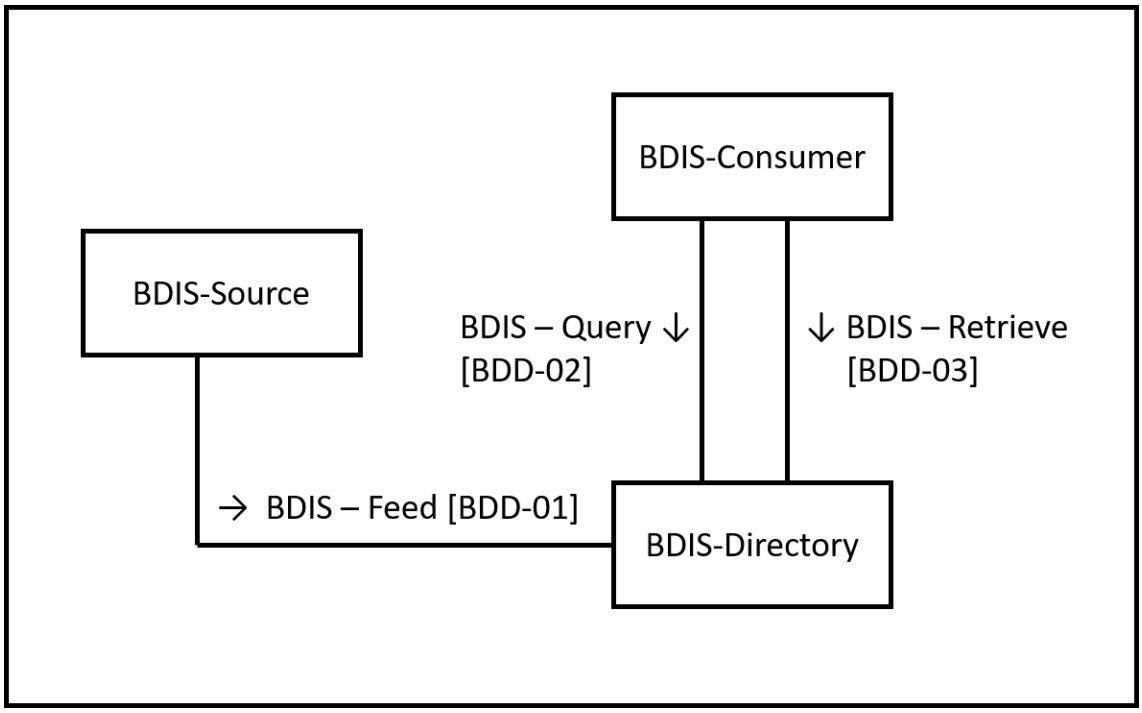


Figure 4.7 – Shows the "Interoperable Big Data Information Set (BDIS) Directory" IBD-Profile with its actors and related transactions.

IBD-Actors

The BDIS-Source actor is responsible for building and transmitting BDIS, according to the described format, by sending a BDD-01 transaction to the BDIS-Directory actor. The latter is responsible for receiving BDIS and storing the raw data and its context information. From a technical point of view, this could be a HL7 FHIR server supporting STU 3.0.1 specification. On the one hand, the BDIS-Consumer sends a BDD-02 transaction to the BDIS-Directory to query for purpose-appropriate BDIS. Based on these queries or without previously querying but based on known IDs derived through other processes, the BDIS-Consumer furthermore uses the BDD-03

transaction to retrieve BDIS from the BDIS-Directory for further analysis. IBD-Actors can be grouped according to the requirements e.g. BDIS-Directory with BDIS-Consumer in case the data storage and the analysis software is a combined product. Some examples are described in the related scenarios section 4.3.1. Table 4.3 shows the actors, related transactions and its optionalities.

Table 4.3 – Provides an overview of the IBD-Profile's actors and transactions.

Actors	Transactions	Optionality
BDIS Source	BDIS-Feed [BDD-01]	R
BDIS Consumer	BDIS-Query [BDD-02]	O *
	BDIS-Retrieve [BDD-03]	O *
BDIS Directory	BDIS-Feed [BDD-01]	R
	BDIS-Query [BDD-02]	R
	BDIS-Retrieve [BDD-03]	R
* at least one shall be implemented		

IBD-Transactions

All described transactions (BDD-01, BDD-02 and BDD-03) are based on the standards as shown in table 4.4. The message content shall generally be transmitted either via using JSON or XML i.e. with a supporting media-type "application/fhir+json" or "application/fhir+xml".

Table 4.4 – Shows the standards applied in the described transactions of the IBD-Profile.

HL7 FHIR	HL7 FHIR STU 3
RFC2616	HTTP/1.1
RFC7540	HTTP/2
RFC3986	Uniform Resource Identifier (URI): Generic Syntax
RFC6585	Additional HTTP Status Codes

The BDIS-Feed transaction (see figure 4.8) is initiated by building a BDIS structure and then sending a BDIS-Feed request message to the BDIS-Directory actor by using

the HTTP Post methods for creating a new resource on the server. However, in terms of the BDIS-Feed transaction the basic data shall be stored on the one hand in an HL7 FHIR observation resource, but on the other hand the BDIS-Extension resource (see 4.3.3) shall contain the context relevant BDIS information. Therefore, the transaction-mechanism of HL7 FHIR STU 3.0.1 shall be used to support necessary data integrity. Hence, the following general URL shall be used according to HL7 FHIR STU 3.0.1:

POST [base] ?_format=[mime-type]

A successful retrieval i.e. all resources are created successfully, shall be replied by sending an HTTP response with Code 200, or in case of not successfully storing data, the status codes defined by HL7 FHIR STU 3.0.1 shall be applied (see HL7 International (2018)).

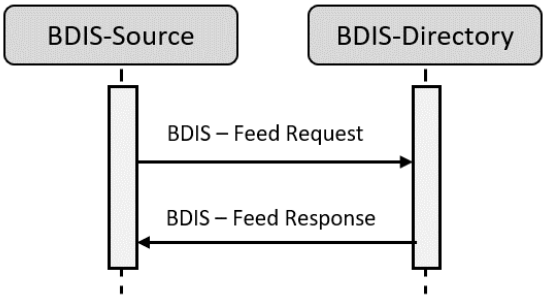


Figure 4.8 – Shows the interaction diagram for the BDIS-Feed [BDD-01] transaction.

Figure 4.9 shows the interaction diagram of the BDD-02 transaction for querying data in a BDIS-Directory.

Hence, the BDIS-Consumer sends a BDIS-Query request message, which is a HTTP Get request parametrized with the specific query parameters. This parameters shall be used as defined by the RESTful API in the HL7 FHIR STU 3.0.1 and also be valid for the defined BDIS-Extension (see 4.3.3) as well. The general URL according to this shall be:

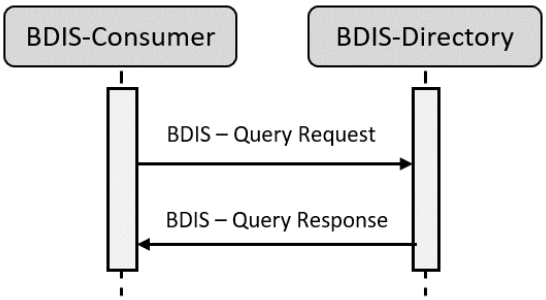


Figure 4.9 – Shows the interaction diagram for the BDIS-Query [BDD-02] transaction.

[base]/Extension?<query>

The BDIS-Query response message shall include a HTTP 200 status code, together with the resource or resources found via the applied query parameters. In case that the BDIS-Query response is not successful, the status codes (see HL7 International (2018)) defined by HL7 FHIR STU 3.0.1 shall be used.

As shown in figure 4.10, the BDIS-Retrieve request message initiates the retrieval of a referenced HL7 FHIR observation resource integrating the BDIS-Extension (see 4.3.3) information by sending a HTTP Get request to the BDIS-Directory actor.

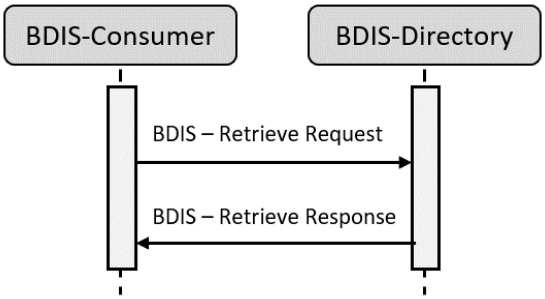


Figure 4.10 – Shows the interaction diagram for the BDIS-Retrieve [BDD-03] transaction.

In case a HL7 FHIR observation resource can be returned successfully, the response shall include a HTTP 200 status code and the data itself. In case that the BDIS-Retrieve is not successful, the response shall include the status codes (see HL7 International (2018)) defined by HL7 FHIR STU 3.0.1.

4.3.3 BDIS structure and BDIS-Extension definition

Figure 4.11 represents the defined structure in context of the IBD-Profile to share BDIS with its defined transactions.

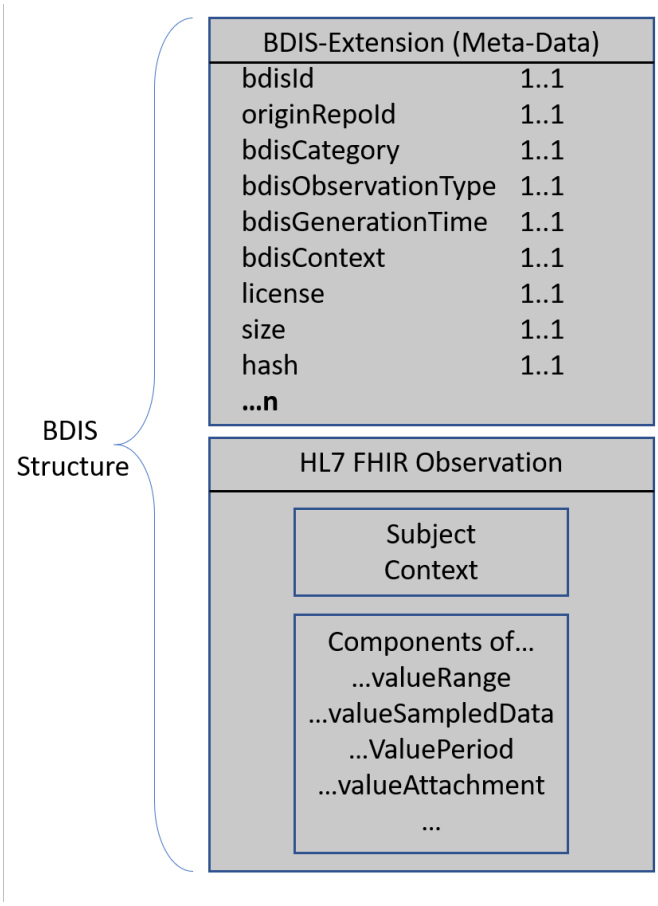


Figure 4.11 – Shows the generic BDIS structure containing the BDIS-Extension (common meta-data) and the HL7 observation resource for data storage in different forms.

A BDIS was separated in the BDIS-Extension used to carry core meta-data like an identifier for the BDIS, where it was originally stored, license about the data usage, its size and more. Below the BDIS-Extension a HL7 FHIR observation resource was used to carry additional data describing the content. Examples are the authoring organization and/or person as well as additional codes supporting semantics to describe the data. Subsequently, this is followed by the raw data itself,

which can be integrated using HL7 FHIR's component approach to form packages of data depending on their different formats. These formats are ranges, periodic data, attachments in form of Base64 encoded data or zip-file and all other formats as defined by HL7 FHIR STU 3.0.1.

For combining data from healthcare with environmental data for example, several "components" can be used. Hence, a component including systolic, one for diastolic, one for mean arterial pressure and one for pollen exposure data are integrated in the HL7 FHIR Observation. Furthermore, the components concept adds the possibility to add data in other formats, as shortly mentioned before. That means that unstructured data might also be added in the BDIS structure. How the data is translated in the BDIS structure is out of scope of the IBD-Profile, as this is an actor internal requirement. The BDIS structure therefore provides a structured concept but also allows integration of non-structured data, although the latter should be clearly avoided.

Figure 4.12 to 4.16 show the most important parts of the definition of the specified BDIS-Extension. A system implementing the IBD-Profile shall include this link to the BDIS-Extension definition in its implemented HL7 FHIR observation resource for data transmission. The specification has a total of 1506 lines of code and is published separately from the STU 3.0.1 package under the following link:

http://healthyio.technikum-wien.at/oid/bdis-extension.xml

The BDIS-Extension was defined according the HL7 FHIR "Resource Structure Definition" specification. To perform correct validation during the feasibility studies, this extension was added to the "extension-definition.xml"-file in the specification package for STU 3.0.1. Figure 4.12 shows the defined meta-data elements in the BDIS-Extension. The identification and storage of the bdis-extension is defined in line 86758 as well as 86762 by its URL. Line 86764 to 86768 show that the definition is currently under draft version for the STU 3.0.1 (Version 3.0.1 as it includes one technical errata). The implementation of the BDIS-Extension is only allowed in

an HL7 FHIR observation resource, which is shown in line 86772 and its type is a complex-type (line 86769) as it is a container element for the specific sub-extensions.

```
86757   <entry>
86758     <fullUrl value="http://healthyio.technikum-wien.at/oid/bids-extension"/>
86759     <resource>
86760       <StructureDefinition xmlns="http://hl7.org/fhir">
86761       <id value="bdis-information"/>
86762       <url value="http://healthyio.technikum-wien.at/oid/bdis-extension"/>
86763       <name value="bdisExtension"/>
86764       <status value="draft"/>
86765       <date value="2018-01-07"/>
86766       <publisher value="URBAUER"/>
86767       <description value="Additional information about Big Data Information Set Exchange"/>
86768       <fhirVersion value="3.0.1"/>
86769       <kind value="complex-type"/>
86770       <abstract value="false"/>
86771       <contextType value="resource"/>
86772       <context value="Observation"/>
86773       <type value="Extension"/>
86774       <baseDefinition value="http://hl7.org/fhir/StructureDefinition/Extension"/>
86775       <derivation value="constraint"/>
```

Figure 4.12 – Shows the meta-data elements in the BDIS-Extension definition.

The BDIS-Extension includes both definitions, a snapshot element for providing a standalone version usable without considering the base StructureDefinition as well as a differential element shown in figure 4.13.

In this case the element is described relative to the StructureDefinition. The latter is used for explanation at is more compressed, but still includes all relevant information. Hence, line 87447 to 87448 show that the BDIS-Extension must not, but can be applied as often as needed. Furthermore, from line 87451 to 87478 the first sub-extension is defined for describing the category of the BDIS i.e. health, environment, transport, etc.. Therefore its place in the structure of the BDIS-Extension (line 87452) is descriptive information (line 87453 to 87455) and its number of occurrence i.e. minimally once and maximally once are described as well as its type being similarly an ”Extension” (line 87459). Next, lines 87462 to 87465 are used to stated that no sub-extension is allowed, whereby lines 87466 to 87471 are used to describe the attribute value ”uri” of the bdisCategory-element to be filled with the fixed value ”bdisCategory”. Finally, the content of the bdisCategory is described to be a CodeableConcept-element, including code-system, code and display name as well as an ”text”-value (lines 87473 to 87478). As the definition

```
87442       <differential>
87443         <element id="Extension">
87444           <path value="Extension"/>
87445           <short value="Extention for adding big data information set information"/>
87446           <definition value="Adds additional information regarding a big data information set."/>
87447           <min value="0"/>
87448           <max value="*"/>
87449         </element>
87450
87451         <element id="Extension.extension:bdisCategory">
87452           <path value="Extension.extension"/>
87453           <sliceName value="bdisCategory"/>
87454           <short value="BDIS data category"/>
87455           <definition value="Category of BDIS."/>
87456           <min value="1"/>
87457           <max value="1"/>
87458           <type>
87459             <code value="Extension"/>
87460           </type>
87461         </element>
87462         <element id="Extension.extension:bdisCategory.extension">
87463           <path value="Extension.extension.extension"/>
87464           <max value="0"/>
87465         </element>
87466         <element id="Extension.extension:bdisCategory.url">
87467           <path value="Extension.extension.url"/>
87468           <type>
87469             <code value="uri"/>
87470           </type>
87471           <fixedUri value="bdisCategory"/>
87472         </element>
87473         <element id="Extension.extension:bdisCategory.valueCodeableConcept">
87474           <path value="Extension.extension.valueCodeableConcept"/>
87475           <type>
87476             <code value="CodeableConcept"/>
87477           </type>
87478         </element>
```

Figure 4.13 – Shows the definition of the first sub-extension element, describing the "bdisCategory".

and structure of the "bdisObservationType" sub-extension element is similar to the specification of the "bdisCategory", this is not explicitly shown in another figure. The purpose is to describe if the data is of historical, actual or predictive nature.

Figure 4.14 shows the definition of the "bdisGenerationTime" sub-extension element. This is used to specify the generation time of the BDIS. The structure definition is similar to these of the "bdisCategory", which is shown in lines 87509 to 87530. Furthermore, the element is a sibling to the bdisCategory sub-extension and its fixed specifier is "bdisGenerationTime". However, the lines 87531 to 87536 are describing that the data inside this element shall be a time-stamp (dateTime).

The purpose of the last sub-extension element in this prototype of the BDIS-Extension definitionu, is to describe the the context of the BDIS in form of a narrative explanation of the context information of the data. Hence, figure

```
87509        <element id="Extension.extension:bdisGenerationTime">
87510          <path value="Extension.extension"/>
87511          <sliceName value="bdisGenerationTime"/>
87512          <short value="The time when the data was generated"/>
87513          <definition value="A time stamp when the data was measured."/>
87514          <min value="1"/>
87515          <max value="1"/>
87516          <type>
87517            <code value="Extension"/>
87518          </type>
87519        </element>
87520        <element id="Extension.extension:bdisGenerationTime.extension">
87521          <path value="Extension.extension.extension"/>
87522          <max value="0"/>
87523        </element>
87524        <element id="Extension.extension:bdisGenerationTime.url">
87525          <path value="Extension.extension.url"/>
87526          <type>
87527            <code value="uri"/>
87528          </type>
87529          <fixedUri value="bdisGenerationTime"/>
87530        </element>
87531        <element id="Extension.extension:bdisGenerationTime.valueDateTime">
87532          <path value="Extension.extension.valueDateTime"/>
87533          <type>
87534            <code value="dateTime"/>
87535          </type>
87536        </element>
```

Figure 4.14 – Shows the definition of the "bdisGenerationTime" sub-extension element.

4.15 shows this element, which structure definition is similar to these of the "bdisCategory" as shown in lines 87538 to 87559. The appearance of th element is only allowed once, as shown in lines 87543 and 87544 and its fixed uri is defined to be "bdisContext". The data-type of for the context is "string" (lines 87560 to 87565) to add support for narrative description.

Figure 4.16 shows the specification of the rule for using the "http://healthyio.technikum-wien.at/oid/bdis-extension" in the url attribute of the BDIS-Extension (lines 87567 to 87573). Additionally, the last element tag (lines 87574 to 87578) describes that the BDIS-Extension shall not have a "value" in the root level, as a sibling to the sub-extension elements "bdisCategory", "bdisObservationType", "bdisGenerationTime" and "bdisContext". Finally, the last tags show the closing tag of the BDIS-Extension and shows its integration int the "extension-definition.xml" of the HL7 FHIR STU 3.0.1 specification, as this

```
87538          <element id="Extension.extension:bdisContext">
87539            <path value="Extension.extension"/>
87540            <sliceName value="bdisContext"/>
87541            <short value="The context of the BDIS"/>
87542            <definition value="Context describing the BDIS."/>
87543            <min value="1"/>
87544            <max value="1"/>
87545            <type>
87546              <code value="Extension"/>
87547            </type>
87548          </element>
87549          <element id="Extension.extension:bdisContext.extension">
87550            <path value="Extension.extension.extension"/>
87551            <max value="0"/>
87552          </element>
87553          <element id="Extension.extension:bdisContext.url">
87554            <path value="Extension.extension.url"/>
87555            <type>
87556              <code value="uri"/>
87557            </type>
87558            <fixedUri value="bdisContext"/>
87559          </element>
87560          <element id="Extension.extension:bdisContext.valueString">
87561            <path value="Extension.extension.valueString"/>
87562            <type>
87563              <code value="string"/>
87564            </type>
87565          </element>
```

Figure 4.15 – Shows the "bdisContext" sub-extension element from the BDIS-Extension definition.

```
87567          <element id="Extension.url">
87568            <path value="Extension.url"/>
87569            <type>
87570              <code value="uri"/>
87571            </type>
87572            <fixedUri value="http://healthyio.technikum-wien.at/oid/bdis-extension"/>
87573          </element>
87574          <element id="Extension.value[x]">
87575            <path value="Extension.value[x]"/>
87576            <min value="0"/>
87577            <max value="0"/>
87578          </element>
87579        </differential>
87580      </StructureDefinition>
87581      </resource>
87582    </entry>
87583  </Bundle>
```

Figure 4.16 – Shows the definition of the url to be used for referencing the BDIS-Extension.

is indicated by the resource, entry as well as bundle, were the latter is the root-container of containing all official HL7 FHIR extensions registered.

Table 4.5 – Shows the defined code-systems used for the BIDS-Extension category and observation type to add semantic interoperability.

Code-System Name: BIDSCategory		
Code-System Root OID: 1.2.40.0.29.99.1		
Display	**Code**	**Comment**
Health Data	HEALTH	Data from the Health Domain
Transport Data	TRANSPORT	Data from the Transport Domain
Environment Data	ENVIRONMENT	Data from the Environment Domain
...		e.g. further codes according to the domains as described by open data portals
Code-System Name: BIDSObservationType		
Code-System Root OID: 1.2.40.0.29.99.2		
Display	**Code**	**Comment**
Historical Data	HDA	Data from the past e.g. for predictive analysis based on that
Actual Data	ADA	Data measured recently e.g. through Telemonitoring
Forecast Data	FDA	Data that represents the forecast to something e.g. flue epidemics or public transport load factors etc.

Two code-systems, as shown in table 4.5, were specified for the "bdisCategory" and "bdisObservationType". The first code-system BDISCategory (OID: 1.2.40.0.29.99.1) shows the codes used to describe the main category of the data e.g. health related data or transport data etc.. The second code-system BDISObservationType (OID: 1.2.40.0.29.99.2) shows the codes used to describe the nature of the data i.e. if the data is historical data and for example used for making predictive analysis or if it represents already predictive data or recently measured data e.g. real time data in terms of telemonitoring. These code-systems were used in the feasibility studies described in the respective chapter 4.4.

4.3.4 Security & Privacy Considerations

Implementation of the IBD-Profile require the use of state-of-the-art security and privacy measures to fulfill the requirements derived from the different layers of laws, regulations and policies. Although there are differences in the data transmitted through IBD-Profile common measures should be applied independent of the fact that some data is pseudonymous and not sensible or even sensitive anymore. By using RESTful architecture TLS/SSL shall be used to encrypt the data in a proper way and provide certificate based software component authentication. In terms of IHE the ATNA-Profile describes this issues and can be used as a extension to the IBD-Profile.

Furthermore authentication as well as authorization are facts to consider when implementing the IBD-Profile. Therefore, HTTP basic authentication or Hash based Message Authentication (HMAC) can be used for authentication purposes connected to RESTful architecture. However, the profile can seamlessly be combined with OAuth 2.0 to add a more secure context of authentication and additionally supporting requirements derived from authorization perspectives. Generally, the type of token to use depends on the system, but as examples OAuth-token or SAML-tokens may be issued and used for authentication and authorization purposes.

In case the IBD-Profile is used in a highly sensitive environment, logging i.e. audits and audit-trails may be integrated. Therefore, each actor may send proper audits to an audit-repository, when sending or receiving any BDD-0* transaction. Subsequently, these can be filtered and/or forwarded, for example in terms of a reactive security model for automated fraud detection. Apart of this, HL7 FHIR STU 3.0.1 additionally provides "Security Labels", which can be used to add security related meta-data to resources and bundles and therefore adding decision criteria for authorization mechanisms.

Independent of the stated examples, IHE provides already well known and applied security profiles, which can be used with the proposed IBD-Profile. Therefore it

can be recommended to use the "Audit Trail and Node Authentication" (ATNA)-Profile for proper certificate based software component authentication and data encryption as well as Audit generation, as already stated. Another IHE profile, which is currently under development as it is only available in trial implementation status, is the Internet User Authorization (IUA)-Profile. This applies OAuth, SAML and other base standards for authorization in context of using RESTful architecture.

4.4 Domain specific Feasibility Studies

Technical feasibility, to prove the concept i.e. the resulting IBD-Profile, was done by prototypically implementing parts of the proposed scenarios. Common attributes to BDIS were implemented in accordance to the defined BDIS-Extension described in 4.3.3 and the raw data was integrated in HL7 FHIR observation resources. In this examples data integration is only shown for data from the specific domain of explanation. However, combination can be done by adding multiple component-tags to the HL7 FHIR Observation, which purpose is to include the relevant Data itself. Due to reasons of clarity and comprehensibility only the most important BDIS-attributes in the context of the feasibility studies are shown and using xml presentation. However, the complete examples (XML and JSON) including the adapted FHIR STU 3.0.1 definition and the HL7 FHIR Validator can be downloaded from:

http://healthyio.technikum-wien.at/oid/BDIS_FHIR_DevPack_v1.0.zip

Semantic interoperability was provided through defining code-lists and value-sets. An Object Identifier (OID) for each code-list/value-set was assigned. OIDs are internationally used and world wide unique identifiers. Therefore the root (OID) and extension (Code) concept used by HL7 standards, was applied i.e. combining an OID with a code makes it a unique Code. OIDs are issued according to ISO/IEC 9834-1 (IEEE Standards Association, 2018). In Austria OIDs for eHealth related

requirements are issued by the eHealth registration authority by using the OID-Portal Austria (Austrian Ministry of Health and Women, 2018).

4.4.1 Scenario: Healthcare Data Integration

The first feasibility study aimed the integration of wearable activity trackers, measuring heart rate, number of steps per minute as well as sleep behavior, to professional medical IT systems as indicated by the resulting IT architecture shown in figure 4.17.

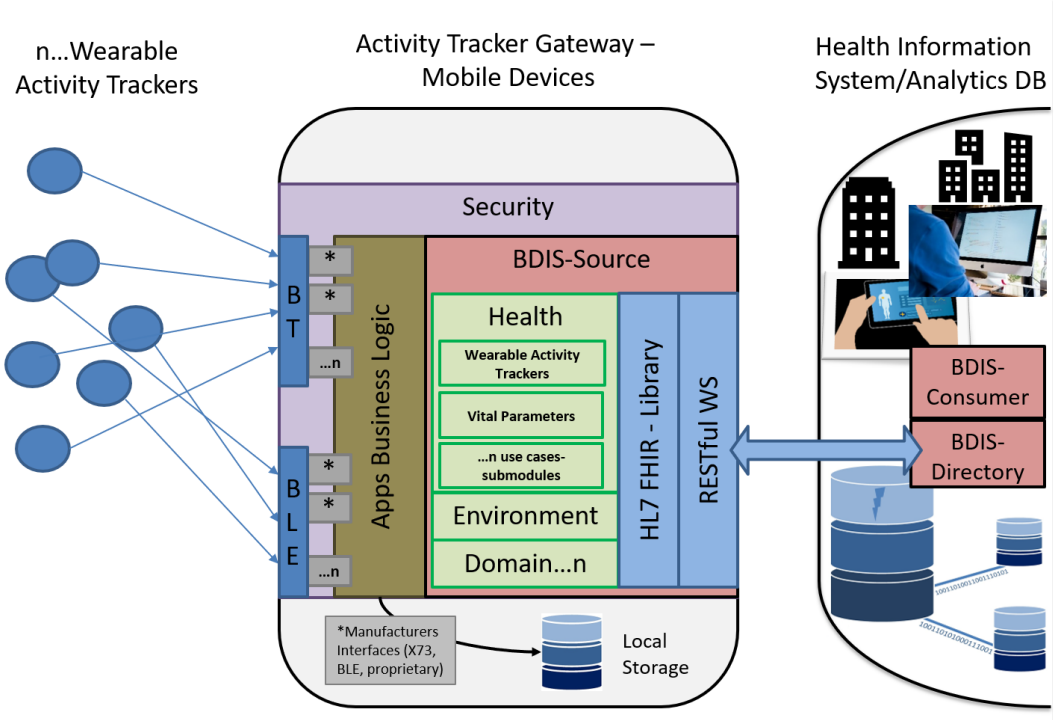


Figure 4.17 – Shows the system architecture of the wearable fitness tracker scenario, based on the work publish at DSAI 2018 (Urbauer et al., 2018).

An arbitrary number of wearable activity trackers may transmit its data using BT or BLE to the Activity Tracker Gateway (ATG). The ATG is an App, running on any mobile device e.g. smartphone, tablet or black-box placed at patients home. As indicated on the left of the ATG by the gray colored boxes, the wearable activity

trackers proprietary interfaces needed to be connected to the Apps business logic, which stored the data in the local storage using SQL-Lite database.

The central part of the ATG is the BDIS-Source in accordance to the conceptualization, which purpose is the proper transformation of the data to the standardized format according to HL7 FHIR STU 3.0.1. In this case, the BIDS-Source shows an implementation example which uses sub-modules (green) for the domain specific mapping i.e. health, environment or transport etc.. Each of this domain specific sub-modules may include a number of mapping routines necessary to fulfill specific scenarios i.e. wearable fitness tracker data integration or vital parameters inside the health domain sub-module. Right to the mapping modules, the light-blue modules indicate a FHIR-Library and a RESTful web-service module, which were used to prepare the HL7 FHIR resources i.e. the observation resource connected with the defined BDIS-Extension. Therefore the HAPI-FHIR open source libraries were used, although support of self-developed libraries was given through the modular approach. Using HAPI-FHIR allowed to use JSON or XML based resource transmission.

Table 4.6 – Shows the defined value-set for the wearable activity tracker scenario to support semantic interoperability, published in the proceedings at the DSAI2018 conference.

Value Set Name: ATC **Value Set Root OID: 1.2.40.0.29.99.3**		
Display	**Code**	**Comment**
Activity Tracking	AT	Observation Type: Activity Tracking
MDC_DIM_STEP_PER_MIN	6752	From IEEE 11073-10441-2013. Describes the "number of steps per minute measured", when observing a person with a wearable activity tracker for activity
Number of seconds of sleep per minute measured	NSSPM	Describes the "number of seconds of sleep per minute measured", when observing a person's sleep behavior
Heart Rate	8867-4	Heart Rate as defined in LOINC

On the right side of figure 4.17, the receiving instance in form of an integrated

BDIS-Directory/Consumer is shown, which generally can be part of a HIS. In this feasibility study the HAPI-FHIR server was used to test the proper data exchange and to test, through its integrated search and retrieve features, the BDIS-Directory/Consumer role. An important fact in this context, was the support for semantical interoperability. Therefore, terminologies like LOINC or UCUM were used in accordance to the HL7 FHIR specification to add semantics to the observation resources transmitted. Nevertheless, a value-set called Activity Tracking Codes (ATC) was defined as there was lack of an appropriate code to properly describe the "number of seconds of sleep per minute". Furthermore, a code for describing the type of observation was added to the value-set, which is shown in table 4.6.

```
<Observation xmlns="http://hl7.org/fhir">
    <contained>
     <Organization xmlns="http://hl7.org/fhir">
      <id value="1"/>
      <name value="ATG"/>
     </Organization>
    </contained>
    <contained>
     <Patient xmlns="http://hl7.org/fhir">
      <id value="2"/>
      <identifier>
        <system value="urn:oid:1.2.40.0.29.99.0"/>
        <value value="1234"/>
      </identifier>
      <name>
       <use value="official"/>
       <family value="Urbauer"/>
       <given value="Philipp"/>
      </name>
      <gender value="male"/>
      <birthDate value="1984-01-26"/>
     </Patient>
    </contained>
```

Figure 4.18 – Shows the contained-elements, describing external resources in-line inside the wearable fitness tracker HL7 FHIR observation resource.

The figures 4.18 to 4.22 represent an example HL7 FHIR observation resource according to the IBD-Profiles definition, for the wearable fitness tracker scenario. Starting with figure 4.18, this shows two contained-elements used for inline-definition of the performing organization and the patient measuring his/her biometric indicators with a fitness tracker. The performing organization in this case is the

ATG as shown in line 5. The organization- and the patient-element include id-elements in line 4 and 10, which is used subsequently for reference purposes in figure 4.20. Furthermore, the patient element includes a complex-type identifier-element, which shows the PID for use inside a HIS, PHR or EHR (see lines 11-14). Finally, the patients demographic data, including name, gender and birth-date information, is carried within the observation resource, which is shown in lines 15-21.

```
24      <extension url="http://healthyio.technikum-wien.at/oid/bdis-extension">
25      <extension url="bdisCategory">
26        <valueCodeableConcept>
27          <coding>
28            <system value="urn:oid:1.2.40.0.29.99.1"/>
29            <code value="HEALTH"/>
30            <display value="Health Data"/>
31          </coding>
32        </valueCodeableConcept>
33      </extension>
34      <extension url="bdisObservationType">
35        <valueCodeableConcept>
36          <coding>
37            <system value="urn:oid:1.2.40.0.29.99.2"/>
38            <code value="ADA"/>
39            <display value="Actual Data"/>
40          </coding>
41        </valueCodeableConcept>
42      </extension>
43      <extension url="bdisGenerationTime">
44        <valueDateTime value="2018-03-07T09:32:02+01:00"/>
45      </extension>
46      <extension url="bdisContext">
47        <valueString value="Continous fitness tracker data measurements."/>
48      </extension>
49    </extension>
```

Figure 4.19 – Shows the implementation of the defined BDIS-Extension used in the wearable fitness tracker data observation resource.

Figure 4.19 shows the BDIS-Extension, starting with the url reference to the public definition file "bdis-extension". The first sub-extension element, shown in lines 25-33, describe the category of the BDIS by using a codeable-concept element. Hence, it includes the proper code from the BDISCategory code-system as described earlier and defined in table 4.5. The second sub-extension (lines 34-42) shows the BDIS observation type, which is similarly to the category described with a codeable-concept element. The proper code is taken from the BDISObservationType code-system shown in table 4.5. In this case it is "Actual Data" as this scenario focuses

on telemonitoring. The third sub-extension, shown in lines 43-45, is the time-stamp component i.e. the BDIS generation time, which in this case shows the 7th of March 2018 at 09:32 and 2 seconds, GMT+1. Finally, lines 46 to 48 show the BIDS context-element describing, that the data in this observation is concerned with continuous fitness tracker measurements.

```
50      <status value="final"/>
51      <code>
52       <coding>
53        <system value="urn:oid:1.2.40.0.29.99.3"/>
54         <code value="AT"/>
55         <display value="Activity Tracking"/>
56       </coding>
57      </code>
58      <subject>
59        <reference value="#2"/>
60      </subject>
61      <effectiveDateTime value="2017-03-14T10:11:47+01:00"/>
62      <performer>
63       <reference value="#1"/>
64      </performer>
65      <component>
66       <code>
67        <coding>
68         <system value="urn:oid:1.2.40.0.29.99.3"/>
69         <code value="8867-4"/>
70         <display value="Heart rate"/>
71        </coding>
72       </code>
73       <valueSampledData>
74        <origin>
75         <value value="0"/>
76         <unit value="beats/minute"/>
77         <system value="http://unitsofmeasure.org"/>
78         <code value="{beats}/min"/>
79        </origin>
80        <period value="60000"/>
81        <lowerLimit value="0"/>
82        <upperLimit value="250"/>
83        <dimensions value="1"/>
84        <data value="56 59 61 120 130 144 123 50 56"/>
85       </valueSampledData>
86      </component>
```

Figure 4.20 – Shows part three of the wearable fitness tracker HL7 FHIR observation resource, describing general meta-information and the first component including the heart rate data.

The next part of the HL7 FHIR observation resource, as described in figure 4.20, shows the general root information elements of the resource and additionally the first component-element, which includes data values and its description. Hence, line

50 describes the status of the observation to be final i.e. closed data collection and the code element (see lines 51-57) describes the purpose of the observation to be an "Activity Tracking" session. In this case, the code "AT" from the ATC value-set is used (see table 4.6). Lines 58-60 as well as 62-64 are used to reference to the performing organization and patient connected to this observation. The definition was done in-line as described earlier, but can also be done externally. This means by using an HL7 FHIR Organization resource, which exactly the same structure but has been previously stored. Then the resource stored at the server is linked with its URI in this HL7 FHIR Observation resource. The effective data shown in line 61, indicates that the measurements were taken nearly a year before the BIDS was composed.

The component-element shows the activity related tracking data i.e. the heart rate information measured. Therefore, as shown in line 66-72, the LOINC code 8867-4 is used to properly describe the data by using the a codeable-concept format. The valueSampledData-element is used for structuring the data, as it is shown in lines 48-60. This was used to allow transmission of multiple sample values in one observation resource, at any chosen point in time e.g. automated transmission. Lines 74-79 show the origin-element used to describe the the zero-value as well as its physical unit derived from UCUM code-system. Subsequently to line 79, the period-element defines the time span between two samples to be 60000 milliseconds and the lowerLimit-element as well as the upperLimit-element the minimal/maximum possible values taken into account for heart rate measurements. Finally, the data-element contains the heart rate values separated by whitespace-character.

The next component-element, as shown in figure 4.21, was used to depict the activity data measured i.e. the number of steps per minute. Supporting semantical interoperability, lead to the necessity of including a proper code from the IEEE 11073-10101 (Point-of-care medical device communication – Part 10101: Nomenclature) in the ATC code-system. The application is shown in lines 88-94, followed by the valueSampledData-element containing the data values. The approach is similar to the component used to store heart rate measurements. Thus,

the data values are stored in the data-element (line 106) and the period, lower/upper limit and dimensions are shown in lines 102-105.

```
87    <component>
88     <code>
89      <coding>
90       <system value="urn:oid:1.2.40.0.29.99.3"/>
91       <code value="6752"/>
92       <display value="MDC_DIM_STEP_PER_MIN"/>
93      </coding>
94     </code>
95     <valueSampledData>
96      <origin>
97       <value value="0"/>
98       <unit value="number/minute"/>
99       <system value="http://unitsofmeasure.org"/>
100      <code value="{#}/min"/>
101     </origin>
102     <period value="60000"/>
103     <lowerLimit value="0"/>
104     <upperLimit value="300"/>
105     <dimensions value="1"/>
106     <data value="7 2 3 2 45 49 39 22 30"/>
107    </valueSampledData>
108   </component>
109   <component>
```

Figure 4.21 – Shows the second component of the wearable fitness tracker HL7 FHIR observation resource, applied to include the "steps per minutes" values.

The last part of the applied HL7 FHIR observation resource is shown in figure 4.22. This component-element contains the data relevant to the sleep-activity related measurement data. A separated code "Number of seconds of sleep per minuted measured" (NSSPM), was defined in the ATC value-set as shown in the coding-element (line 110-116). This was done as no applicable code was found in an Internet based research as well as by usage of the terminology-server of the Austrian healthcare system (Austrian Ministry of Health and Woman, 2018). The physical unit of the data, was defined to be number/minute. Similar to the heart rate and the number of steps per minute data values, a valueSampledData-element was used for defining the period, ranges and dimensions (line 124-127) of the data values itself (line 128). The values and its ranges were selected to allow deep-sleep and weak-sleep analysis done in the way it was investigated during the market analysis regarding the wearable fitness trackers.

```
109     <component>
110      <code>
111       <coding>
112        <system value="urn:oid:1.2.40.0.29.99.3"/>
113        <code value="NSSPM"/>
114        <display value="Number of seconds of sleep per minute measured"/>
115       </coding>
116      </code>
117      <valueSampledData>
118       <origin>
119        <value value="0"/>
120        <unit value="number/minute"/>
121        <system value="http://unitsofmeasure.org"/>
122        <code value="{#}/min"/>
123       </origin>
124       <period value="60000"/>
125       <lowerLimit value="0"/>
126       <upperLimit value="60"/>
127       <dimensions value="1"/>
128       <data value="0 0 0 1 2 49 55 59 58"/>
129      </valueSampledData>
130     </component>
131 </Observation>
```

Figure 4.22 – Shows the final component, containing the value for "number of seconds of sleep per minute", for sleep analysis.

The technical validation was done by using the official HL7 FHIR Validator together with the HL7 FHIR STU 3.0.1 specification. However, to support proper validation for the defined extension, the specification was previously extended by the definition of the BIDS-Extension described in 4.3.3. Subsequently the validation was done with the command-line tool. The prove of successful validation against the extended HL7 FHIR STU 3.0.1 specification is shown in 4.23.

```
C:\FHIR_Validator>java -jar org.hl7.fhir.validator.jar PhD_FHIR_Resource_Trackers.xml -defn definitions.xml.zip
  .. load FHIR from definitions.xml.zip
  .. connect to tx server @ http://tx.fhir.org/r3
    (vnull-null)
  .. validate
Success...validating PhD_FHIR_Resource_Trackers.xml:  error:0 warn:0 info:0
```

Figure 4.23 – Shows the successful result of the BDIS conform wearable fitness tracker HL7 FHIR observation resource conformance validation by application of the HL7 Validator including the extended HL7 FHIR STU 3.0.1 specification.

4.4.2 Scenario: Environmental Data Integration

In order to test the applicability of the IBD-Profile with non-medical data, the second feasibility study focused on a scenario for the combination of pollen exposure data (severity levels of pollen exposure) as well as vital parameters i.e. systolic-, diastolic and mean atrial pressure as well as oxygen saturation and pulse. The resulting IT architecture, shown in 4.24, consists of two components i.e. telemonitoring component and the infrastructure component.

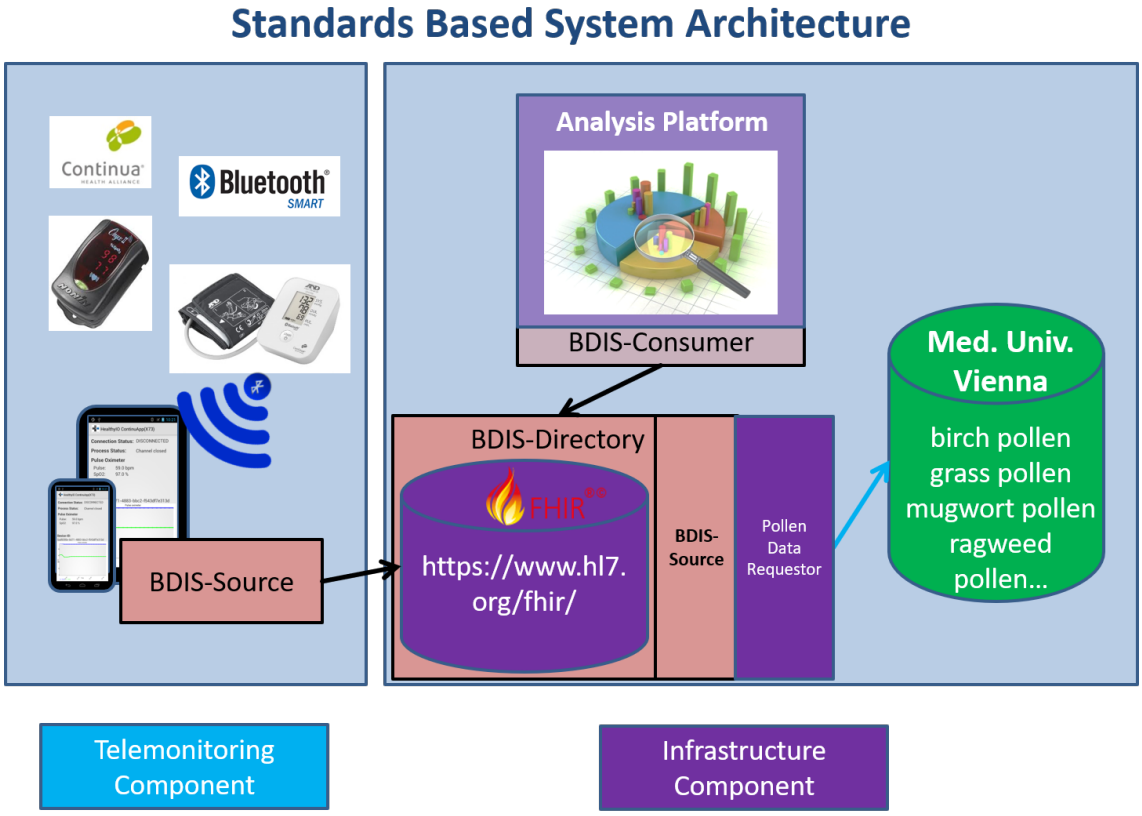


Figure 4.24 – Overview about the system architecture of the pollen exposure data integration scenario based on the already published version in (Urbauer et al., 2017).

Telemonitoring Component: In order to continuously measure oxygen saturation and pulse, the PCHA certified Onyx Vantage 9590 Finger Pulse Oximeter from the company Nonin was used. The data transmission between the device and an Android

App, was done using Bluetooth and the Health Device Profile (HDP) as well as Secure Simple Pairing (SSP) to pair the smartphone/tablet and the device before communication took place. As an application layer protocol, the IEEE 11073-20601 Optimized Exchange Protocol using the specified Medical Device Encoding Rules (MDER) according to the Continua guidelines, was used. Hence, the device acted as an X73-Agent which is the generic term for a Personal Health Device (PHD) according to the guidelines of the PCHA. The measured vital parameters were transmitted as described to the Android App running on a OnePlus 3 smartphone, representing the X73-Manager according the guidelines. For measurement and transmission of blood pressure data i.e. systolic-, diastolic- and mean arterial pressure as well as pulse, the A&D Medical Blood Pressure Monitor UA-651ble (CHA Certified) was used. In this case the Bluetooth Low Energy (BLE) technology was used as required by the Continua guidelines. Hence, the App included an X73 & BLE mapping library, which is capable of decoding as well as extracting the measured data from the data formats. The BDIS-Source transforms the measured data to the standardized format according to extended HL7 FHIR STU 3.0.1. Therefore, an observation resource connected with the BDIS-Extension was generated by using the HAPI-FHIR library for Android.

Infrastructure Component: In this case, the focus lied on the infrastructure components of the system. The BDIS-Directory was depicted trough the public open source HAPI-FHIR server. Hence, measured data from the two PHD's was stored in the BDIS-Directory. The server supports HL7 FHIR STU 3.0.1 and therefore provides the necessary requirements for the BIDS-Consumer to use the BDD-02 transaction to query data from the server. In this case only the queries were conducted, but no analysis platform was developed. Furthermore, figure 4.24 shows that for the pollen exposure data integration process, the BIDS-Directory is grouped with an BDIS-Source. This was necessary, as the "PollenDataRequester"-component (PDR) purpose was to periodically request pollen exposure data, but the mapping process to the standardized HL7 FHIR format was subsequently done by the BDIS-Source. The received pollen data from the MUV database indicated information about pollen exposure on a daily basis. Hence, the values describe

severity levels of exposure for different particles like birch-, grass-, mugwort- and/or ragweed-pollen. The value of the severity level is a between 0.0-4.0 (dimensionless). The used service provided data describing pollen exposure for up to three days i.e. today, tomorrow and the day after tomorrow. For receiving proper data, the location information was sent to the MUV service (based on RESTful architecture) to retrieve regionally correct information. After requesting data from the pollen exposure database of the MUV, the data was then mapped by the BDIS-Source to the BDIS conform HL7 FHIR format i.e. HL7 FHIR observation resource with enhanced BDIS-Extension using the HAPI-FHIR library as shown in figures 4.25 and 4.26.

Mapping the data to the BDIS conform HL7 FHIR observation resource, required the definition of codes as shown in table 4.7.

Table 4.7 – Shows the defined code-system for the pollen exposure integration scenario to support semantic interoperability.

Code-System Name: PEI		
Code-System Rood OID: 1.2.40.0.29.99.4		
Display	**Code**	**Comment**
Pollen Forecast Data	PollenForecast	Indicates a pollen exposure forecast
BETULA	BETU	Birch exposure
ALNUS	ALNU	Alder exposure
CORYLUS	CORY	Hazel exposure
POACEAE	POAC	Grasses exposure
...		

The
code-system "Pollen Exposure Information" (PEI) with the OID 1.2.40.0.29.99.4 was defined to be used for pollen exposure data integration. On the one hand, the PEI code-system includes a code necessary to identify the observation resource itself as a container for pollen forecast data (PollenForecast). On the other hand, for each type of particle of pollen exposure a code was defined e.g for birch pollen particles

the code BETU was used in the observation resource.

Figure 4.25 and 4.26 shows the resulting HL7 FHIR observation resource used in this feasibility study to prove standards based transmission of pollen exposure data. In figure 4.25, two inline-defined contained-elements were placed followed by the BDIS-Extension.

```
1  <Observation xmlns="http://hl7.org/fhir">
2    <contained>
3      <Location xmlns="http://hl7.org/fhir">
4        <id value="1"/>
5        <position>
6          <longitude value="48.239229"/>
7          <latitude value="16.378234"/>
8        </position>
9      </Location>
10   </contained>
11   <contained>
12     <Organization xmlns="http://hl7.org/fhir">
13       <id value="2"/>
14       <name value="Data Provider"/>
15     </Organization>
16   </contained>
17   <extension url="http://healthyio.technikum-wien.at/oid/bdis-extension">
18     <extension url="bdisCategory">
19       <valueCodeableConcept>
20         <coding>
21           <system value="urn:oid:1.2.40.0.29.99.1"/>
22           <code value="ENVIRONMENT"/>
23           <display value="Environment Data"/>
24         </coding>
25       </valueCodeableConcept>
26     </extension>
27     <extension url="bdisObservationType">
28       <valueCodeableConcept>
29         <coding>
30           <system value="urn:oid:1.2.40.0.29.99.2"/>
31           <code value="FDA"/>
32           <display value="Forecast Data"/>
33         </coding>
34       </valueCodeableConcept>
35     </extension>
36     <extension url="bdisGenerationTime">
37       <valueDateTime value="2017-03-14T10:57:34+01:00"/>
38     </extension>
39     <extension url="bdisContext">
40       <valueString value="Combination of pollen exposure- and vital parameter data."/>
41     </extension>
42   </extension>
```

Figure 4.25 – Shows the first part of the HL7 FHIR observation resource, including the BDIS-Extension, used for pollen exposure data integration.

The first contained-element (line 2-11) is used to describe the location for which the pollen exposure data was requested. Thus, longitude and latitude values (lines 6-7) are part of the position-element in the location resource in-line-definition.

The id-element in line 4, is used to reference the in-line-definition of the location resource later in the BDIS observation's meta-data (see figure 4.26). The second contained-element is used to add information about the institution providing the pollen exposure data. Similarly, this element defines an id-element as it is referenced later in the meta-data.

Line 17-42 shows the BDIS-Extension as it was used in this scenario. Hence, the url references to the public definition file "bdis-extension". The first sub-extension element, shown in lines 18-26, describe the category of the BDIS by using a codeable-concept element. Therefore, it includes the proper code from the BDISCategory code-system as described earlier and defined in table 4.5. The second sub-extension (lines 27-35) shows the BDIS observation type, which is similarly to the category, described with a codeable-concept element. The proper code was taken from the BDISObservationType code-system shown in 4.5. In this case the code is FDA as forecast data was used i.e. pollen exposure of today, tomorrow and the day after tomorrow. The third sub-extension, shown in lines 36-38, is the time-stamp component i.e. the BDIS generation time, which in this case it shows the 14th of March 2017 at 10:57 and 34 seconds, GMT+1. Finally, lines 39 to 41 show the BIDS-context-element describing the context of the data measurements i.e. pollen exposure data shall be combined with vital parameters.

Figure 4.26 shows the meta-data of the BDIS observation resource, including the status (line 43) of the data to be final as well as the code-element (line 45-49) describing the observation type in more detail. Hence, the "PollenForecast" code from the PEI code-system (see 4.7) was used. Subsequently, the subject-element (line 51-54) was used to reference to the in-line-definition of the location recourse and the performer-element (line 59-61) is used to reference the in-line-definition of the performing organization. Line 55-58 shows the time period for the pollen exposure data i.e. three days forecast.

After the meta-data the component-elements were placed. In this example there is only one due to better clarity. However, the general structure shown here is used for each type of pollen exposure particle. First, line 63 to 69 contains the

```
43    <status value="final"/>
44    <code>
45      <coding>
46        <system value="urn:oid:1.2.40.0.29.99.4"/>
47        <code value="PollenForecast"/>
48        <display value="Pollen Forecast Data"/>
49      </coding>
50    </code>
51    <subject>
52      <reference value="#1"/>
53      <display value="Region of Pollen Forecast"/>
54    </subject>
55    <effectivePeriod>
56      <start value="2017-03-14"/>
57      <end value="2017-03-16"/>
58    </effectivePeriod>
59    <performer>
60      <reference value="#2"/>
61    </performer>
62    <component>
63      <code>
64        <coding>
65          <system value="urn:oid:1.2.40.0.29.99.4"/>
66          <code value="BETU"/>
67          <display value="BETULA"/>
68        </coding>
69      </code>
70      <valueSampledData>
71        <origin>
72          <value value="0"/>
73        </origin>
74        <period value="86400000"/>
75        <lowerLimit value="0"/>
76        <upperLimit value="4"/>
77        <dimensions value="1"/>
78        <data value="2 2 3"/>
79      </valueSampledData>
80    </component>
81  </Observation>
```

Figure 4.26 – Shows the second part of the HL7 FHIR observation resource, including the component elements holding the meta-data and pollen exposure data.

code for describing the actual particle type, which in this case was birch pollen as the BETU code from the PEI code-system (see 4.7) indicates. Subsequently, the valueSampledData-element (lines 70-79) contains the severity value for pollen exposure and its properties. Thus, the origin value is 0 and the range of the severity value is 0-4 in the dimensions 1. The period defines the timely difference between the values to be 24 hours, were the physical unit is milliseconds. Finally, the data element contains the values for the severity level of birch pollen exposure for the three days.

The technical validation was done by using the official HL7 FHIR Validator together with the STU 3.0.1 specification. However, to support proper validation for the defined extension the specification was previously extended by the integration of the BIDS-Extension as described in 4.3.3. Subsequently the validation was done with the command-line tool. The prove of successful validation against the extended HL7 FHIR STU 3.0.1 specification is shown in figure 4.27.

```
C:\FHIR_Validator>java -jar org.hl7.fhir.validator.jar PhD_FHIR_Resource_Environment.xml -defn definitions.xml.zip
  .. load FHIR from definitions.xml.zip
  .. connect to tx server @ http://tx.fhir.org/r3
    (vnull-null)
  .. validate
Success...validating PhD_FHIR_Resource_Environment.xml:  error:0 warn:0 info:0
```

Figure 4.27 – Shows the successful result of the BDIS conform pollen exposure HL7 FHIR observation resource conformance validation by application of the HL7 Validator including the extended HL7 FHIR STU 3.0.1 specification.

4.4.3 Scenario: Transport Data Integration

The final feasibility study focused on the integration of public transport data. Therefore, figure 4.28 shows the designed architecture of the system. The green unit describes the mobile device, in this case specifically the smartphone application. On the left, the light blue units indicate the PHDs acting as data sources for providing vital parameters. On the right side, the light gray unit indicates the possibility of integrating different Transportation Data Sources (TDS). On the bottom of the architecture, the BDIS-Directory for long-term data storage is shown and an example of a consuming service in form of an analysis system including a BDIS-Consumer is described.

The light blue unit indicates a generic integration of PHDs. However, in the pilot a blood pressure device as well as a pulse oximeter was used, similar to the environment scenario. Hence, the PCHA certified Onyx Vantage 9590 Finger Pulse Oximeter from the company Nonin was used to measure oxygen saturation and pulse. The data transmission between the device and an Android App, was done using Bluetooth and the Health Device Profile (HDP) as well as Secure Simple

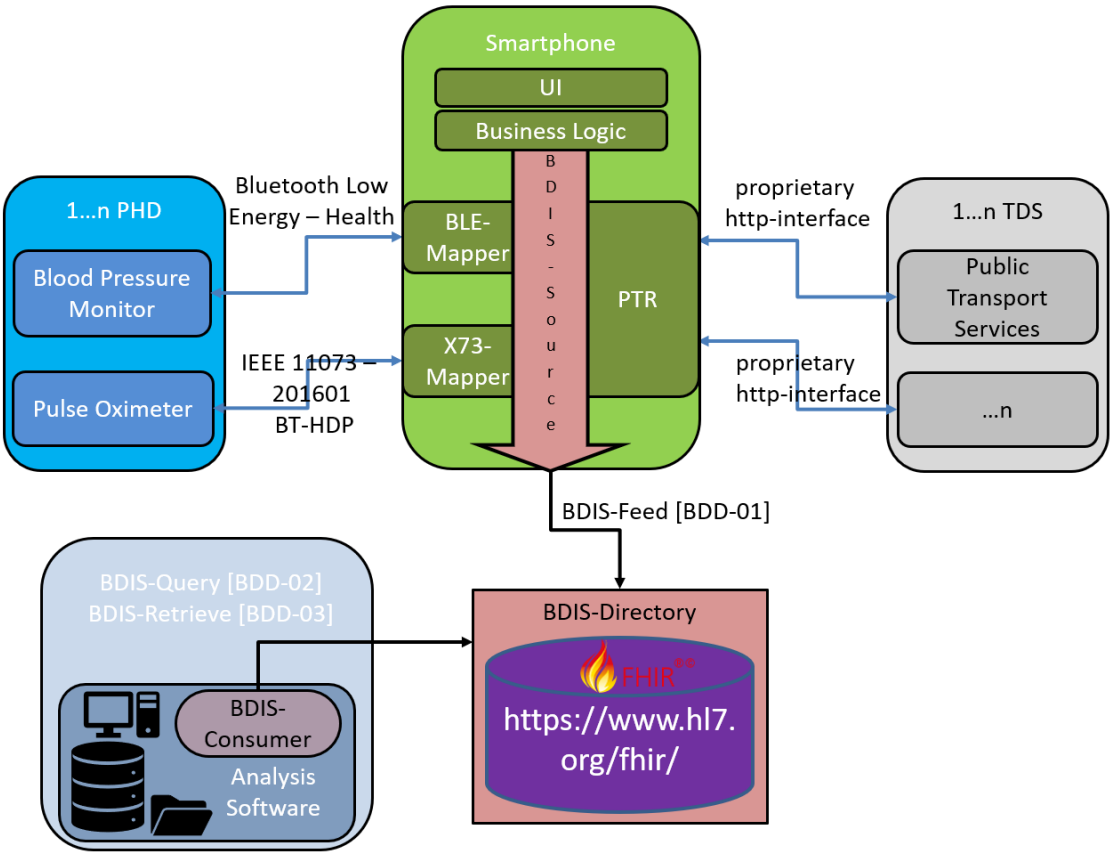


Figure 4.28 – Shows the system architecture of the public transport data integration scenario.

Pairing (SSP) to pair the smartphone/tablet and the device before communication took place. As an application layer protocol, the IEEE 11073-20601 Optimized Exchange Protocol using the specified Medical Device Encoding Rules (MDER) according to the Continua guidelines, was used. The A&D Medical Blood Pressure Monitor UA-651ble (CHA Certified) was used to measure systolic, diastolic and mean arterial pressure. In this case the Bluetooth Low Energy (BLE) technology was used for transmission in accordance to the Continua guidelines. The Application included an X73 & BLE mapping library, which is capable of decoding as well as extracting the measured data from the data formats. The BDIS-Source transformed the measured data to the standardized format according to HL7 FHIR STU 3.0.1. Therefore, an observation resource with specified BDIS-Extension was generated by using the HAPI-FHIR library for Android. In this case the initiation was done by

the patient, when performing measurements.

Taking the gray box into account, the indicated Public Transport Data Requester (PTR) was actively requesting data from the proprietary interfaces of the VPTS or other data providers via using HTTP requests. Once done, the data was forwarded to the BDIS-Source, which mapped the data to the HL7 FHIR observation resource with BDIS-Extension. On the bottom of the architecture, the IT infrastructure components for data storage and analysis are shown. Hence, the centralized BDIS-Directory, which was emulated using the public open-source HAPI FHIR server to store the BDIS conform HL7 FHIR observation resource via BDIS-Feed. As indicated by the gray-blue component, an analytic software including a BDIS-Consumer could be integrated for later analysis of the data. However, this was not done in this study.

Table 4.8 – Shows the defined code-systems for the public transport data integration scenario to support semantic interoperability.

Code-System Name: PT		
Code-System Root OID: 1.2.40.0.29.99.5		
Display	**Code**	**Comment**
Public Transport Departures	PTD	Departures of public transport line
Ultra-Low-Floor	ULF	Disability-friendly high degree
Low-Floor	LFL	Disability-friendly low degree
High-Floor	HFL	Not disability-friendly
Code-System Name: VPTS		
Code-System Root OID: 1.2.40.0.29.99.6		
Display	**Code**	**Comment**
37A Direction Engerthstrasse —Traisengasse	37A_ET	Identifies line and direction
U6 Direction Siebenhirten	U6_SH	Identifies line and direction
2 Direction Dornbach	2_DB	Identifies line and direction
31 Direction Schottenring	31_SR	Identifies line and direction
...		

Integrating public transport data in the BDIS conform HL7 FHIR observation

resource format, required the definition of code-system and codes to foster semantic interoperability. Hence, on the top of table 4.8 the "Public Transport" (PT) code-system identified with an OID according to the recommendation of ITU-T X.660 — ISO/IEC 9834-1 (IEEE Standards Association, 2018), is shown. This code-system defines a code for identifying data regarding departures of public transports as well as the types of public transport vehicles. This was important to indicated for disabled people to know which line is sufficient for their purposes. On the bottom of table 4.8, the "Vienna Public Transport Services" (VPTS) code-system was defined. The purpose of this list was to clearly identify the public transport lines and its moving detections. Taking the 37A_ET as an example, this indicates the line 37A going in the direction of Engerthstrasse/Traisengasse. As a complete list of lines for the VPTS would go beyond the scope of this work, it is abbreviated to a rational number.

Figure 4.29 to 4.32 show the resulting HL7 FHIR observation resource used in this feasibility study to prove standards based transmission of data in the context of public transport. Starting with figure 4.29 the first contained-element (line 2-7) was used for in-line-definition of the organization providing the transport related data. This element includes a description and an id-element, necessary for referencing it in the observation resource meta-data. The second contained-element, shown in lines 8-24 describes the users/patients location. That includes the address information of his/her location, which is on the one hand described in narrative-elements (lines 11-18) and on the other hand by using a position-element for the geo-information i.e. longitude & latitude (see lines 19-22). Similarly to the organization resource, the in-line-definition requires the specification of an id-element for proper referencing in the meta-data of the observation resource.

Figure 4.30 shows the BDIS-Extension used for indicating a transport related data observation. The first sub-extension element, shown in lines 26-34, describe the category of the BDIS by using a codeable-concept element. Therefore, it includes the proper code from the BDISCategory code-system as described earlier and defined in table 4.5. The second sub-extension (lines 35-43) shows the BDIS observation type which is, similarly to the category, described with a codeable-concept element. The

```
<Observation xmlns="http://hl7.org/fhir">
    <contained>
     <Organization xmlns="http://hl7.org/fhir">
      <id value="1"/>
      <name value="VPTS"/>
     </Organization>
    </contained>
    <contained>
     <Location xmlns="http://hl7.org/fhir">
      <id value="2"/>
      <address>
        <use value="home" />
        <text value="Hoechstaedtplatz 6 Vienna AUT" />
        <line value="Hoechstaedtplatz 6" />
        <city value="Vienna" />
        <postalCode value="1200" />
        <country value="AUT" />
      </address>
      <position>
        <longitude value="48.239210"/>
        <latitude value="16.378198"/>
      </position>
     </Location>
    </contained>
```

Figure 4.29 – Shows the in-line-defined resources in the public transport HL7 FHIR observation resource.

```
    <extension url="http://healthyio.technikum-wien.at/oid/bdis-extension">
    <extension url="bdisCategory">
      <valueCodeableConcept>
        <coding>
          <system value="urn:oid:1.2.40.0.29.99.1"/>
          <code value="TRANSPORT"/>
          <display value="Transport Data"/>
        </coding>
      </valueCodeableConcept>
    </extension>
    <extension url="bdisObservationType">
      <valueCodeableConcept>
        <coding>
          <system value="urn:oid:1.2.40.0.29.99.2"/>
          <code value="FDA"/>
          <display value="Forecast Data"/>
        </coding>
      </valueCodeableConcept>
    </extension>
    <extension url="bdisGenerationTime">
      <valueDateTime value="2018-04-02T18:52:06+01:00"/>
    </extension>
    <extension url="bdisContext">
      <valueString value="Public transport data to be combined with vital parameters."/>
    </extension>
  </extension>
```

Figure 4.30 – Shows the BDIS-Extension used in the public transport HL7 FHIR observation resource.

proper code was taken from the BDISObservationType code-system shown in 4.5. In this case the code is FDA as forecast data was provided to the person using the Android App for efficiently finding route guidance. The third sub-extension, shown in lines 44-46, is the time-stamp component i.e. the BDIS generation time, which in this case shows the 2nd of april 2018 at 18:52:06 GMT+1. Finally, lines 47 to 49 show the BIDS context-element describing the context of the data measurements i.e. public transport data to be combined with vital parameters.

Figure 4.31 shows the meta-data of this BDIS conform HL7 FHIR observation resource as well as the first component-element containing public transport data.

The code-element, shown in lines 52-58 contains the type of the observation. Thus, the code PTD from the PT code-system indicates that the data is about public transport departures. Subsequently, the subject-element and the performer-element reference the in-line-defined location- and organization resources. The effectiveDateTime-element (line 63) holds the relevant time information when the data was requested from the VPTS. Lines 67 to 87 show the first component-element of this observation resource, containing the public transport departure data. Hence, the code-element (line 68-74) includes a code defined in the VPTS code-system for a specific line and its movement direction. In this case it was the line 37A going in the direction Engerthstrasse/Traisengasse. The valueQuantity-element, as shown in line 75-79, includes the time in minutes (using the unit-code "min" from UCUM code-system) when the 37A was passing by the station near the user/patients request location. Subsequently, the lines 80-86 show an interpretation-element for describing the type of public transport regarding its support for disabled persons. Therefore it includes a coding-element using one of the defined codes (ULF, LFL or HFL) from the PT code-system, in this case LFL i.e. indicating a low-floor unit.

Finally, figure 4.32 shows a second component-element indicating again the line 37A in the direction of Engerthstrasse/Traisengasse (lines 89-95), but this time passing by thirteen minutes form the request-time stored in the effectiveDatetime-element. Furthermore, this time it was a high-floor unit i.e. providing less support for disabled

```
51      <status value="final"/>
52      <code>
53       <coding>
54        <system value="urn:oid:1.2.40.0.29.99.5"/>
55         <code value="PTD"/>
56         <display value="Public Transport Departures"/>
57       </coding>
58      </code>
59      <subject>
60        <reference value="#2"/>
61        <display value="Place of Origin"/>
62      </subject>
63      <effectiveDateTime value="2018-01-03T09:23:58+01:00"/>
64      <performer>
65       <reference value="#1"/>
66      </performer>
67      <component>
68       <code>
69        <coding>
70         <system value="urn:oid:1.2.40.0.29.99.6"/>
71         <code value="37A_ET"/>
72         <display value="37A Direction Engerthstrasse|Traisengasse" />
73        </coding>
74       </code>
75       <valueQuantity>
76        <value value="2"/>
77        <system value="http://unitsofmeasure.org"/>
78        <code value="min"/>
79       </valueQuantity>
80       <interpretation>
81        <coding>
82         <system value="urn:oid:1.2.40.0.29.99.5"/>
83         <code value="LFL"/>
84         <display value="Low-Floor"/>
85        </coding>
86       </interpretation>
87      </component>
```

Figure 4.31 – Shows meta-data and the first component-element, holding data of a specific public bus with the line number 37A.

persons.

The technical validation was done by using the official HL7 FHIR Validator together with the STU 3.0.1 specification. However, to support proper validation for the defined extension the specification was previously extended by the integration of the BIDS-Extension described in 4.3.3. Subsequently the validation was done with the command-line tool. The prove of successful validation against the extended HL7 FHIR STU 3.0.1 specification is shown in figure 4.33.

```
88      <component>
89       <code>
90        <coding>
91         <system value="urn:oid:1.2.40.0.29.99.6"/>
92         <code value="37A_ET"/>
93         <display value="37A Direction Engerthstrasse|Traisengasse" />
94        </coding>
95       </code>
96       <valueQuantity>
97        <value value="13"/>
98        <system value="http://unitsofmeasure.org"/>
99        <code value="min"/>
100      </valueQuantity>
101      <interpretation>
102       <coding>
103        <system value="urn:oid:1.2.40.0.29.99.5"/>
104        <code value="HFL"/>
105        <display value="High-Floor"/>
106       </coding>
107      </interpretation>
108     </component>
109 </Observation>
```

Figure 4.32 – Shows the second component of the public transport HL7 FHIR observation resource, used for transmitting the "steps per minutes" values.

```
C:\FHIR_Validator>java -jar org.hl7.fhir.validator.jar PhD_FHIR_Resource_Transport.xml -defn definitions.xml.zip
  .. load FHIR from definitions.xml.zip
  .. connect to tx server @ http://tx.fhir.org/r3
    (vnull-null)
  .. validate
Success...validating PhD_FHIR_Resource_Transport.xml:  error:0 warn:2 info:0
  Warning @ Observation.component[1].interpretation (line 80, col22) : None of the codes provided are in the value s
et http://hl7.org/fhir/ValueSet/observation-interpretation (http://hl7.org/fhir/ValueSet/observation-interpretation,
 and a code should come from this value set unless it has no suitable code) (codes = urn:oid:1.2.40.0.29.99.5#LFL)
  Warning @ Observation.component[2].interpretation (line 101, col22) : None of the codes provided are in the value
set http://hl7.org/fhir/ValueSet/observation-interpretation (http://hl7.org/fhir/ValueSet/observation-interpretation
, and a code should come from this value set unless it has no suitable code) (codes = urn:oid:1.2.40.0.29.99.5#HFL)
```

Figure 4.33 – Shows the successful result of the BDIS conform public transport HL7 FHIR observation resource conformance validation, by application of the HL7 Validator including the extended HL7 FHIR STU 3.0.1 specification.

4.5 Experts Review

The experts review was performed during the European Connect-a-thon in April 2018 in The Hague. This allowed to interview leading experts in the area of interoperability standards in IT and medical IT. Table 4.9 shows the data describing the characteristics of the international experts who participated in the review. The total number of participants were eleven, were six came from the health domain i.e.

were participating in projects concerned with medical IT. Five others came from other domains working on IT projects with interoperability standards. However, all experts already conducted IT projects in other domains too. 63.6% percent of the participants stated the interest of their employers regarding the topic of this thesis. In each case, six persons indicated to work as engineers and managers apart of one participating user and one from a different interest group. As furthermore shown in table 4.9, all of the participating experts had cross-distributed knowledge in the described IHE domains i.e. IT infrastructure, radiology, laboratory but as well as standards in common sense (HL7, DICOM). Apart of the questionnaire the participating experts were asked about their time working in the professional field of application of interoperability standards. The result was that each of the experts had at least fifteen years of professional experience in this field.

Table 4.9 – Shows the data describing the characteristics of the international experts participating in the review at the EU Connect-a-thon 2018 in The Hague.

Number of Participants	11	Engineer	6
Strategic Interest in Topic	7	Management	6
Health-Domain	6	User	1
Other IT	5	Other	1
ITI	5	LAB	1
RAD	5	HL7	5
PCC	2	DICOM	5
PCD	1	Others	4

Table 4.10 provides an overview about the results for the yes/no-questions asked during the review.

Therefore, 100% stated that IHE processes are be applicable for BDIS data sharing in their opinion. 72.7% of the participants stated that they think that the developed profile including actors and transactions is sufficient to share BDIS. The remaining three stated that it is principally possible, but further investigation might be needed regarding the possibility of data streaming. The first question focused on the technical basis used in the IBD-Profile. Therefore five out of eleven participants

Table 4.10 – Shows the results of the yes/no-questions, answered during international experts review performed at the EU Connect-a-thon 2018 in The Hague.

Question	Yes	No	Comments
Are IHE processes applicable to share Big Data Information Sets (BDIS)?	11	0	-
Do you think that sharing of Big Data Information Sets (BDIS) according to the Actor/Transaction Diagram is feasible?	8	-	Other 3: depends on purpose and data. streaming needs inverstigation
Do you think that the following standards may be applied in this concept? (HL7 FHIR STU 3, HTTP/1.1, HTTP/2, URI: Generic Syntax, Additional HTTP Status Codes)	5	-	5 von 6 Engineers, Nr.6 depends on exact details; Others didn't comment

confirmed the selected technical basis and one said it depends on the detailed requirements. The others, did not comment on this questions.

Table 4.11 shows the results of the descriptive questions. The aim was to evaluate if the IBD-Profile fulfills the requirements a system must fulfill regarding BDIS sharing, from the point of view of the international experts. Therefore, important characteristics were the definition of meta-data describing the raw data, not forcing implementers to use a defined data model for the raw data. On the other hand, they indicated the importance of a possibility to structure the raw data and adding filtering options to search the data. Other comments were made on using lightweight technologies i.e. JSON, description of scenarios to show applicabilities and providing a broad governance structure.

Regarding security and privacy challenges and coverage through IHE, the experts generally stated that IHE provides an applicable basis, although access control is not covered through IHE profiles. Definition of semantics in the BDIS shall be used to provide access control criteria. Furthermore, they stated the importance of handling intellectual property questions and the support for the new EU General Data Protection Regulation (GPDR).

Table 4.11 – Shows the results of the descriptive questions, answered during international experts review performed at the EU Connect-a-thon 2018 in The Hague.

Question	Answers
What are the main important characteristics a BDIS exchange concept needs to fulfill to become widely accepted?	o Do not use a fixed data model o Important to describe whom to adress and domains o Meta-Data separate from BDIS data o Best possible governance Structure o Data needs to be defined, structured and matched source by source o Unstructured data only if not otherwise possible =>provide structured format o Quality of data, filtering options, easy access o JSON, scalable, standalone interface
Which security challenges do you see? Coverable by IHE Profiles?	o Anonymization, pseudonymization, data security, avoid leaks o Privacy depends on type of data, Intelectual Property issues o GPDR conform o Coverable by IHE, but no profile for access rights o Access Rights according to semantic fields in BDIS
Which scenarios do you see in your opinion for this application?	o Research o Health-Analytics, Evidence Based-Medicine, quality monitoring in healthcare institutions (huge effort today!) o Apps, Numerus scenarios =>Enrichment is the key. o Secondary use. Horizon 2020: personal health, riskgroups of patients o Financial allocation in medicine, pricing of services, Identifying KPIs for political decision making & evaluation o Public health, public health epidemic
Apart from IHE and related standards: Do you see any alternative interoperability specifications for this purpose?	o IHE-QRPH o NoSQL

The question of applicable scenarios for the profile was used to identify the importance and areas of applicability of the topic and profile. Hence, research (e.g. Horizon 2020) as a general topic, public health, quality monitoring, evidence-based medicine were stated to be important. Apps were stated to be important tools for collecting and sharing through its broad fields of application. Especially the term "secondary-use" of data for research and analysis was identified.

Last but not least, financial and political aspects were stated to be important i.e. financial allocation on projects and identifying Key Performance Indicators (KPI) for political decision making and evaluation processes. On the question if the experts see any other interoperability specifications, the IHE-QRPH profile as well as NoSQL was stated.

5 Discussion

As described in chapter 1, the term "big data" is a broad definition, which lead to the necessity of narrowing the focus to open data platforms in Europe to provide a comprehensible environment in this work. The results of the investigative analysis of these data platforms and formats, reinforced this decision. Comparing the numbers regarding the amount of collected data sets in the EU for the domains of health, transport and environment, clearly indicate that the number of environmental data sets exceed the numbers of the other two domains by far. Taking a closer look to the UK, which was the leader of open data set providers, this shows that they provide 5.6 times the amount of environmental data sets compared to transport related data sets. In case of Germany, which is the country with the second highest overall number of data sets, this number was 4.8 times higher. The reason for this may be that environmental data like pollution- or pollen-exposure data is very often measured by governmental institutions instead of private companies tending to prohibit access to data sources from the public. This finding fits also in the domain of transport related data, as parking garages are in most cases private and therefore not providing data for research. Similarly, it is the case with logistic companies measuring movement data of their transport vehicles. A further reason why the amount of environment data is much higher, is the data sensitivity level regarding

data privacy. Health data is much more sensitive than environmental pollution data and therefore typically not provided as open data. Only in the UK and Spain, the amount of provided health data sets was higher than the amount of transport related data sets. The reason for this might be legal regulations, but additionally a more open mindset of the citizens. Nevertheless, the investigated numbers regarding the amount of data sets in the specific domains indicated, that these could be used for this work.

The results of the subsequently followed data format analysis, shows a large number of different data formats. Figure 4.3 provides an overview about the most used formats in all three domains. From the top 10 used formats HTML and CSV were applied to the highest extend. Unfortunately, from a standardization point of view, especially the CSV format is very insufficient to be used for providing harmonized structural and semantical format and rules. This is underpinned by the processes shown in medical IT standardization for EHR and PHR data exchange. HTML on the other hand is generally a format for visualization purposes and therefore XML or JSON would support these purposes in a better way. However, XML as well as JSON were underrepresented compared to HTML, CSV or PDF. The reason for this might be its easy application and use. However, during investigation of the formats structure and semantics proved the lack of a common approach for the different domains. In rare situations standards were applied, but these were used for specific domain data for example defining geopoints.

Another important question was the identification of the communication protocols provided by the platforms. The RESTful architecture under the definition of CKAN, was used in most cases of investigated open data platforms. That is a light-weight communication technology providing a solid basis for data sharing of different formats. CKAN is an initiative that provides a definition of set of meta-data for the open data area. However, there is still a lack of a common syntactical and semantical definition for the data itself as well as clear process definition and it has a clear focus on open data platforms.

The selection process of standards and the technical basis was done in two steps

starting by the definition of user requirements. Therefore the definition of user requirements helped to reduce the amount of medical interoperability standards to a manageable set. The remaining selected standards focused on the processes of exchanging data as well as the data itself with its syntactically and semantically interoperable requirements. Especially the decision of using standards, which were message-, documents- or resources-based, was a crucial question at this point in time. This stage already highlighted IHE, PCHA and HL7 FHIR compared to the other standards like HL7 V2/V3, HL7 CDA or openEHR. Although HL7 FHIR lacked of process definitions it was clearly the most promising standard in this case as it fulfilled all other defined user requirements. Nevertheless, indications that a combination of openEHR or IHE with HL7 FHIR would provide a strong basis for this work, could be identified at the end of this step.

This indication was verified through the results of the analysis and the definition of the detail requirements. Taking the initial data collection process in case of medical domain as an example, within a PHR system, PHDs transmit data via mobile devices like smartphones to an EHR systems and each of this components have very specific technical requirements. This communication is not only covered by on SDOs and related standards, like by IHE profiles. Within the PCHA Continua guidelines, the first communication steps required to use BLE or X73 standards, which were introduced to optimize the exchange of information between low-power medical devices and routing devices. Subsequently, data shall be transmitted from the Application Hosting Device (AHD) to EHR systems to aggregate the information received in multiple messages. In the feasibility study for the healthcare data integration scenario, a concrete implementation of an abstract definition of an AHD was the ATG. As an example IHE provides the IHE DEC profile for this purpose, which enables to send event-driven messages containing health or health-administrative information to healthcare service providers using an EHR. An alternative or extensible approach is to use XDS/XDR profiles of IHE to generate and send medical documents like CDA documents directly on/from the user's AHD. Therefore, a health report is generated on these devices and brought into an XDS based system within an affinity domain. However, such systems need

huge flexibility in terms of exchanging different types of documents and content. By using the mentioned IHE components, it can be ensured to foster interoperability between PHR systems and EHR systems since these IHE profiles are based on international standards. This is a crucial factor to build a widely supported and future-proof system for data exchange as several vendors are adopting these standards around the world in multiple implementations. This is shown by the IHE and PCHA databases for products and implementations (IHE International, 2018c; Personal Connected Health Alliance (PCHA), 2018a). By March 2018, the latter database showed 118 Continua certified available devices. Therefore it can be expected that applying IHE as well as PCHA provides a solid basis for fulfilling these requirements. However, implementations always depend on local (country, organizational, personal) situations and requirements, which need to be considered.

When its about security requirements, the healthcare domain and health related data is an outstanding field as health data is highly sensitive. The IT landscape in healthcare is very diverse and data from measurements and health care reports can be transmitted from different sources to various destinations in various ways. Medical information related to persons and health profiles are extremely sensitive information and therefore the primary security objectives in accordance to the CIA-Triad/Extended CIA-Triad (confidentiality, integrity, availability, non-repudiation) need to be rigorously fulfilled. Taking the example of PHRs and EHRs, these systems are highly complex and need software component authentication, user authentication, logging, authorization of all involved persons and organizations, data encryption and user safety. IHE and its technical frameworks, provide a lot of different profiles and process definitions, for example the Audit Trail and Node Authentication (ATNA), Cross-Enterprise User Assertion (XUA), Internet User Authorization (IUA) or the Enterprise-User Authentication (EUA) profiles. These enable the establishment of state-of-the-art secure healthcare environments with a strong focus on information security and patient privacy protection.

IHE underlies a continuous iterative process, which fulfills new requirements derived over time by adding trial implementation profiles to always improve or extend

security and privacy requirements step by step. However, content of privacy consents and rules or regulations are touched only slightly, as the focus of IHE is clearly a technical one. Table 4.2 shows, that the need for software component authentication via certificates, audit logging and encrypted communication using public/private key cryptography via TLS, can be coverd by the ATNA profile. Requirements for user authentication is coverd by the EUA profile from IHE, but focuses strongly on IT-Infrastructure based systems e.g. EHRs. Taking PHR systems into account, i.e. mobile environments, this may not be sufficient. Regarding authorization requirements, IHE provides two profiles called XUA and IUA covering the base technologies of OAuth and SAML (Security Assertion Markup Language). These focus on issuing and distribution of security assertions for authentication and authorization in a distributed environment based on SOAP or RESTful web-services. However, in older systems SOAP protocols are widely used and these can be connected with SAML. Modern systems and especially mobile environments need IUA and OAuth to be connected with RESTful approaches as SOAP is not common here any more.

Summarizing, the analysis outcomes showed that specifications of IHE and PCHA can be used to fulfill specific security requirements. However, some security requirements, like end-to-end encryption is not yet clearly defined or solved, although approaches exist. Nevertheless, IHE as well as PCHA provide applicable concepts, which can be used as a basis to fulfill security and privacy requirements derived from individual projects.

A lot of legal requirements can be derived from the EU Data Protection Directive 95/46/EC and the GDPR. Technical requirements, can mostly be covered by the security measures described in the chapter before, but excludes explicitly processes and policies. This focuses strongly on the exchange of sensitive data. Another aspect is, as already stated, the data generation itself and the persons responsible for it. Hence, the question of the legal status of generated content arises. In medical terms this is especially interesting as there is a clear difference between data generated by medical professionals or layperson data from patients in home

environments. The latter may not be used as a solid basis in decision making processes. Nevertheless, these crowdsourced data is a very important source of data for generating big data information sets with a certain semi-controlled data quality. This is not a technical issue, but it is a legislative issue which has to be taken into account when implementing solutions. It is important to answer when and/or how laypeople are entitled to introduce their personal generated data in such systems or into the overall data distribution in terms of big data information set exchange.

Another crucial factor is that systems are interconnected, which need to fulfill specific legal requirements to cover security and privacy concerns defined on regional, national or international levels. At least the latter, as already stated, is covered in the EU to some degree by the Directive 95/46/EC, but national and regional aspects are still very diverse. Hence, the different nature of the data (health vs. environment vs. transport etc.) require different aspects of security and privacy. Taking health data as an example, high sensitivity needs a very high level of security through these requirements. Hence, IT components in this systems need to be designed to handle this highly sensitive data properly. The process for data administration must use appropriate IT security technologies. A continuous and iterative implementation as well as improvement process is necessary, as legal regulations may change and adaption is necessary. Especially this is of high importance as data is collected over longer periods of time. Therefore standards and harmonized specifications are highly beneficial as they provide explicit guidelines and are designed for long durability, iterative extensibility and improvement.

Mutually authenticated partners and systems as well as data quality is of importance and malign data insertion needs to be prevented in any case. Logging might be the tool of usage for this approach to support a reactive security model and improve the trustworthiness of system components during exchange. However, this may rise challenges when consumer devices such as smart-phones are used as they need to be integrated transparently and according to security regulations. In order to advice users about this processes, expert systems might interpret the available data and help with further processes. This solutions and developed components for data

exchange need to cover different operating systems and their multiple versions on mobile devices and the variety of different smartphone and tablet models, which increases the overall complexity. The analysis showed, that legal situations are very complex and not even clear due to the mix of national and regional requirements with international onces, although international effort is done as the GDPR shows.

The more detailed resolution of the requirements as well as the separation of SDOs, profiles and standards provided the basis for the final decision to make a combinatorial approach of using IHE and HL7 FHIR in a common concept. In this step a focus was placed on the technical-, security- and legal & regulatory requirements, as these were the most important once for defining a technical concept. The decision to use IHE or openEHR was made based on the fact that IHE provided a broader range of profiles covering the specific needs derived from security and technical requirements. However, when it came to legal and regulatory requirements, the coverage depth of IHE profiles for these requirements, decreased. An example for this was the integration of policies. IHE provides a basic setup covering a patients consent, but over this only refers to use XACML as a possible technology to cover management of policies without specific process and/or content definitions. However, looking at the profile level in detail it became evident that especially for the technical requirements IHE profiles didn't provide complete support for the use of HL7 FHIR. Profiles used for the exchange of data in this context were XDS/XDR, which focused on exchange of documents, and DEC (or others) focusing on message exchange. Both of these profile groups mostly fulfill the detailed technical criteria, but the support for mobile environment was largely decreased. This was enforced as the basic technologies used, e.g. SOAP based WS etc., would not support larger amounts of data in an efficient way. HL7 FHIR's approach of using resources once more emphasized its qualities for this works approach. Taking this into account, the final decision was made to use HL7 FHIR as a communication standard and the process definition according to IHE, which would support to re-use IHE's profiles for security & privacy through profile or actor grouping.

Based on the selection process a business domain overview including the description

of scenarios and the definition of the framework components was implemented. This resulted in the definition of the IBD-Profile, its actors and transactions together with HL7 FHIR and RESTful architecture. The aim was to provide a generic profile, which supports data-independent scenarios in the domain of big data. Hence, as analytics in terms of big data is finally the most frequent application, the IBD-Profile focused on the definition of an import/export system with sharing functionality in a controlled environment. It included a directory for storage of BDIS data sets and offering these data to be used in analytics in a standards based way. The directory integrates the storage of meta-data as well as the data itself providing the data in XML or JSON format. Hence, the profile doesn't apply a registry-repository separated approach to interconnect different BDIS data sinks. The reason for this was that using RESTful as a base technology and HL7 FHIR as a base standard, both by nature provide query possibilities and the integration of meta-data per definition. Therefore, a client may search and retrieve data simultaneously from different BDIS directories to collect relevant information for analysis. Furthermore, a registry/repository approach reveals the question of who i.e. which institution runs and pays for the registry and do we need multiple ones in the end too.

From a security perspective, the profile allows the seamless integration of OAuth, which in this terms is a crucial requirement to support query/retrieve of BDIS from different BDIS directories linked to a federation. However, the concrete definition of policies and especially its attributes was not done as it is beyond scope of this work. Additionally the definition of the profile allows efficient integration of other security measures covered by IHE like including audits in terms of reactive security measures using IHE ATNA profile. Generally the recommendation of security measures are not forcing to use specific implementations, but more providing a guidance and proving the seamless integration with examples. Nevertheless, it has to be stated that the security mechanisms were not implemented in the prototypes as the focus lied on the prove of concept by testing the technical basis with the data derived from the scenarios.

The described scenarios for healthcare, environment and transport data integration

and the designed architectures provided not only the basis for describing different actor grouping examples, but moreover to study technical feasibility. Therefore, the prototypes successfully proved the integration of healthcare, environmental and transport related data using the IBD-Profile and its connected processes and standards. Taking a closer look to the health data integration study, the propose can be used for any wearable activity tracker system, independent of the place of application on the body as the biometric indicators used were the most prominent ones, which were measured by common devices. The support for semantic interoperability was a crucial factor in this study. Searching for codes on public tools like the Austrian terminology server, revealed that in some cases applicable codes could not be found. An example was the "seconds of sleep per minute" attribute, were several codes could be identified, but they were designed for specific applications like in the context of EEG. However, the aim was to provide the stated type of data resolution to support analysis like its made on manufacturers platforms (e.g. polar flow). Furthermore, this format fits to the other indicators for defining clear transmission blocks, which can be dynamically transmitted at any point in time. Similar observations were made in the environmental and transport data integration studies. Hence, in both cases code-lists/value-sets were defined to support semantic interoperability. Specifically for the transport data scenario, it has to be mentioned that such a code-system has to be developed for each city specifically. A specific minimal data set might be defined for this data, but this is not scope of this work. Furthermore, disability specific information might also be included in the stations, not in the vehicle itself. That might also rise the need for dynamically adapt the HL7 FHIR Observation structure.

Based on implementation experiences of CDA, HL7 V2 and HL7 FHIR standards, it can be stated that the latter allows huge flexibility for novel approaches, although still providing support of interoperability. In the transport data integration study, an approach was to use a sampledValueData element in the observation resource examples to aggregate measurement values with the same nature just changing over time, like ECG potentials. This approach would have allowed a much more compact FHIR resource as the XML-meta data describing the value itself would be reduced

noticeably. Unfortunately, it was not possible without harming or extending the specification as the disability information could not be integrated reasonably in this approach. The shown example HL7 FHIR observation resource could finally be design without any additional changes to the specification. This flexibility, in terms of using multiple instances of the components-tag of the HL7 FHIR Observation resource, allows the combination of several data from other domains. Furthermore, it includes a based-on-tag for references to other HL7 FHIR resources, which allows the reference of other HL7 FHIR Observation resources and therefore the combination of data. Management of ids is not scope of the IBE-Profile. A recommendation for this requirement is the use of IHE identity management profiles, which provide processes and examples for this purpose. In the actual examples IDs where mostly auto-generated examples. In practical examples OIDs shall be used as also recommended by IHEs identity management profiles. Additionally, it has to be stated that using XML generates a lot of overhead data in respect the data itself. This may lead to problems in case of using huge amounts of data. Possible approaches to overcome this problems, are using JSON instead and apply data-segmentation. Finally, from the perspective of this studies, it can be stated that HL7 FHIR is a promising standard for the integration of data from other domains.

As the IBD-Profile was designed to be used as a generic model independent of the domain, this approach was used to define necessary meta-data attributes in the defined BDIS-Extension for describing the data in the HL7 FHIR observation. Although these attributes were sufficient to carry out the three feasibility studies, more attributes may be defined in further steps. An example could be that a finer granular description of the BDIS context might be useful for more efficient search procedures. In the actual prototype only a narrative description was integrated, although integration of codes for sub-categories might improve this status. The advantage is that the defined BDIS-Extension can be adapted by simply adding attributes to the developed definition. After such changes, the STU 3.0.1 definition has to be adapted again, as it was similarly done in this work. During the phase of testing, a copy of the STU 3.0.1 was extended and integrated in the official FHIR Validator. Without this step a proper validation of developed application examples

would not be possible. Additionally, the BDIS-Extension was placed on a public server (see 4.3.3) to allow all communication partners applying the IBD-Profile, to use the definition for integration in their software. After this, the code samples could be successfully tested according their conformance to the extended specification. However, the process defined by HL7 for defining HL7 FHIR extensions, normally require to register any extension on their public service. This step was bypassed as the work focused primarily on prototyping. However in future steps this process can be trigged.

The first questions of the experts review were focusing on identifying the experts appropriateness for this topic. The division of participants in the two main groups of engineers, but as well as managers and additionally users, provides a broader point of view on the topic through the experts. Similarly the domain knowledge distribution i.e. 6 in the health domain and 5 in other IT relevant domains provide more general view on the topic. Nevertheless it has to be stated that the most experts had experiences from different domains, but one had to be chosen as a main knowledge domain. The results regarding the knowledge of the experts furthermore may support consideration of broader perspectives and standards. The high number in ITI, RAD as well as DICOM and HL7 on the other hand indicate that the experts have solid knowledge of standards used in this work. Apart of this the experts were asked according their years of professional work on the topic of interoperability standards. Each of the participants worked at least 15 years on that topic. Summarizing this, it was concluded that these experts were capable to be integrated in this review and additionally the importance of the topic was confirmed through the fact that 64% stated that they have strategic interest in this topic.

In the next step of the review, the yes/no-questions were used to directly get feedback to the developed approach in this work. Hence, all experts stated that IHE processes are applicable to develop a standardized solution for sharing BDIS. Furthermore the huge percentage of conformation on the IBD-Profile can be interpreted as a first successful step to providing a process for sharing of BDIS. The experts had only concerns and uncertainties of a possible application for the purposes of streaming

of data. The IBD-Profile in its current status is not focusing on this needs, but further development is definitely necessary. The last question of the second step was focusing on evaluating the decisions of the base standards. This question was the most important one to confirm the correct selection of standards. This targets the decision for HL7 FHIR, Http, RESTful architecture and as a technological basis for a lightweight process for the exchange of data context of Big Data. Based on the answers of the experts, it finally can be stated that five out of six engineers confirmed the proper selection of base standards for this topic and the sixth expert did not reject, but stated that it might depend on the purpose of application. Answers to this questions came from the experts representing only engineers or mixed roles like manager & engineers. This indicated that only engineers answered this question, as detailed technical knowledge is needed to do so.

The last step of the review contained the section with narrative questions. The purpose of this questions was to gain indications out of the experts answers to confirm the importance of the thesis topic, stated security considerations and identify alternative standards. Nevertheless, the most important aim was to evaluate that the developed IBD-Profile supports the most important characteristics of a BDIS exchange concept. Summarizing the answers, it can be stated that the IBD-Profile fulfills all the experts characteristics. They stated that meta-data should be used separated from the BDIS data fulfilling a wide governance. This is fulfilled by the defined BDIS structure through the extension. Additionally it fulfills the stated characteristics regarding the requirement of having structured data itself. By the use of HL7 FHIR observation resource, it is possible to structure different kinds of BDIS in forms like periodic or sampled data structures (or any other kind supported by HL7 FHIR as stated in 4.3.3). This supports syntactic, but also semantic interoperability to the highest possible degree through definition of the structure not only on a meta-data level (by the BDIS-Extension), but also by using the HL7 FHIR Observation specification to structure data and connect it to codes. BDIS-Extension meta-data attributes allow coding, which adds level of semantic interpretation. This which was another stated characteristic from the experts. This furthermore fulfills the characteristic of filtering the data for search and investigation based on

semantics, but also supports the quality of data to some degree. Additionally the characteristic of using JSON, which is apart of XML the second format supported by HL7 FHIR and the IBD-Profile, and easy access through using JSON/XML and HTTP is covered. Finally, the importance to describe scenarios for application of the profile and whom to address was stated by an expert. This was done by defining three scenarios describing the integrating of data from different domains and the performance of feasibility studies based on these scenarios. Based on this, it can be stated that the characteristics from the experts are fulfilled to a very high degree by this works approach.

The topic security and privacy was covered with the sixth question of the questionnaire used during the experts review. The experts confirmed the importance of considering data security and privacy especially in connection to the new GDPR from the EU. Additionally, the topics of anonymization, pseudonymization and the handling of intellectual property requirements were stated to be of importance. All of this requirements can be integrated in the developed IBD-Profile via using the defined meta-data attributes shown in the BDIS-Extension (see figure 4.11)or the attributes defined in the HL7 FHIR observation resource specification. Finally, the experts confirmed that IHE provides sufficient security profiles to integrated security and privacy measures in accordance to the stated requirements. However, the experts also confirmed that there is no access control profile defined by IHE focusing on the management (i.e. issuing, storing, maintaining) of policies used for deciding based on and enforcing rules for access control. Nevertheless, in this work this fact was stated during the standard selection process and a solution i.e. recommendation of XACML, was also recommended.

The results from the question regarding the fields of application, confirmed additionally the importance of the thesis topic and the spacious application areas. Generally the topic of research with a special focus on medical research areas like health analytics, evidence based medicine and quality monitoring in healthcare as well as public health were stated to be very important areas. This was underpinned by the experts as they confirmed international initiatives, e.g. in research proposals

and projects in Horizon 2020 focusing on topics of analysis of data for personal health to improve healthcare processes of risk groups of patients. Especially the secondary-use of data from IT systems and projects were stated as important scenarios. Exactly this fact is integrated in the scenarios defined in this work i.e. integration of health, environment and transport data. Also the importance of Apps as data providers/gateways for mapping was stated by an expert. That was fulfilled by this works approach too, as the three defined scenarios in the IBD-Profile are strongly based on Apps working as gateways and providing data mappings to the standards based BDIS format. Finally, it was stated that well structured data and its exchange is important in terms of identifying Key Performance Indicators (KPI) for political and financial decision making processes. The IBD-Profile provides structuring and sharing of data together with semantics to support analytic processes independent of the purpose. Based on this fact and the feasibility studies proving its applicability, it can be concluded that these scenarios are supported to a very high degree by the IBD-Profile.

Finally, the last question focused on the identification of other interoperability standards for the approach of this work. The answers did not deliver any huge alternatives. One stated was to use NoSQL for data storage, which still may be integrated in the BDIS-Directory. A second alternative commented was the IHE domain of Quality, Research and Public Health (QRPH). The latter focuses on sharing information regarding quality improvement and providing population based health surveillance. However, taking a closer look did not indicate any kind of overlap or incompatibility with the contained profiles from the perspective of the IBD-Profile. Nevertheless, it might be considered that the IBD-Profile could be integrated in this healthcare specific domain in terms of secondary-use of data.

6 Conclusion and Future Work

The intention of this thesis was to explore the hypotheses of combining big data information sets, from a methodical point of view of how these data sources can be combined in a standardized way. This connects to the research question, if interoperability standards from the medical IT domain are applicable to other domains to support exchange and interpretation of BDIS. An investigative analysis was performed to get an overview about the open data platforms and its formats, followed by a selection process for medical interoperability standards according to well defined criteria. Subsequently, the next objective was to establish a big data business domain overview and the conceptualization of re-usable, modular and combinable IT framework components for standardized exchange of BDIS. This was followed by prototyping and execution of feasibility studies in the domains of health, transport and environment as a prove of concept of the framework components and especially the data syntax and semantics. Technical validation was done by defined state-of-the-art testing procedures using validation tools. Finally, the last objective was to evaluate the concept and the approach by application of an international experts review.

As a result of the investigative analysis and the selection process, IHE processes and

HL7 FHIR were selected as standards to define a BDIS sharing concept. The results of the experts review confirmed this choice and therefore it can be concluded that the selected standards were correct to provide an approach to cover as many scenarios as possible. This was furthermore confirmed by the fact that the scenarios and related feasibility studies showed successful results in technical implementation. This results were confirmed from the perspective of technical validity with conformance tests. Hence, it can be concluded that based on this successful results further scenarios from other domains can be implemented according the defined IBD-Profile. The experts review results additionally showed that the recognized criteria for defining the IBD-Profile overlapped to a very high degree with the experts stated characteristics. Thus, it can be concluded that the most important characteristics necessary to support a broad field of application were met during the design process of the IBD-Profile. The experts review results furthermore confirmed the considerations made during the design process for the security and privacy implementations. Although the GPDR, as an example for applications in the EU, put a huge load of requirements on IT systems in terms of security and privacy, it can be concluded that the recommended technologies and approaches can be used as a starting point to meet these requirements.

Apart of the successful results and its confirmation through the experts review results, two points were identified to be improved in future steps. The first one is that it might make sense to undock the meta-data i.e. the BDIS-Extension from the HL7 FHIR observation resource. The reason for this is that a separated actor for storing only the meta-data may support the interconnection of existing data repositories through an registry/repository approach. This would require to first query the registry for data and its location and then subsequently request the information from the distributed data sinks. This would decreasing the numbers of search requests as in the actual setup each BDIS-Directory has to be queried for data (BDD-02) followed by the retrieval requests (BDD-03). Hence, in future steps undocking the meta-data from the HL7 FHIR observation resource might be considered and tested.

The second point is concerned with streaming of data. As there might be scenarios in terms of secondary-use of data e.g. ECG measurements, the actual concept of this approach might not fulfill criteria of streaming data. However, to provide a solution for this possible gap, further detailed scenarios need to be analyzed and technical tests need to be performed to provide an efficient solution and answer to this question.

Additionally, during the last quarter of this work, the European Interoperability Reference Framework (EIRA) came into public focus. Hence, EIRA provides guidance and a tool set regarding development of interoperable public services in the EU. Therefore, future steps need to be done in conformance to this framework, as public and governmental data plays a key role in big data science. Furthermore, the last month during the establishment of the thesis, the EU started the Big Data Test Infrastructure (BDTI)(European Union, 2018) under the CEF Digital to support public entities in understanding, using, managing and testing Data in this terms. This shows that the work done in this thesis is of large scale interest and therefore underpins the work done.

Summarizing this, it can be concluded that the developed IBD-Profile defines a set of framework components for standards based sharing of BDIS, which was successfully tested and validated through international experts. This supports interoperability from a syntactically and semantically point of view for meta-data as well as the raw data using interoperability standards derived from the medical IT domain. Hence, it can be stated that interoperability standards from the medical IT domain are applicable to other domains to support exchange of BDIS.

Bibliography

Ahmed, E., Yaqoob, I., Hashem, I. A. T., Khan, I., Ahmed, A. I. A., Imran, M., and Vasilakos, A. V. (2017). The role of big data analytics in Internet of Things. *Computer Networks*, 129:459–471. 2

Aliakbarpoor, Y., Comai, S., and Pozzi, G. (2017). Designing a HL7 compatible personal health record for mobile devices. In *2017 IEEE 3rd International Forum on Research and Technologies for Society and Industry (RTSI)*, pages 1–6. IEEE. 28

Andress, J. (2014). *The basics of information security : Understanding the fundamentals of InfoSec in theory and practice*. 2nd edition. 32

Apache (2018). Welcome to Apache Hadoop! [online available:] http://hadoop.apache.org/, (last access: 2018-06-10). 64

Austrian Ministry of Health and Woman (2018). Terminologie-Browser. [online available:] https://termpub.gesundheit.gv.at/TermBrowser/gui/main/main.zul;jsessionid=83712581E4EEC956C8278F084A64ACB8.vm{_}pub{_}0, (last access: 2018-02-16). 90

Austrian Ministry of Health and Women (2018). OID Portal Österreich. [online

available:] https://www.gesundheit.gv.at/OID{_}Frontend/index.jsp, (last access: 2018-11-30). 84

Bender, D. and Sartipi, K. (2013). HL7 FHIR: An agile and RESTful approach to healthcare information exchange. In *Proceedings of CBMS 2013 - 26th IEEE International Symposium on Computer-Based Medical Systems*, pages 326–331. 24

Bloomfield, R. A., Polo-Wood, F., Mandel, J. C., and Mandl, K. D. (2017). Opening the Duke electronic health record to apps: Implementing SMART on FHIR. *International Journal of Medical Informatics*, 99:1–10. 28

Boone, K. W. (2011). *The CDA(TM) Book*. Springer. 21

Bresnick, J. (2018). A FHIR Future Burns Brightly for Population Health Management. [online available:] https://healthitanalytics.com/features/a-fhir-future-burns-brightly-for-population-health-management, (last access: 2018-02-25). 19

Briante, O., Campolo, C., Iera, A., Molinaro, A., Paratore, S. Y., and Ruggeri, G. (2014). Supporting augmented floating car data through smartphone-based crowd-sensing. *Vehicular Communications*, 1(4):181–196. 1, 63

Brull, R. (2013). 5 Things to Know About HL7 FHIR - Health Standards. [online available:] http://healthstandards.com/blog/2013/03/26/hl7-fhir/, (last access: 2018-02-25). 19

CEN/TC 278 (2018). About us - CEN/TC 278 Intelligent Transport Systems (ITS). [online available:] http://www.itsstandards.eu/tc278, (last access: 2018-03-11). 29

CKAN (2018). About - ckan. [online available:] https://ckan.org/about/, (last access: 2018-03-11). 58

Clarke, M., de Folter, J., Verma, V., and Gokalp, H. (2017). Interoperable End-to-End Remote Patient Monitoring Platform based on IEEE 11073 PHD and ZigBee

Health Care Profile. *IEEE Transactions on Biomedical Engineering*, pages 1–1. 27

Cloudera (2018). Machine Learning — Analytics — Cloud - Cloudera. [online available:] https://www.cloudera.com/, (last access: 2018-06-10). 64

COBWE-Project (2016). The Potential of Crowdsourced Environmental Data to Support Policy Making :. 1

Data Go Jp (2018). Data - DATA GO JP. [online available:] http://www.data.go.jp/data/en/, (last access: 2018-06-10). 56

Davis, N. A. and LaCour, M. (2014). *Health Information Technology*. Elsevier Saunders, St. Lous, third edit edition. 11

Doelger, R. and Geissler, T. (2012). DATEX II - The standard for ITS on European Roads DATEX II - CEN TS 16157. 30

Dolin, R. H., Alschuler, L., Boyer, S., Beebe, C., Behlen, F. M., Biron, P. V., and Shabo (Shvo), A. (2006). HL7 Clinical Document Architecture, Release 2. *Journal of the American Medical Informatics Association*, 13(1):30–39. 19

Drazen, E. and Rhoads, J. (2011). Using Tracking Tools to Improve Patient Flow in Hospitals Executive Summary. *California Healthcare Foundation*. 2

eHealth Network (2015). eHealth Network - Refined eHealth European Interoperability Framework. xxi, 10, 11

ELGA (2018). ELGA: Meine elektronische Gesundheitsakte. [online available:] https://www.elga.gv.at/index.html, (last access: 2018-02-23). 3

EpSos-Project (2016). About: European Patients Smart Open Services (epsos). [online available:] http://www.epsos.eu/home/about-epsos.html, (last access: 2018-02-23). 3

European Commission (2018a). About ISA2-ISA2. [online available:] https://ec.europa.eu/isa2/isa2{_}en, (last access: 2018-02-24). 8

European Commission (2018b). Data protection in the EU - European Commission. [online available:] https://ec.europa.eu/health/data{_}collection/data{_}protection/in{_}eu{_}en, (last access: 2018-03-09). 31

European Commission (2018c). Interoperability — Digital Single Market. [online available:] https://ec.europa.eu/digital-single-market/en/interoperability, (last access: 2018-12-02). 3

European Commission (2018d). New European Interoperability Framework - Promoting seamless services and data flows for European public administrations. *Publications Office of the European Union*. 8, 40

European Commission (2018e). The New European Interoperability Framework - Overview. [online available:] https://ec.europa.eu/isa2/eif{_}en, (last access: 2018-02-24). 3, 8

European Data Portal (2018). Home page - European Data Portal. [online available:] https://www.europeandataportal.eu/en/, (last access: 2018-03-11). 57

European Patients Forum. The new EU Regulation on the protection of personal data: what does it mean for patients? A guide for patients and patients' organisations 2 The new EU Regulation on the protection of personal data: what does it mean for patients? Technical report. 31

European Telecommunications Standards Institute (2018). Intelligent Transport Systems - What are Intelligent Transport Systems? [online available:] http://www.etsi.org/images/files/ETSITechnologyLeaflets/IntelligentTransportSystems.pdf. 29

European Union (2018). BDTI What is Big Data Test Infrastructure. [online available:] https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/BDTI+What+is+Big+Data+Test+Infrastructure, (last access: 2018-11-30). 127

Ferranti, J. M., Musser, R. C., Kawamoto, K., and Hammond, W. E. (2006). The Clinical Document Architecture and the Continuity of Care Record: A Critical

Analysis. *Journal of the American Medical Informatics Association*, 13(3):245–252. 20

Government of Canada (2018). Open Data. [online available:] https://open.canada.ca/en/open-data, (last access: 2018-06-10). 56

Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., and Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, 26(3):191–209. 2

Health Level Seven International (2014). About Health Level Seven International. [online available:] http://www.hl7.org/about/index.cfm?ref=common. 18

Health Level Seven International (2018a). Comparison-cda - FHIR v3.0.1. [online available:] https://www.hl7.org/fhir/comparison-cda.html, (last access: 2018-03-08). 24

Health Level Seven International (2018b). Index - FHIR v3.0.1. [online available:] https://www.hl7.org/fhir/, (last access: 2018-03-08). xxi, 24, 25, 26

Health Level Seven International (2018c). Introduction to HL7 Standards. [online available:] http://www.hl7.org/implement/standards/index.cfm?ref=nav, (last access: 2018-02-25). 18, 22, 23, 35

Health Level Seven International (2018d). Observation - FHIR v3.0.1. [online available:] https://www.hl7.org/fhir/observation.html, (last access: 2018-02-17). 25

Health Level Seven International (2018e). Reference Information Model (RIM) Downloads. [online available:] http://www.hl7.org/implement/standards/rim.cfm, (last access: 2018-02-28). 22

Health Level Seven International (2018f). Summary - FHIR v3.0.1. [online available:] https://www.hl7.org/fhir/summary.html, (last access: 2018-01-14). 25

Health Level Seven International (2018g). Validation - FHIR. [online available:] https://www.hl7.org/fhir/validation.html, (last access: 2018-02-15). 47

Helbing, D. (2017). Digitization 2 . 0 : A New Game Begins. 1

Herzog, J., Pohn, B., Forjan, M., Sauermann, S., and Urbauer, P. (2014). Education for eHealth-a status analysis. In *Studies in Health Technology and Informatics*, volume 198, pages 172–179. 34

HL7 Austria (2013). *ELGA CDA Implementierungsleitfäden HL7 Implementation Guide for CDA(R) R2: Allgemeiner Implementierungsleitfaden für ELGA CDA Dokumente*. 21

HL7 International (2018). Http Status Codes - FHIR v3.0.1. [online available:] https://www.hl7.org/fhir/http.html, (last access: 2018-12-02). 73, 74

Hong, J., Morris, P., and Seo, J. (2017a). Interconnected Personal Health Record Ecosystem Using IoT Cloud Platform and HL7 FHIR. In *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 362–367. IEEE. 27

Hong, N., Wang, K., Yao, L., and Jiang, G. (2017b). Visual FHIR: An Interactive Browser to Navigate HL7 FHIR Specification. In *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 26–30. IEEE. 28

Huang, E.-W., Hsiao, S.-H., and Liou, D.-M. (2003). Design and implementation of a web-based HL7 message generation and validation system. *International Journal of Medical Informatics*, 70(1):49–58. 23

Huang, E.-W., Wang, D.-W., and Liou, D.-M. (2005). Development of a deterministic XML schema by resolving structure ambiguity of HL7 messages. *Computer Methods and Programs in Biomedicine*, 80(1):1–15. 23

IBM (2018). IBM Watson. [online available:] https://www.ibm.com/watson/, (last access: 2018-06-10). 64

IEEE Standards Association (2004). *Health Informatics - Personal health device communication Part 10101: Nomenclature*. IEEE. 12

IEEE Standards Association (2012). *Health Informatics - Personal health device communication Part 00103: Overview*. 13

IEEE Standards Association (2014). *Health Informatics - Personal health device communication Part 20601 : Application profile - Optimized Exchange Protocol*. 12

IEEE Standards Association (2018). ISO/IEC 9834-1:2012 - Information technology – Procedures for the operation of object identifier registration authorities: General procedures and top arcs of the international object identifier tree. [online available:] https://www.iso.org/standard/58055.html, (last access: 2018-11-30). 83, 101

IEEE Xplore Digital Library (2018). IEEE. http://ieeexplore.ieee.org/Xplore/home.jsp. 26, 41

IHE International (2018a). About IHE. [online available:] http://ihe.net/IHE{_}Process/, (last access: 2018-02-25). 13

IHE International (2018b). Gazelle - INTEROPERABILITY & CONFORMANCE TESTING FOR E-HEALTH INFORMATION SYSTEMS. [online available:] https://gazelle.ihe.net/, (last access: 2018-02-23). 13, 14

IHE International (2018c). IHE Connectathon Results Browser. [online available:] https://connectathon-results.ihe.net/, (last access: 2018-03-25). 42, 114

IHE International (2018d). IHE Member Organizations. [online available:] http://www.ihe.net/Member{_}Organizations/, (last access: 2018-02-25). 14

IHE International (2018e). IHE Process. [online available:] http://ihe.net/IHE{_}Process/, (last access: 2018-02-25). 13, 42

IHE International (2018f). Pre-Connectathon/MESA Software - IHE. [online available:] http://ihewiki.wustl.edu/wiki/index.php/Pre-Connectathon/MESA{_}Software, (last access: 2018-02-25). 13

Institute of Electrical and Electronics Engineers (2018). IEEE - About IEEE. [online available:] https://www.ieee.org/about/about{_}index.html, (last access: 2018-02-25). 12

International Organization for Standardization (2018). ISO/TC 204 - Intelligent transport systems. [online available:] https://www.iso.org/committee/54706.html, (last access: 2018-03-11). 29

Iroju, O., Soriyan, A., Gambo, I., and Olaleke, J. (2013). Interoperability in Healthcare: Benefits, Challenges and Resolutions. *International Journal of Innovation and Applied Studies ISSN*, 3(1):2028–9324. 7

ISO/TC 204 Intelligent Transport Systems (2013). *ISO/TS 18234-1: Intelligent transport systems – Traffic and travel information via transport protocol experts group, generation 1 (TPEG1) binary data format – Part 1: Introduction, numbering and versions (TPEG1-INV)*. ON. 29

ISO/TC 204 Intelligent Transport Systems (2016). *ISO/TS 21219-1: Intelligent transport systems – Traffic and travel information (TTI) via transport protocol experts group, generation 2 (TPEG2) – Part 1: Introduction, numbering and versions (TPEG2-INV)*. 29

Janki, Z. R., Szabo, Z., Bilicki, V., and Fidrich, M. (2017). Authorization solution for full stack FHIR HAPI access. In *2017 IEEE 30th Neumann Colloquium (NC)*, pages 000121–000124. IEEE. 28

Jardim, S. V. (2013). The Electronic Health Record and its Contribution to Healthcare Information Systems Interoperability. *Procedia Technology*, 9:940–948. 7

Lamprinakos, G. C., Mousas, A. S., Kapsalis, A. P., Kaklamani, D. I., Venieris, I. S., Boufis, A. D., Karmiris, P. D., and Mantzouratos, S. G. (2015). Using FHIR to develop a healthcare mobile application. In *Proceedings of the 2014 4th International Conference on Wireless Mobile Communication and*

Healthcare - "Transforming Healthcare Through Innovations in Mobile and Wireless Technologies", MOBIHEALTH 2014, pages 132–135. 28

Legal Information System of the Republic of Austria (2012). 111. Bundesgesetz: Elektronische Gesundheitsakte-Gesetz-ELGA-G. [online available:] https://www.ris.bka.gv.at/Dokumente/BgblAuth/BGBLA{_}2012{_}I{_}111/BGBLA{_}2012{_}I{_}111.html, (last access: 2018-03-09). 31

Legal Information System of the Republic of Austria (2018a). RIS - Datenschutzgesetz 2000 - Bundesrecht konsolidiert, Fassung vom 09.03.2018. [online available:] https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=bundesnormen{&}Gesetzesnummer=10001597, (last access: 2018-03-09). 30, 31

Legal Information System of the Republic of Austria (2018b). RIS - Gesundheitstelematikgesetz 2012 - Bundesrecht konsolidiert, Fassung vom 09.03.2018. [online available:] https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen{&}Gesetzesnummer=20008120, (last access: 2018-03-09). 31

Maeng, D. D., Starr, A. E., Tomcavage, J. F., Sciandra, J., Salek, D., and Griffith, D. (2014). Can Telemonitoring Reduce Hospitalization and Cost of Care? A Health Plan's Experience in Managing Patients with Heart Failure. *Population Health Management*, 17(6):340–344. 64

Marton, P., Urbauer, S. P., and Haindl, P. (2017). Evaluierung von Fitness-Tracker APIs für das Monitoring von gesundheitsrelevanten Indikatoren. Master Thesis, UAS Technikum Wien. xix, 44, 45

McKinsey Global Institute (2011). Big data: The next frontier for innovation, competition, and productivity. 2

MongoDB (2018). MongoDB for GIANT Ideas — MongoDB. [online available:] https://www.mongodb.com/, (last access: 2018-06-10). 64

Müller, M. L., Ückert, F., Bürkle, T., and Prokosch, H.-U. (2005). Cross-institutional data exchange using the clinical document architecture (CDA). *International Journal of Medical Informatics*, 74(2-4):245–256. 20

Nyotowidjojo, I., Erickson, R. P., and Lee, K. S. (2016). Crowd-Sourcing Syncope Diagnosis: Mobile Smartphone ECG Apps. *The American Journal of Medicine*, 129(4):e17–e18. 1, 63

Open Data Bits (2014). Combine Chicago Crime Data with Census Data for New Insights. [online available:] http://opendatabits.com/combine-chicago-crime-data-with-census-data-for-new-insights/, (last access: 2018-02-23). 2

Oxford Dictionaries (2018). digitization - Definition of digitization in English by Oxford Dictionaries. [online available:] https://en.oxforddictionaries.com/definition/digitization, (last access: 2018-02-23). 1

Pahontu, R., Schneider, G., Bergh, B., and Merzweiler, A. (2015). An IHE based gateway architecture to link healthcare IT with medical devices in the operating room. In *2015 17th International Conference on E-health Networking, Application & Services (HealthCom)*, pages 586–589. IEEE. 27

Personal Connected Health Alliance (PCHA) (2016). H.811 Personal Health Devices Interface Design Guidelines. page 114. xxi, 17, 22, 48

Personal Connected Health Alliance (PCHA) (2018a). Continua Certified Product Showcase — Personal Connected Health Alliance. [online available:] http://www.pchalliance.org/product-showcase, (last access: 2018-02-16). 17, 42, 114

Personal Connected Health Alliance (PCHA) (2018b). Personal Connected Health. [online available:] http://www.pchalliance.org/, (last access: 2018-01-10). 16

Personal Connected Health Alliance (PCHA) (2018c). Resources — Personal Connected Health Alliance. [online available:] http://www.pchalliance.org/resources, (last access: 2018-02-16). 17

Press, G. (2014). 12 Big Data Definitions: What's Yours? — Forbes Tech. [online available:] https://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours/{#}50a073b313ae, (last access: 2018-02-23). 2

Public Health Data Standards Consortium (2018). General Information about the Public Health Data Standards Consortium. [online available:] http://www.phdsc.org/standards/health-information/D{_}Standards.asp{#}ListofOrganizations, (last access: 2018-02-25). 11

PubMed. NCBI, US National Library of Medicine. http://www.ncbi.nlm.nih.gov/pubmed. 41

Ragged, B. G. (2010). *Information Security Management: Concepts and Practice - Bel G. Raggad - Google Books*. CRC Press. 32

Rodrigues, J. (2010). *Health Information Systems: Concepts, Methodologies, Tools, and Applications ... - Google Books*. IGI Global. 22

Rodrigues, J. J. P., Sendra Compte, S., de la Torra Diez, I., Rodrigues, J. J. P., Sendra Compte, S., and de la Torra Diez, I. (2016). Clinical Document Architecture. In *Advances on Sensors for Health Systems*, pages 33–51. Elsevier. 19

Ruiz-Alarcon-Quintero, C. (2016). Harmonization of Transport Data Sources According to INSPIRE Data Specification on Transport Networks. *Transportation Research Procedia*, 18:320–327. 30

Sanchez, Y. K. R., Demurjian, S. A., and Baihan, M. S. (2017). Achieving RBAC on RESTful APIs for Mobile Apps Using FHIR. In *2017 5th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud)*, pages 139–144. IEEE. 28

ScienceDirect (2018). Science, health and medical journals. http://www.sciencedirect.com/. 26, 41

Shaw, J. (2014). Understanding big data leads to insights, efficiencies, and saved lives — Harvard Magazine. [online available:] https://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal, (last access: 2018-02-23). 2

Shields, P. M. and Rangarajan, N. (2013). A Playbook for Research Methods: Integrating Conceptual Frameworks and Project Management. *New Forums Press Inc*, page 292. 38, 41

Sloane, E. B. and Gehlot, V. (2016). Improved population health surveillance and chronic disease management using secure email: Application of the DIRECT, IEEE 11073, HITSP, and IHE standards and protocols. In *2016 18th Mediterranean Electrotechnical Conference (MELECON)*, pages 1–3. IEEE. 27

Spelman, T., Gray, O., Trojano, M., Petersen, T., Izquierdo, G., Lugaresi, A., Hupperts, R., Bergamaschi, R., Duquette, P., Grammond, P., Giuliani, G., Boz, C., Verheul, F., Oreja-Guevara, C., Barnett, M., Grand'Maison, F., Edite Rio, M., Lechner-Scott, J., Van Pesch, V., Fernandez Bolanos, R., Flechter, S., Den Braber-Moerland, L., Iuliano, G., Amato, M. P., Slee, M., Cristiano, E., Saladino, M. L., Paine, M., Vella, N., Kasa, K., Deri, N., Herbert, J., Moore, F., Petkovska-Boskova, T., Alroughani, R., Savino, A., Shaw, C., Vucic, S., Santiago, V., Bacile, E. A., Skromne, E., Poehlau, D., Cabrera-Gomez, J. A., Lucas, R., and Butzkueven, H. (2014). Seasonal variation of relapse rate in multiple sclerosis is latitude dependent. *Annals of Neurology*, 76(6):880–890. 2

Tolk, A., Diallo, S. Y., and Turnitsa, C. D. (2013). Applying the Levels of Conceptual Interoperability Model in Support of Integratability, Interoperability, and Composability for System-of-Systems Engineering. 7, 13, 17, 18

Tseng, T.-W., Yang, C.-Y., and Liu, C.-T. (2016). Designing Privacy Information Protection of Electronic Medical Records. In *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 75–80. IEEE. 27

United States Department of Transportation (2018). Intelligent Transportation

Systems - Interoperability. [online available:] https://www.its.dot.gov/research{_}areas/interoperability.htm, (last access: 2018-03-11). 29

University Health Network (2018). HAPI FHIR - The Open Source FHIR API for Java. [online available:] http://hapifhir.io/, (last access: 2018-01-12). 28, 47

Urbauer, P., Frohner, M., David, V., and Sauermann, S. (2018). Wearable Activity Trackers Supporting Elderly Living Independently: A Standards based Approach for Data Integration to Health Information Systems. In *Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion*, Thessaloniki. DSAI 2018. xxiii, 84

Urbauer, P., Frohner, M., Forjan, M., Pohn, B., Sauermann, S., and Mense, A. (2012). A Closer Look on Standards Based Personal Health Device Communication: A Resume over Four Years Implementing Telemonitoring Solutions. *European Journal for Biomedical Informatics*, 8(3):65–70. 27, 64

Urbauer, P., Herzog, J., Pohn, B., Forjan, M., and Sauermann, S. (2014). Certification programs for ehealth-status quo. In *Studies in Health Technology and Informatics*, volume 198, pages 164–171. xxi, 34

Urbauer, P., Kmenta, M., Frohner, M., Mense, A., and Sauermann, S. (2017). Propose of standards based IT architecture to enrich the value of allergy data by telemonitoring data. In *Studies in Health Technology and Informatics*, volume 236, pages 136–143. xxiv, 92

Urbauer, P., Sauermann, S., Frohner, M., Forjan, M., Pohn, B., and Mense, A. (2015). Applicability of IHE/Continua components for PHR systems: Learning from experiences. *Computers in Biology and Medicine*, 59:186–193. 27

U.S. Department of Health & Human Services (2018). HIPAA for Professionals — HHS.gov. [online available:] https://www.hhs.gov/hipaa/for-professionals/index.html, (last access: 2018-03-09). 31

USA.GOV (2018). The home of the U.S. Government's open data. [online available:] https://www.data.gov/, (last access: 2018-06-10). 56

Yang, C.-Y., Liu, C.-T., and Tseng, T.-W. (2015). Design and Implementation of a Privacy Aware Framework for Sharing Electronic Health Records. In *2015 International Conference on Healthcare Informatics*, pages 504–508. IEEE. 27

About the Author

Philipp Urbauer was awarded with the designation Engineer after absolving the HTL Ottakring in 2003 and obtained his B.Sc. and M.Sc. in Biomedical Engineering from the UAS Technikum Wien in Austria. Currently he is Program Director for the Short Cycle Study "Software Testing" and Deputy Program Director of the Bachelor Study Program "Transport and Environment". As a researcher he is founding member of the Research Focus "Secure Services, eHealth and Mobility" at the UAS Technikum Wien, working on the interdisciplinary topic of IT-systems interoperability. In this role, he participates in standardization work (IEEE 11073 - Personal Health Device - work group, IHE and HL7) and attended IHE Connect-a-thons and PCHA Plugfests as vendor and monitor. He took part in a leading role regarding the design and development of the Object Identifier (OID)-Repository and the Terminology-Server for the Austrian Healthcare System. Today he supports the systems-maintenance as a consulter and is responsible for the eHealth-Austria-registration authority. His thematic interests are interoperability of IT-systems, medical information systems, crowd-sourced healthcare, open data applications and software testing were he is furthermore certified as an ISTQB software tester.