Engineering Privacy for Mobile Health Data Collection Systems in the Primary Care

Mobile health (mHealth) systems empower Community Health Workers (CHWs) around the world, by supporting the provisioning of Community-Based Primary Health Care (CBPHC). In particular, Mobile Health Data Collection Systems (MDCSs) are used by CHWs to collect health-related data about the families that they treat, replacing paper-based approaches. Although MDCSs improve the efficiency of CBPHC, existing solutions lack adequate privacy and security safeguards.

To bridge this knowledge gap between the research areas of mHealth and privacy, we start by asking: How to design secure and privacy-preserving systems for Mobile Health Data Collection Systems? To answer this question, an engineering approach is chosen to analyse and design privacy and security mechanisms for MDCSs.

Among the main contributions, a comprehensive literature review of the Brazilian mHealth ecosystem is presented. On the privacy engineering side, the contributions are a Privacy Impact Assessment (PIA) for the GeoHealth MDCS and three mechanisms: SecourHealth, a security framework for data encryption and user authentication; an Ontology-based Data Sharing System (O-DSS) that provides obfuscation and anonymisation functions; and, an electronic consent (e-Consent) tool for obtaining and handling informed consent.
Engineering Privacy for Mobile Health Data Collection Systems in the Primary Care

Leonardo Horn Iwaya

Faculty of Health, Science and Technology
Computer Science

DOCTORAL THESIS | Karlstad University Studies | 2019:1
Engineering Privacy for Mobile Health Data Collection Systems in the Primary Care

Leonardo Horn Iwaya
Engineering Privacy for Mobile Health Data Collection Systems in the Primary Care
Leonardo Horn Iwaya

DOCTORAL THESIS

Karlstad University Studies | 2019:1

urn:nbn:se:kau:diva-70216

ISSN 1403-8099


© The author

Distribution:
Karlstad University
Faculty of Health, Science and Technology
Department of Mathematics and Computer Science
SE-651 88 Karlstad, Sweden
+46 54 700 10 00

Print: Universitetstryckeriet, Karlstad 2018

WWW.KAU.SE
Engineering Privacy for Mobile Health Data Collection Systems in the Primary Care

Leonardo Horn Iwaya
Department of Mathematics and Computer Science

Abstract

Mobile health (mHealth) systems empower Community Health Workers (CHWs) around the world, by supporting the provisioning of Community-Based Primary Health Care (CBPHC) – primary care outside the health facility into people’s homes. In particular, Mobile Health Data Collection Systems (MDCSs) are used by CHWs to collect health-related data about the families that they treat, replacing paper-based approaches for health surveys. Although MDCSs significantly improve the overall efficiency of CBPHC, existing and proposed solutions lack adequate privacy and security safeguards. In order to bridge this knowledge gap between the research areas of mHealth and privacy, the main research question of this thesis is: How to design secure and privacy-preserving systems for Mobile Health Data Collection Systems? To answer this question, the Design Method is chosen as an engineering approach to analyse and design privacy and security mechanisms for MDCSs. Among the main contributions, a comprehensive literature review of the Brazilian mHealth ecosystem is presented. This review led us to focus on MDCSs due to their impact on Brazil’s CBPHC, the Family Health Strategy programme. On the privacy engineering side, the contributions are a Privacy Impact Assessment (PIA) for the GeoHealth MDCS and three mechanisms: (a) SecourHealth, a security framework for data encryption and user authentication; (b) an Ontology-based Data Sharing System (O-DSS) that provides obfuscation and anonymisation functions; and, (c) an electronic consent (e-Consent) tool for obtaining and handling informed consent. Additionally, practical experience is shared about designing a MDCS, GeoHealth, and deploying it in a large-scale experimental study. In conclusion, the contributions of this thesis offer guidance to mHealth practitioners, encouraging them to adopt the principles of privacy by design and by default in their projects.

Keywords: Privacy, data protection, information security, mobile health, community-based primary health care, privacy impact assessment, consent management, anonymisation.
Acknowledgements

I take this opportunity to express my gratitude and appreciation to the amazing people that helped throughout my doctoral studies. Without your help this would never have been possible.

First and foremost, I would like to thanks my wonderful family for all the encouragement. Despite the geographical distance your love has always been deeply felt, making life so much easier and enjoyable. To my cherished parents and grandparents, thank you for supporting me in all my endeavors and thank you for being the best teachers that one could ever have – education is indeed freedom. To my beloved wife, Corinne, thank you for lifting my spirits and splashing colour in my life. I sincerely cannot thank you enough.

I am also grateful to my inspiring supervisors Simone Fischer-Hübner, Leonardo Martucci, and Rose-Mharie Åhlfeldt. Thank you so much for your attention and dedication during my entire PhD education at Karlstad University. You have helped me polish my skills and comprehend science at its highest level. Also, thank you for your understanding and genuine camaraderie. Your influence is not only positive to my scientific career but also to my personal development as an individual.

I would also like show my gratitude to collaborators in foreign institutions. Especially, thank you to professors Tereza Cristina Carvalho and Marcos Simplicio at University of São Paulo and former co-workers at the Laboratory of Computer Networks and Architecture. Also thanks to academics at University of Trento, particularly Ronald Chenu-Abente, Alethia Hume and Fausto Giunchiglia. And thank you to Jane Li from the Australian eHealth Research Centre (CSIRO). Having worked with so many brilliant researchers has truly enriched my academic experience.

Many thanks as well to all my friends, near and far, that always managed to keep in touch. Thanks to my fellow students and colleagues at Datavetenskap that create such an incredible working environment. In particular to my colleagues in the Privacy and Security research group, for sharing so many empathetic conversations about the PhD life. An enthusiastic thanks to the board games crew for the all the fun, food and laughter. Thanks to the rock climbing, innebandy and shorinji kempo teams for rescuing me from a desk-bound life. And I am also grateful for my Brazilian friends in my hometown that have always maintained contact and received me back with open arms as time went on.

This research was, besides Karlstad University, partly funded by:

- the Seventh Framework Programme for Research of the European Community in the project SmartSociety (under grant no. 600854);
- the Swedish Knowledge Foundation in the project High Quality Networked Services in a Mobile World (HITS);
- the Nordic Digital Health Center (NDHC) – a cooperative project between the county of Värmland, the municipality of Karlstad, Karlstad University and Stiftelsen Compare Karlstad;
• and, the DigitalWell Research – a partnership between Region Värmland and Karlstad University (Dnr RV2017-545).

Karlstad (Sweden), November 2018

Leonardo Horn Iwaya
List of Appended Papers


Comments on my Participation

**Paper I**  I am the main author and the proposer of the publication. I participated in the entire writing process but specially Sections “Mobile health in Brazil” and “Analysis and Discussion”.

**Paper II**  I participated in the entire writing process but specially in Section “Implementation”, i.e., SecourHealth’s implementation and testing.
Paper III  I participated in the entire writing process but specially in Subsection “GeoHealth’s security features”.

Paper IV  I am the main author and the proposer of the publication. I participated in the entire writing process and I am also the researcher who led the privacy impact assessment.

Paper V  I am the main author and the proposer of the publication. I participated in the entire writing process and I am also the researcher who led the privacy impact assessment.

Paper VI  I participated in the entire writing process but specially in Sections “Background and related work” and “Obfuscation and anonymisation for HIS”. I am also the main author of the publication.

Paper VII  I am the main author and the proposer of the publication. I participated in the entire writing process and I am also the researcher who led the design of the consent management system.

Other publications


Contents

List of Appended Papers vii

INTRODUCTORY SUMMARY 1

1 Introduction 3

2 Background 4
  2.1 Primary Health Care ................................. 5
  2.1.1 Community-Based Primary Health Care ................. 5
  2.1.2 Brazil’s Family Health Strategy ....................... 6
  2.2 Mobile Health ......................................... 9
       2.2.1 Mobile Health Data Collection Systems .. ......... 11
       2.2.2 eSUS and MDCSs in Brazil ...................... 12
  2.3 Privacy, Security and mHealth ......................... 14
       2.3.1 Privacy Law for Mobile Health ................... 16
       2.3.2 Legal Frameworks EU GDPR and BR LGPD .......... 17
  2.4 Privacy Engineering ..................................... 19
       2.4.1 Privacy by Design ............................... 19
       2.4.2 Privacy Impact Assessments .................... 20
       2.4.3 Privacy-Enhancing Technologies ................. 23
       2.4.4 Security and Privacy-Enhancing Mechanisms for MDCSs 25

3 Research Question 29

4 Research Method 30

5 Contributions 32
  5.1 Limitations ............................................ 34

6 Related Work 34
  6.1 Mobile Health for LMICs .............................. 34
  6.2 Security for MDCSs .................................... 35
  6.3 Design and Deployment of MDCSs ..................... 35
  6.4 Privacy Impact Assessment for MDCSs ................. 36
  6.5 Obfuscation and Anonymisation of Health Data .......... 37
  6.6 Consent Management Systems .......................... 37

7 Summary of Appended Papers 38

8 Conclusions and Future Work 41

Paper I
Mobile Health in Emerging Countries: A Survey of Research Initiatives in Brazil 55
1 Introduction 60
  1.1 Related work .................................................. 61
  1.2 Organization .................................................. 62

2 An overview of health care in Brazil 62

3 Key research areas in mHealth 63

4 Mobile health in Brazil 65
  4.1 Health surveys & surveillance ................................. 66
  4.2 Patient records .................................................. 69
  4.3 Patient monitoring .............................................. 71
  4.4 Decision support systems ..................................... 74
  4.5 Treatment compliance ......................................... 75
  4.6 Awareness raising ............................................. 76
  4.7 Summary ......................................................... 76

5 Analysis and discussion 76

6 Conclusions 82

PAPER II
SecourHealth: A Delay-Tolerant Security Framework for
Mobile Health Data Collection 94

1 Introduction 97

2 Scenario assumptions and requirements 99
  2.1 Tolerance to delays and lack of connectivity .............. 100
  2.2 Protection against device theft or loss .................... 100
  2.3 Secure data exchange between mobile device and server .. 101
  2.4 Lightweight and low cost solution ......................... 101
  2.5 Device sharing ................................................ 101
  2.6 Usability ....................................................... 102

3 The SecourHealth framework 102
  3.1 Preliminaries and notation .................................. 102
  3.2 User registration .............................................. 103
  3.3 Offline user authentication .................................. 104
  3.4 Secure Data Storage .......................................... 106
    3.4.1 No forward secrecy (\(K_{nofs}\)) ......................... 106
    3.4.2 Weak forward secrecy (\(K_{wfs}\)) ....................... 106
    3.4.3 Strong forward secrecy (\(K_{sfs}\)) ....................... 107
    3.4.4 Key generation and usage – summary .................. 108
  3.5 Data exchange with server .................................. 110
# Analysis

4.1 Tolerance to delays and lack of connectivity ........................................ 112  
4.2 Protection against device theft or loss .................................................. 112  
4.3 Secure data exchange between mobile device and server ......................... 113  
4.4 Lightweight and low cost solution ........................................................... 113  
4.5 Device sharing ......................................................................................... 113  
4.6 Usability ..................................................................................................... 113  

# Implementation

5.1 Platform characteristics ............................................................................ 115  
5.2 Cryptographic algorithms adopted ............................................................ 115  
5.3 Pilot application ........................................................................................ 115  
5.4 Benchmark results .................................................................................... 115  

# Related work

6.1 Future work ............................................................................................... 119  

## Paper III

Georeferenced and Secure Mobile Health System for Large Scale Data Collection in Primary Care  

1 Introduction .................................................................................................. 128  
2 Related projects in Brazil and worldwide ..................................................... 130  
3 Scenario description and requirements ........................................................ 131  
3.1 Data gathering in the family health strategy .............................................. 131  
3.2 Potential improvements to the data gathering process ............................. 132  
4 GeoHealth ..................................................................................................... 133  
4.1 GeoHealth-mobile ..................................................................................... 135  
4.2 GeoHealth-web .......................................................................................... 136  
4.3 Security features ......................................................................................... 137  
5 Status and discussion ................................................................................... 138  
5.1 System deployment .................................................................................... 138  
5.2 Continuous training .................................................................................. 139  
5.3 Improved efficiency and data quality ....................................................... 140  
5.4 Use of collected data .................................................................................. 141  
5.5 Cost of the solution .................................................................................... 141  
5.6 Comparison to other systems .................................................................... 142  

6 Conclusions .................................................................................................... 143
### Paper IV

**mHealth: A Privacy Threat Analysis for Public Health Surveillance Systems**

1. Introduction 151
2. Related work 152
3. Methodology 153
4. System characterisation - GeoHealth for public health surveillance 154
5. Privacy Threat Analysis 155
   - 5.1 Privacy Targets 155
   - 5.2 Identification of Privacy Threats 155
     - 5.2.1 Threats to data quality 155
     - 5.2.2 Threats to processing legitimacy & informed consent 157
     - 5.2.3 Threats to data subject’s information right 157
     - 5.2.4 Threats to data subject’s right to access 158
     - 5.2.5 Threats to data subject’s right to object 158
     - 5.2.6 Threats to data security 159
     - 5.2.7 Threats to accountability 159
6. Results and discussion 160
7. Conclusion 162

### Paper V

**Mobile Health Systems for Community-Based Primary Care: Identifying Controls and Mitigating Privacy Threats**

1. Introduction 168
2. Prior Work & Background 169
   - 2.1 MDCSs Worldwide and in Brazil 170
   - 2.2 PIA Frameworks 170
   - 2.3 Security & Privacy of MDCSs 171
3. Methods 171
4. Results 172
   - 4.1 Characterisation of the System 173
   - 4.2 Definition of Privacy Targets 173
   - 4.3 Evaluation of Degree of Protection Demand for each Privacy Target 174
   - 4.4 Identification of Threats for each Privacy Target 175
4.5 Identification and Recommendation of Controls to Protect Against Threats ........................................ 177

5 Discussion ..................................................... 177
5.1 Transparency and Intervenability .......................... 178
5.2 Informed Consent .......................................... 180
5.3 Data Objection and Deletion ............................ 180
5.4 Security Mechanisms ...................................... 181
5.5 Anonymisation ............................................. 181

6 Limitations .................................................. 182

7 Conclusion .................................................. 183

Appendices .................................................... 187

A Characterisation of the system .............................. 187
A.1 System View ............................................... 188
  A.1.1 GeoHealth-Mobile .................................. 188
  A.1.2 GeoHealth-Web .................................... 188
A.2 Functional View .......................................... 188
  A.2.1 Overall Goals and Business Process ............... 188
  A.2.2 GeoHealth’s Detailed Use Case .................... 189
A.3 Data View ................................................ 192
A.4 Physical Environment View ............................. 192

B Evaluation of degree of protection demand ............. 193
B.1 P1 - Quality of Data Processing ........................ 194
B.2 P2 - Processing Lawfulness and Informed Consent .... 197
B.3 P3 - Information right of data subject (ex ante Transparency) ........................................ 198
B.4 P4 - Access right of data subject (ex post Transparency) ........................................ 199
B.5 P5 - Intervenability ....................................... 200
B.6 P6 - Data subject’s right to object ..................... 201
B.7 P7 - Security of data .................................... 203
B.8 P8 - Accountability ..................................... 204

C Identification of privacy threats ............................ 204
C.1 Threats to data quality .................................... 206
C.2 Threats to processing legitimacy & informed consent .. 208
C.3 Threats to data subject’s information right ............ 208
C.4 Threats to data subject’s right to access ............... 209
C.5 Threats to intervenability ............................... 209
C.6 Threats to data subject’s right to object ............... 209
C.7 Threats to data security .................................. 210
C.8 Threats to accountability ............................... 211
  C.8.1 Threats to notification/reporting duties ............ 211
D Technical and non-technical controls 212
  D.1 Control descriptions 212

PAPER VI
Ontology-based Obfuscation and Anonymisation for Privacy: A Case Study on Healthcare 218

1 Introduction 222

2 Data protection regulations and legislation 223

3 Background and related work 224
  3.1 EHR’s data elements 224
  3.2 Conventional anonymisation methods 225
  3.3 Ontology-based approaches 226
    3.3.1 Access control and context obfuscation 226
    3.3.2 Ontology-Based Anonymisation 227
  3.4 Ontology-based Identity Management and Access Control 228
  3.5 Putting things together 228

4 Obfuscation and anonymisation for HIS 228
  4.1 Ontology-based Data Sharing Service 229
    4.1.1 UC1: Privacy-preserving patient treatment 230
    4.1.2 UC2: Patient’s granular control of E/PHR 230
    4.1.3 UC3: Reminder or alert systems 230
    4.1.4 UC4: Medical research repository 230
    4.1.5 UC5: Nationwide HIS network 230
  4.2 SO and DA functions 231
    4.2.1 Primary use and semantic obfuscation 232
    4.2.2 Secondary use and data anonymisation 233

5 Future work 234

6 Conclusions 235

PAPER VII
E-Consent for Data Privacy: Consent Management for Mobile Health Technologies in Public Health Surveys and Disease Surveillance 238

1 Introduction 241

2 Related work 243

3 Methods 243
4 Results 244
  4.1 Legal and technical requirements 244
  4.2 Design considerations 245
  4.3 Consent interface 247
  4.4 Generating Consent Receipts 248
  4.5 Usability evaluation 248

5 Discussions 250

6 Conclusions 251
“Feliz é aquele que pode recordar com saudade o passado, mas infeliz, com certeza, é quem tem medo de lembrá-lo.”

— A Saga de um Imigrante Japonês. (1985)
S. Shumin

“Happy is the one who can longingly recall the past, but unhappy, surely, is the one who is afraid to remember it.”
1 Introduction

Mobile phones are seemingly ubiquitous in today’s society. Virtually everyone has a mobile phone, no matter if they live in high-income countries or in low- and middle-income countries (LMICs). As a matter of fact, it has been reported that mobile phones are already more accessible than basic sanitation [77]. Throughout the developing world they are used more than any other kind of modern technology [58]. Not surprisingly, many industries, governments and researchers are creating new business models based on the power of mobile emerging markets, a phenomena referred as Mobile for Development (M4D) [47]. Examples are mobile-based systems for financial services, agriculture, digital identity, and health care.

Mobile health (or mHealth) technologies – the use of mobile devices to support the delivery of health care – leverage from this ubiquity and show great potential to enhance public health and clinical practice in LMICs. In this context, hundreds of mHealth projects have already been developed, targeting issues such as HIV, malaria, tuberculosis (TB), diabetes and antenatal care [3]. Existing solutions provide a range of innovative applications, e.g. for patient follow-ups, staff training, drug supply chain, patient education, disease surveillance, data collection and reporting.

On the one hand, mHealth systems make communication among health care providers, researchers and patients easier and offer great promise for improving quality of life [58]. On the other hand, the increasing collection and disclosure of health-related data raise various privacy issues. Several concerns exist on privacy and security of mHealth applications, due to the fact the collected data often reveal highly sensitive personal information, such as social interactions, location, emotion, and other health conditions [62]. This is particularly alarming for nationwide mHealth systems for health surveys and disease surveillance, used to support public health care.

For instance, a case was reported in Haiti in which the government requested access to electronic medical records of HIV positive individuals [115]. The purpose was to build a national database to calculate HIV prevalence in the country. Although many organisations complied with the government’s request, others were uncomfortable disclosing such types of data without patient’s consent. In this kind of situation, the potential absence of privacy laws and the fact that often such systems are being used to treat vulnerable populations just worsens the situation. Furthermore, many mHealth systems today process a much wider range of health conditions than just HIV.

That brings us to the category of mHealth applications known as Mobile Health Data Collection Systems (MDCSs), which are the main focus of this thesis. In brief, such systems usually support Community Health Workers (CHWs) in primary care activities [59]. As part of their job, they visit families at their homes to provide basic health and medical care. However, they are also responsible for conducting health surveys and reporting various health indicators back to their managers at district- or national-level systems. Health surveys are used to collect data about the families, including information about hous-
ing characteristics, sanitary conditions, chronic diseases, and illnesses. Data collection and reporting that was once done with pen and paper is now being replaced by electronic forms.

Ensuring privacy and security for these systems is challenging [59]. Project managers may face barriers due to the lack of understanding about the relevant privacy laws. At the same time, developers may lack practical experience on translating privacy principles from legal frameworks to actual privacy-preserving mechanisms. In any case, concerns with privacy and security severely hinder projects from being fully deployed and scaling-up.

Taking this into account, in this thesis an engineering approach is adopted to bridge the knowledge gap between mHealth practitioners and privacy specialists. Departing from the research question, “How to design secure and privacy-preserving systems for Mobile Health Data Collection Systems?”, this thesis offers a series of analyses and novel mechanisms. A review of the Brazilian mHealth ecosystem is presented, highlighting the importance of CHWs and existing solutions. The review also provides insights about mHealth market, research, players and maturity of initiatives. In addition, practical experience with the implementation and deployment of a large-scale MDCs is also shared. First-hand involvement with the CHWs, health managers and developers is essential for designing realistic solutions. Most importantly, privacy and security analyses for MDCs were carried out, allowing the proposal of different mechanisms. Specifically, security mechanisms for authentication, access control and storage, as well as strategies for obfuscation and anonymisation of data, and management of informed consent.

The remainder of this Introductory Summary is structured as follows. Section 2 provides the research background, defining terms, concepts and technologies covered in the thesis. Section 3 introduces the thesis research question. Section 4 describes the research method that was predominantly adopted in this work. Section 5 states the scientific contributions achieved with this research. Section 6 discusses the relevant related work. Section 7 contains a list of the appended publications. Section 8 presents the research conclusions and directions for future work.

2 Background

Privacy and health informatics are by nature multidisciplinary research areas. This section introduces the main concepts and terms used throughout the thesis. First, the health-related concepts are presented. Primary Health Care (PHC) and Community-Based Primary Health Care (CBPHC) are defined, laying the foundation to introduce Brazil’s Family Health Strategy (FHS) programme. Second, the mHealth technology used to support CBPHC activities is explained, covering mHealth initiatives carried out around the world and in Brazil more specifically. And finally, we introduce the privacy-related concepts, relevant legal frameworks, and approaches to engineering privacy in MDCs.
2.1 Primary Health Care

Primary Health Care (PHC) is a model for delivering essential health care services, introduced in the declaration of the International Conference on Primary Health Care held in Alma-Ata, Kazakhstan in 1978. This declaration urged from multiple sectors of society urgent action to protect and promote the health of all people, underlining the importance of PHC. Article VI of the 1978 Declaration of Alma-Ata defines PHC as follows [83]:

“Primary health care is essential health care based on practical, scientifically sound and socially acceptable methods and technology made universally accessible to individuals and families in the community through their full participation and at a cost that the community and country can afford to maintain at every stage of their development in the spirit of self-reliance and self-determination. It forms an integral part both of the country’s health system, of which it is the central function and main focus, and of the overall social and economic development of the community. It is the first level of contact of individuals, the family and community with the national health system bringing health care as close as possible to where people live and work, and constitutes the first element of a continuing health care process.”

This definition carries in itself the notion of equity in the provision of health care, i.e. universally accessible. It also stresses the community involvement, their full participation and responsibility towards improving health outcomes. And emphasises that PHC is not only related to the health system, but also to other areas that play a role in health, such as social and economic development. The Alma-Ata declaration defined PHC forty years ago but it is still relevant to today’s health care systems [43], capturing an ideal to strive for.

2.1.1 Community-Based Primary Health Care

Community-Based Primary Health Care (CBPHC) is a strategy for PHC that leverages from the community involvement. It is the joint work of community members, local leaders, public and private organisations to extend health services beyond health facilities into the communities and their homes. More recently, experts have agreed on the following definition of CBPHC [89]:

“CBPHC is a process through which health programs and communities work together to improve health and control disease. CBPHC includes the promotion of key behaviors at the household level as well as the provision of health care and health services outside of health facilities at the community level. CBPHC can (and of course should) connect to existing health services, health programs, and health care provided at static facilities (including health centers and hospitals) and be closely integrated with them.”
In [89], the authors also stress that CBPHC includes multi-sectoral approaches to health improvement – beyond the health services *per se* – with positive impact on education, income, nutrition, living standards, and empowerment. That is, other kinds of programmes that positively correlate with health outcomes. Besides, they also identify three different types of CBPHC interventions [89]: (1) health communication, i.e. disseminating health information to individuals, families and communities; (2) social mobilisation and community involvement for planning, delivering, evaluating and using health services; and, (3) provision of health care in the community (e.g. preventive or curative care).

Interventions are put into operation by various primary care workers and community members, but most remarkable is the role of Community Health Workers (CHWs). CHWs are known by several names in different countries and their working conditions also vary in terms of performed activities, wages and received training [16]. A widely accepted definition for CHWs was proposed by the World Health Organization (WHO) [84]:

“Community health workers should be members of the communities where they work, should be selected by the communities, should be answerable to the communities for their activities, should be supported by the health system but not necessarily a part of its organisation, and have shorter training than professional workers.”

CHWs serve as an entry point into the formal health system [20]. In rural and under-served communities they are most likely the only link between the population and the health system [103]. As part of primary care teams, they help people navigate the maze of health services, helping with referrals, health promotion, community engagement and requesting specialised services [85] (see Figure 1).

CHWs visit families at their homes in order to provide primary care. They keep track of these visitations reporting back to their supervisors and district or national-level systems. Notwithstanding, they also conduct demographic and health surveys helping to monitor various health indicators of the population. Taking advantage of digitisation, many paper-based approaches for health surveys and surveillance are now being replaced by computer applications, running on CHWs’ mobile phones. Thus, empowering CHWs and enhancing the quality and efficiency of the entire CBPHC programmes.

Low- and middle-income countries (LMICs) greatly benefit from CBPHC programs and the work performed by CHWs. In this thesis we further explore the CBPHC programme implemented in Brazil.

### 2.1.2 Brazil’s Family Health Strategy

The history of the Brazilian Family Health Strategy (FHS)\(^1\) started in 1991 when the Ministry of Health established the *Programa de Agentes Comunitários*

\(^1\)formerly know as Family Health Programme
de Saúde (PACS, Programme of Community Health Agents\(^2\)). The objective of PACS was the reduction of infant and maternal mortality, especially in the Brazilian North and Northeast regions, extending the coverage of health services to the poorest and most disadvantaged areas [97, 110]. Based on the successful experience with the PACS in the state of Ceará the FHS was conceived as a federal programme in 1993 from a meeting on “Family Health”, convened by the Brazilian Health Minister with the support of United Nations Children’s Fund (UNICEF) [97]. This was the beginning of the Brazilian approach for CBPHC.

The FHS aims at providing preventive and basic health care. To do so, Primary Care Teams (PCTs) are organised, usually consisting of a physician, a nurse, and about six CHWs – and sometimes supported by a dental and oral health team [111] (see Figure 2). Every four or five PCTs also receive support of other health professionals (e.g. psychologists, pharmacists, physiotherapists) for specialised care [111]. Each PCT covers three to four thousand families, with maximum 150 families per CHW. Recent numbers show that Brazil has over 256,000 CHWs delivering primary care to 62 percent of the population (see Table 1).

Besides providing primary care, CHWs are in charge of registering all families in their catchment area, monitoring living conditions (e.g. housing characteristics and sanitation) and health status (e.g. chronic patients, pregnant women, children’s vaccination). They also resolve low-level problems, such as checking patient’s medication adherence or dealing with patient referrals to specialised health services. Enrolled families receive monthly visits from a dedicated CHWs regardless of need [111]. Concomitantly, they collect demo-

\(^2\)Community Health Agents (CHAs) is the most common denomination for CHWs in Brazil.
Figure 2: Family Health Strategy organisation (adapted from [111]).
graphic and health data about the families, that are later digitised at the health facility, and transmitted to the Sistema de Informação em Saúde para a Atenção Básica (SISAB, Health Information System for Primary Care). The data is used to assist the PCTs’ ongoing work in the community, as well as to generate performance metrics and to support activities of public health surveillance, usually conducted by health managers at a district- or national-level.

In what follows, the technology used to support the work of CHWs in the scope of CBPHC is reviewed. Particularly, mHealth initiatives that improve the data collection in PHC.

2.2 Mobile Health

Today, mobile devices are ubiquitous. Profound technologies that “weave[d] themselves into the fabric of everyday life until they are indistinguishable from it” [112]. According to the GSM Association, the number of connections (excluding cellular IoT) totalled 7.8 billion globally in 2017 and will reach 9.0 billion by 2025 [46]. Smartphones represent 57 percent of the total mobile connections and by 2025 this number will increase to 77 percent. Even more stunning, for a few years already, reports have shown that more people have access to mobile phones than basic sanitation such as toilets [77]. Considering such impressive scalability, it is no surprise that mobile devices and smartphones would play a significant role in the field of health information systems.

Mobile health (mHealth) results from the composition of a range of technologies, such as mobile computing, medical sensors, and portable devices to ensure health care [55]. Also relying on wireless networking capabilities and the increasing miniaturisation of computers. A definition of mHealth was given by [61]:

“[...] mobile health, or mHealth, refers to the use of mobile technologies – wearable, implantable, environmental, or portable – by individuals who monitor or manage their own health, perhaps with the assistance of individual caregivers or provider organizations. The technology might support clinical care – including diagnosis and disease management – or wellness goals such as losing weight, eating a healthy diet, quitting smoking, or becoming physically fit.” [61].

Existing mHealth systems can be used by nurses, doctors, caregivers, or

<table>
<thead>
<tr>
<th>1998</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,000 PCTs</td>
<td>39,000 PCTs</td>
</tr>
<tr>
<td>60,000 CHWs</td>
<td>256,000 CHWs</td>
</tr>
<tr>
<td>-</td>
<td>30,000 oral health teams</td>
</tr>
<tr>
<td>7,000,000 people (4%)</td>
<td>120,000,000 people (62%)</td>
</tr>
</tbody>
</table>

Table 1: FHS’s evolution from 1998 to 2014 (numbers from[66]).
personal trainers putting the necessary information about their clients at point of their fingertips. Patients can use mHealth to better manage their medical conditions and communicating with health care providers. A wide range of mHealth apps for planning diets and exercises exist, tailored for fitness and well-being. At the same time, mHealth technologies are also used to assist various public health activities (e.g. disease surveillance, health census and alert systems).

Countries of all income levels have already deployed mHealth systems to some extent; but systems can be fairly different depending on where you live. High-income countries take advantage of more sophisticated settings, enabling teleconsultation and remote monitoring of patients from the comfort of their homes [86]. On the other hand, low- and middle-income countries (LMICs) take benefit from the rapid growth of the mobile market and network infrastructure. Many initiatives use mobile’s SMS systems for health campaigns (raising health awareness), treatment adherence, follow-ups, and reminders to assure medication compliance [21, 86]. Also, mHealth systems are often used to support frontline health workers, including CHWs, nurses and midwives in the promotion of primary health care [21, 74, 103]. The WHO performed a global survey in an attempt to review existing mHealth initiatives and their levels of adoption within all its member states, revealing a wide range of mHealth applications categories [86] (see Figure 3).
In this thesis we are particularly interested in mHealth systems for CBPHC used for public and medical health. These systems usually fall into a joint category of “Health Surveys & Surveillance”, since health surveillance depends on data collection in the first place.

### 2.2.1 Mobile Health Data Collection Systems

Public health surveillance is the continuous, systematic collection, analysis and interpretation of health-related data needed for the planning, implementation, and evaluation of public health practice – to reduce morbidity and mortality and to improve health [42]. In public health\(^3\), surveillance is necessary to [87]:

(a) serve as early warning systems to impede public health emergencies;
(b) document the impact of an intervention, or track progress towards specified goals; and,
(c) monitor and clarify the epidemiology of health problems, to allow priorities to be set and to inform public health policy and strategies.

In other words, although surveillance is a rather loaded term in the privacy community, there would be no public health without surveillance systems.

Health surveys are the active approach for public health surveillance. Large field surveys are common practice and offer a reliable way to collect data in LMICs [108]. Paper-based data collection are still the standard method in many countries, but they are being replaced by mobile applications with electronic forms, improving the overall efficiency of the process as well as data quality and responsiveness. The category of mHealth applications for health surveys is known as Mobile Health Data Collection Systems (MDCSs) [41].

Various frontline health workers benefit from using MDCSs as part of their daily activities of data collection and reporting. In brief, during the home visitations, they can use smartphones to collect data about the families, transmit it to the health facility through the network (e.g. 3G or 4G), and keep a history of their visitations as well as allowing health managers to visualise in real-time health reports of the community (as illustrated in Figure 1).

Across the globe many countries are starting to use MDCSs to replace the archaic paper forms (for a survey, see [101]). Some remarkable examples of MDCSs are:

- **Fiji** – the eSTEPS [122] for comparing the data quality provided by paper-based and electronic health questionnaires.

- **South Africa** – the Personal Data Collection Toolkit (PDAct) for interviewer-administered and respondent-administered data collection [100] and the Mobile Researcher [108] to support CHWs.

---

\(^3\)“Public health is the science and art of preventing disease, prolonging life and promoting physical health and efficiency through organized community efforts for the sanitation of the environment, the control of community infections, the education of the individual in principles of personal hygiene, the organization of medical and nursing service for the early diagnosis and preventive treatment of disease, and the development of the social machinery which will ensure to every individual in the community a standard of living adequate for the maintenance of health.” – Charles-Edward Amory Winslow (1920) [116].
Figure 4: Conventional process and data flows for MDCSs (adapted from [56]).

- Tanzania – the Data Entry at the Point of Collection (DEAPOC) for large household survey in remote areas [102].

- Thailand – a mHealth application to improve antenatal care and expanded programme on immunisation services for the under-served population [57].

- Ghana – the Mobile Technology for Community Health (MOTEC) Platform [44, 63], used by nurses and CHWs for recording and tracking the care delivered to women and newborns, and generating reports mandated by the country’s health authorities.

There are also standardised, general purpose tools that help in the task of designing forms and sending them to mobile devices, such as the Magpi framework [67] and the Open Data Kit [2]. And more recently, the World Health Organization (WHO) together with group of academic and research institutions and technology partners are developing the Open Smart Register Platform (OpenSRP [82]), which has been used to empower frontline health workers to electronically register and track the health of their entire client population.

2.2.2 eSUS and MDCSs in Brazil

Similarly, many MDCSs are being developed and deployed in Brazil. And given the importance of the FHS for CBPHC, it is natural that various MDCSs focus on CHWs and the data gathering for the Brazil’s national level Health Information System for Primary Care (SISAB) database. The FHS is one of the most important programmes of the Brazilian universal public health system, Sistema Único de Saúde (SUS). In the past, MDCSs were mainly developed by research groups within universities and health care institutions. Some examples of SISAB-oriented MDCSs are:
• [12]: an early architecture that implemented electronic forms used by CHWs in the FHS for handhelds computers, such as Palm Top and Hibook Reader.

• Colibri Project [91]: a set of tools for data collection, public health surveillance and expert systems for monitoring communities within the scope of the FHS.

• LISA-MCP [37]: a mHealth module that implemented one of the forms used by CHWs in the FHS (i.e. family registration) – as part of a larger project called LARIISA, a intelligent system for decision-making in public health governance.

• GeoHealth [98]: a system that allowed data collection and analysis in the scope of the FHS, with security and geo-referencing functionalities, deployed and tested in a large-scale study.

In the past years, the Brazilian Ministry of Health put forward a new strategy and created the electronic SUS (e-SUS). This nation-wide project aims to develop, restructure and integrate information systems in order to enable individualised electronic health records of the population. As part of this new strategy, we have e-SUS Atenção Básica (e-SUS AB, e-SUS Primary Care) to restructure the information systems used for primary care at the national level. To do so, the Department of Primary Care (DAB) released two main system modules: (1) the Prontuário Eletrônico do Cidadão (PEC, Citizen’s Health Record) electronic health record for the health facilities; and, (2) the Coleta de Dados Simplificada (CDS, Simplified Data Collection), the electronic forms used by CHWs in the primary care. These systems can be integrated, but in this thesis more attention is given to the CDS module.

The CDS module runs in a computer at the health facility, so that CHWs can digitise the paper forms or synchronise data directly from mobile phones. Health facilities can keep a local server – if they can afford it – with all the collected data. Every month they have to send the data to the SISAB, following specific standards for exporting data.

More recently, the DAB itself has released its first MDCS known as e-SUS AB Território, developed in collaboration with Bridge Laboratory [11]. Other systems, however, made by private companies were also being used across the country, e.g. eSUS+ ACS [107] and ACS Lite eSUS AB ePHealth [32]. Although the Ministry of Health now provides its own system, the decision of choosing and implementing a MDCS is still made at the municipality level. Many cities already had an information system in place, so they are not required to change it. In any case, the government provides a data standard for transmitting data to the national-level database, so that, the municipalities still have to upload monthly all the collected data – regardless if it is collected by paper or electronic forms.

Now that the main concepts related to health care and mHealth systems have been introduced, we proceed with the background on privacy and security, for mHealth in general as well as for MDCSs in specific.
2.3 Privacy, Security and mHealth

In today’s digital society where personal data is the new oil – a commodity in the digital economy – the concept of privacy grows in importance as far as our digital footprint becomes more unique and traceable, impacting our “real” life. Privacy remains however as an elusive concept virtually impossible to be defined [104]. There are multiple forms of privacy and they are perceived differently according to cultural, legal, and personal values. In an effort to steer away from the intricate philosophical discussion yet providing a workable explanation for privacy, one can refer to privacy in two simple ways. Privacy as the ability to seclude oneself physically; or, privacy as ability to have control over your own personal information. Broadly speaking, that means that privacy has at least two categories: (i) physical privacy, and (ii) informational privacy. In this thesis we are particularly focused on the latter, i.e. information privacy (also commonly referred as data privacy). A definition for privacy on such terms was provided by Westin (1967) [113]:

“Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others. Viewed in terms of the relation of the individual to social participation, privacy is the voluntary and temporary withdrawal of a person from the general society through physical or psychological means, either in a state of solitude or small-group intimacy or, when among larger groups, in a condition of anonymity or reserve.”

Privacy, in this thesis, considering Westin’s definition [113], is restricted to the privacy of individuals (natural persons); and information privacy refers to their personal data. Closely related to privacy, yet not equivalent, is the concept of security. As part of today’s information and communication technology, computer security is the discipline of protection of information systems (i.e. hardware, software and information) from theft, damage, disruption or misdirection. Usually computer security boils down to three core concepts [9]: (1) confidentiality, the concealment of information or resources; (2) integrity, the trustworthiness of data resources; prevention of improper or unauthorised change; and, (3) availability, the ability to use the information or resource desired. Consequently, it is natural to see an overlap between the concepts of information privacy and security, specially with regards to confidentiality. In practical terms, confidentiality refers to the authorised access or disclosure of information, comprising notions of secrecy, access-control, sharing, protection of information. Therefore, when it comes to information about a person, information security becomes one of the means for achieving information privacy. For example, individuals expect that their emails should not be read by others than the recipients, that is a privacy claim. Encryption is the security technology used as a mean to achieve such privacy goal. Nonetheless, some privacy concepts go beyond security, such as openness and transparency of personal data processing; or even, the right to access and redact one’s information in a system. Another classical example is the protection of meta communic-
ation data. Usually data confidentiality only concerns the actual content of data, but not necessarily the meta data about who is communicating with whom from where. For such reasons, it is more accurate to say that security is essential for achieving privacy, but only addressing the security issues is not enough.

Privacy can be also defined as a collective value and a human right. As a social value, privacy is essential for a functioning democracy; privacy and democracy reinforce each other [95]. Ensuring people’s dignity and autonomy is necessary for meaningful democratic participation. Privacy is also explicitly mentioned in Universal Declaration of Human Rights (Art. 12) [4], “No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation”. Privacy also has precedent as freedom of speech (and to hold back information), (Art. 19) [4]: “Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers”. Individuals under extensive scrutiny have their rights and freedoms limited, harming democratic values of the society. Privacy is essential element for societal development and therefore key in any endeavours of digital development as well. To truly bridge the digital divide, privacy cannot be seen as an optional feature, but rather as a fundamental part of the design process. And that is what we refer in this thesis as privacy for development, i.e. privacy-by-design and by default for true digital development.

Furthermore, privacy and data protection are sine qua non for high quality health care. The importance of privacy, as mentioned in [22], has been already manifested ages ago in one of the most widely known of Greek medical texts, the Hippocratic Oath:

“What I may see or hear in the course of the treatment or even outside of the treatment in regard to the life of men, which on no account one must spread abroad, I will keep to myself, holding such things shameful to be spoken about.” – excerpt from the Hippocratic Oath [31].

High quality health care requires individuals to share their personal health information with health care professionals [22]. Furthermore, information should be complete and accurate. If patients cannot trust that their information will be kept secure, they will be reluctant to share it (or even to use the service). If health professionals cannot trust the organisation to keep records secure they will not put complete information. In both cases this leads to inferior health care. It is therefore paramount that privacy and security concerns are addressed during the design of any health information system.

Having stressed the importance of privacy in the context of health care, we now expand this discussion with more specific guidelines and legal reports that address privacy for mHealth systems in LMICs.
2.3.1 Privacy Law for Mobile Health

The right to privacy, as already mentioned, is enshrined in various existing laws and even in the declaration of human rights. Such documents and other relevant regulations and standards are the starting point for project leaders and developers to understand the requirements on privacy and data protection. Some countries (i.e. separate legal jurisdictions), however, do not have a specific legislation for data privacy [45, 6]. This usually does not imply a legal void in the area. Privacy rights might stem from the constitution or consumer rights; and in the case of health care, from medical codes of conduct, and so on. Nonetheless, in such countries, project managers can benefit from some specific publications, helping them to understand privacy in the context of mHealth. For instance, [52] presents a list of five guiding principles for mobile privacy, recommended for projects in LMICs:

**Principle 1** Address Surveillance Risks – Projects should take steps to ensure that user data is secure from third party surveillance, e.g. user discriminatory profiling can be made by mobile operators and government.

**Principle 2** Limit Data Collection and Use – Projects should limit data collection to what is absolutely necessary for the project’s goal, e.g. by employing access control, data retention policy, and not collect unnecessary data.

**Principle 3** Promote and Facilitate Transparency – Projects should be transparent about what data is collected, how it is shared, and how it might be used in the future, e.g. user notifications, data transfer policies, audit trails of others that also have access to the data.

**Principle 4** Incorporate User Feedback – Projects should give users the ability to access, amend, and/or delete their data, e.g. create user interfaces, create communication channels to receive feedback from users.

**Principle 5** Assume Responsibility – Projects should assume accountability for potential risks and harms incurred via their projects and platforms, e.g. perform risk assessment, plan incident response, notify data breaches.

These guiding principles [52] offer a good starting point for developers and project leaders working with mHealth (specifically). However, in practice, privacy and security still require a case-by-case evaluation. There is no one-size-fits-all guideline, given the complexity, multiplicity of actors, jurisdictions, and highly culture-specific dimensions of privacy [75]. The ideal case, however, would be to strive for excellence by following regulations such as the EU GDPR, that sets a high standard for privacy. Moreover, many countries in the EU have also specialised national health privacy laws which however need to comply with the GDPR.
2.3.2 Legal Frameworks EU GDPR and BR LGPD

Guiding publications and other reports ([52, 75]), offer helpful advice to project leaders and developers working with mHealth in LMICs. However, these guidelines should be used – if possible – in combination with national legal frameworks that are relevant to your mHealth ecosystem. Today, many countries already have a overarching privacy law in force or at least there is a bill in discussion [6].

A well-know example is the European General Data Protection Regulation (EU GDPR). The EU GDPR replaces the previous Directive 95/46/EC [1] and was designed to: (a) harmonise privacy and data protection laws across Europe; (b) protect and empower all EU citizens privacy and data protection; and, (c) reshape the way organizations across the region approach privacy and data protection. However, it does not apply only to EU member states but also to organisations (i.e. data controllers and data processors) outside EU that offer goods and services to, or that monitor, individuals in the EU. For this reason, the EU GDPR casts its net globally and many other countries are updating their privacy laws so that an adequate level of data security and protection is guaranteed. Thus, allowing cross-border data transfers for business activities, and avoiding penalties and administrative fines4. The EU GDPR is considered the state-of-the-art on privacy law and it impacts organisations all around the world.

The Brazilian Lei Geral de Proteção de Dados (BR LGPD, General Law for Data Protection) is inspired in the EU GDPR. As of August 2018, Brazil

---

4“For especially severe violations, listed in Art. 83(5) GDPR, the fine framework can be up to 20 million euros, or in the case of an undertaking, up to 4% of their total global turnover of the preceding fiscal year, whichever is higher.” (Art. 83 [34])
has seen its bill approved in the senate and sanctioned by the president [10]. Organisations have 18 months to adapt to the new law. However, one of the major differences is that the president vetoed the creation of an Autoridade Nacional de Proteção de Dados (ANPD, National Data Protection Authority), a pillar for the implementation of the new policy. ANPD would be responsible for ensuring compliance with the law, that is, who would supervise and apply sanctions if the processing of personal data by some entity was not in accordance with the legislation. Thus, now it is expected that the executive (government) – instead of the senate – will send a new bill, specific for the creation of the ANPD.

With that being said, this section does not intend to provide a lengthy explanation about the EU GDPR, but rather cover just the main concepts that will be part of “privacy vocabulary” in this thesis. These concepts are used in both EU GDPR and BR LGPD. For the purpose of the regulations, they are defined as follows (Art. 4 GDPR [34]):

- **“personal data”** means any information relating to an identified or identifiable natural person (“data subject”); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

- **“processing”** means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;

- **“controller”** means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law;

- **“processor”** means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller;

- **“consent”** of the data subject means any freely given, specific, informed and unambiguous indication of the data subject’s wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her;

- **“data concerning health”** means personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status;
• “special categories” means personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person’s sex life or sexual orientation.

Maintaining these concepts in mind help to understand the vocabulary used in the following sections as well as the publications that are part of this thesis. Other explanations are provided throughout the document, but we refer the reader to the original texts for further details [34, 10].

2.4 Privacy Engineering

Guidelines, principles and policies help in understanding what should be done, but often it is not clear how to actually do it. High-level policies may even contain some degree of vagueness, which allows for legal interpretation – a positive feature that strengthens the policy enforcement. Hence, it is not the existing policies that will concretely explain how to address privacy and data protection in mHealth projects. The translation from policies to engineering methods, techniques and tools is a task for interdisciplinary groups. Just like “quality”, “safety”, and “environmental” regulations that needed to be put into practice over the past decades, now the same is happening to privacy.

Privacy engineering is therefore this emerging field that aims to close the gap between existing policies and the current practice, systematising and evaluating approaches to address privacy in the engineering of information systems [49]. In computer science, the area of software engineering is the one concerned with all aspects of software development process, from start to finish: requirements, design, implementation, verification, maintenance, and eventual discard. Thus, it is natural that the field of privacy engineering also borrows from software engineering. A definition for privacy engineering was proposed by Güseres et al (2016) [49]:

“[…] privacy engineering as the field of research and practice that designs, implements, adapts, and evaluates theories, methods, techniques, and tools to systematically capture and address privacy issues when developing sociotechnical systems.”

This section presents three themes on privacy engineering, that are repeatedly discussed throughout the thesis. Starting with the seminal idea of Privacy by Design (PbD), followed by the methodologies for Privacy Impact Assessments (PIAs) and some relevant Privacy-Enhancing Technologies (PETS).

2.4.1 Privacy by Design

Privacy by Design (PbD) is an approach geared towards systems engineering, urging that privacy should be taken into account throughout the entire engineering process. Some refer to PbD as a technical approach to a social problem
The concept was originally proposed by Ann Cavoukian during the 90’s [13], and later formulated “as a holistic concept that may be applied to operations throughout an organization, end-to-end, including its information technology, business practices, processes, physical design and networked infrastructure” (in the “Resolution on PbD” [106]).

For accomplishing PbD, Cavoukian [13] lays down seven principles:

1. **Proactive** not reactive; **preventative** not remedial.
2. Privacy as the **default setting**.
3. Privacy **embedded** into design.
5. End-to-end security – **full life-cycle protection**.
6. **Visibility** and **transparency** – keep it open.
7. **Respect** for user privacy – keep it **user-centric**.

These principles characterise various privacy properties rather than practical instructions [25], so that further explanation was later provided on how to operationalise PbD [14]. Some of the methods and techniques used to realise PbD are further explained in the next sections. Most interestingly, however, is that PbD inspired a great deal of research in the area of privacy engineering. And as a matter of fact, the idea of PbD was also incorporated in the EU GDPR Art. 25 Data protection by design and by default [34]. More precisely, Data Protection by Design (DPbD) – a form of PbD – requiring data controllers to “[...] implement appropriate technical and organisational measures, [...] which are designed to implement data-protection principles [...] ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed”. For simplicity, only the term PbD is used in this thesis, as well as PIA instead of Data Protection Impact Assessment (DPIA).

### 2.4.2 Privacy Impact Assessments

For someone to understand privacy, it is crucial to comprehend its technical aspects (e.g. user profiles, data flows, data holders), its security implications, and also consider particular and cultural elements around privacy in a given context. Privacy Impact Assessments (PIAs) are a pragmatic manner to perform such analysis, and arguably, one of the most well-structured methods that organisations can employ to realise PbD in their projects. Actually, PIAs are encouraged or made mandatory in various legal frameworks for privacy and data protection across countries (e.g. New Zealand, Canada, Australia, Hong Kong, European Union) [18]. Although there is no internationally accepted definition for PIA, in this thesis we consider the following two definitions:

[PIA is] “a process whereby the potential impacts and implications of proposals that involve potential privacy-invasiveness are surfaced and examined.” [18].
Or a more detailed construction:

[PIA is] “a methodology for assessing the impacts on privacy of a project, policy, programme, service, product or other initiative and, in consultation with stakeholders, for taking remedial actions as necessary in order to avoid or minimise negative impacts. A PIA is more than a tool: it is a process which should begin at the earliest possible stages, when there are still opportunities to influence the outcome of a project. It is a process that should continue until and even after the project has been deployed.” [120].

In other words, PIAs are some sort of risk analysis for privacy. Multiple technical and organisational methods are used to conduct a PIA, involving for instance: project planning, system documentation, privacy risk analysis, reporting and action plans. Once you have all the methods selected and systematised, you can create a PIA methodology. The methodologies are also commonly known as frameworks.

Many PIA frameworks have already been proposed. Some are recommended to a specific jurisdiction and legal framework, whereas others aim for a specific industry sector, or for a general methodology. A few examples of these PIA frameworks are listed as follows.

- **Sector-specific frameworks:**
  - [81, 80]: Privacy and Data Protection Impact Assessment Framework for RFID Applications (PIA RFID), 2011.
  - [33]: Data Protection Impact Assessment Template for Smart Grid and Smart Metering systems (PIA Smart Grids), Smart Grid Task Force (EU Commission), 2014.

- **General frameworks:**
  - [53]: UK – Conducting privacy impact assessments code of practice, Information Commissioner’s Office (ICO), 2014.

The PIA RFID and PIA Smart Grids are examples of sector-specific frameworks. For instance, the PIA RFID framework [81, 80] was developed by industry players and endorsed by the Article 29 Working Party[^5] [120]. Such frameworks are designed to assist organisations in assessing the potential privacy impacts of their projects and in taking appropriate action to mitigate these impacts.

approach creates a PIA template that is pertinent for a specific industry sector, which are particularly important for industries that have projects involving new technologies – of which privacy impacts are not fully understood yet.

However, the PIA RFID was later generalised in a systematic methodology [79] and it is no longer limited to RFID applications. Other well-known PIA frameworks were proposed by data protection authorities in different countries, such as the ICO’s PIA [53], the OAIC’s PIA [78], and the CNIL’s PIA [19]. And more recently, the ISO also released a standard for PIAs numbered ISO/IEC 29134:2017 [54].

Most frameworks are fairly similar, but the systematic methodology for PIAs proposed in [79] is probably the most rigorous of all. It is akin to existing standards for risk management, such as ISO/IEC 27000 series NIST Special Publications 800 series [79], which is beneficial if the PIA should be integrated into an organisation’s risk management processes. However, because this approach entails a fairly extensive privacy threat analysis it can demand more resources and time if compared to more streamlined approaches, such as the ICO’s and OAIC’s PIA frameworks.

In the context of mHealth, developers could benefit from PIA frameworks, and actually for MDCs they should be seen as mandatory due to the large array on sensitive personal data that is collected. In fact, some authors already proposed a code of conduct on privacy for mHealth applications, suggesting the creation of a PIA template for mHealth apps [71].

In general, a PIA should be considered every time a new project starts or when there are significant changes to a existing system that processes personal data (see Figure 6). It usually starts with a threshold analysis, to decide if a PIA is really necessary. If the data processing is likely to result in a high-risk to the rights and freedoms of natural persons (data subjects) (EU GDPR Art. 35(1) [34]) a PIA must be carried out by the data controller. In such case, the PIA needs to be planned and stakeholders should be consulted throughout the process. Once the plan is ready, the system’s scope and purposes should be described, and most importantly, all the processed personal data has to be mapped, using a data flow diagram. Subsequently, all the privacy threats should be identified and technical or organisational controls should be assigned to minimise, mitigate or eliminate the identified threats. Lastly, a residual risk analysis is performed, clearly stating to what extent the risks are controlled and/or accepted by the organisation. All the documentation is compiled in a final PIA Report, that is submitted to the data protection authority, and usually encouraged to be disseminated in two other versions (internal and external).

By carrying out the PIA, project managers and system designers are compelled to rethink all the personal data that is collected, and encouraged to minimise data collection to satisfy just a limited and specific set of purposes. The PIA also helps to identify privacy pitfalls before development starts and allowing changes in the early stages. Lastly, it helps to identify the technical and organisational privacy-preserving controls that should be put in place. Although organisational controls are essential for privacy, the scope of this thesis
concentrates on the technical controls. In particular, the privacy-preserving technologies for MDCSs, including PETs as well as security mechanisms.

2.4.3 Privacy-Enhancing Technologies

At this point, concrete systems concepts can be finally introduced, i.e. software components used to safeguard the right to privacy of an individual. The technical approaches for protecting privacy, usually in accordance with privacy laws, are widely known as Privacy-Enhancing Technologies (PETs). PETs are essential part of the privacy engineering field.

An early definition for PETs used the term “privacy-enhancing to refer a variety of technologies that safeguard personal privacy by minimising or eliminating the collection of identifiable data” [51]. This definition was later extended “to include security technologies to protect the confidentiality, integrity and availability of personal data.” [35]; and other authors also add, “preventing unnecessary or unwanted processing of personal data, without the loss of the functionality of the information system” [109].

PETs are also described in terms of the privacy properties that they provide, such as anonymity, pseudonymity, unlinkability, undetectability and unobservability – besides the security properties of confidentiality, integrity and availability. These privacy properties can be defined as follows [90]:

- **Anonymity** of a subject means that the subject is not identifiable within a set of subjects, the anonymity set. For instance, allow a user to use a resource or service (e.g. send or receive messages) without disclosing the user’s identity.

- **Pseudonymity** is the use of *pseudonyms* as identifier. A subject is *pseudonymous* if a pseudonym is used as identifier instead of one of its real names. For instance, to protect the user’s identity in cases where anonymity cannot be provided (e.g. if the user has to be held accountable
for his/her activities).

- **Unlinkability** of two or more items of interest (IOIs, e.g. subjects, messages, actions, ...) from an attacker’s perspective means that within the system (comprising these and possibly other items), the attacker cannot sufficiently distinguish whether these IOIs are related or not. For instance, ensures that a user may make use of resources and services without other being able to link these uses together (e.g. user sends various messages that cannot be associated).

- **Undetectability** of an IOI from an attacker’s perspective means that the attacker cannot sufficiently distinguish whether it exists or not. For instance, messages sent over a channel are not sufficiently discernible from random noise.

- **Unobservability** of an IOI (1) undetectability of the IOI against all subjects uninvolved in it and (2) anonymity of the subject(s) involved in the IOI even against the other subject(s) involved in that IOI. For instance, ensures that a user may use a resource or service without others, especially third-parties, being able to observe that the resource or service is being used (e.g. user sends a message but nobody would be able to know that).

In health care systems, anonymisation or pseudonymisation of health records is common, especially when using patient’s health data for secondary purposes. Unlinkability is also useful, for instance, in laboratory tests when the patient’s pseudonym should be unlinkable to the patient’s real name for the laboratory personnel. However, the privacy properties of undetectability and unobservability are usually impractical or even not desired in a clinical context. PETs for health care tend to emphasise aspects of confidentiality and integrity, using encryption and access control mechanisms, for privacy and safety reasons. Patients are also normally concerned about having control over their data. In fact, this is known as the “privacy as control” paradigm. Privacy as control starts from the assumption that the disclosure of personal data is inevitable in an increasingly networked world, and thus, the organisations have to establish trust through transparency and policy enforcement mechanisms [48].

PETs is an area of its own, with various categories and countless mechanisms. This section does not aim to cover PETs in a broad way, but rather focus on a few important security and privacy-enhancing mechanisms that are part of this thesis, in the context of MDCSs. First, security mechanisms, including key management, authentication, encrypted communication and storage are described. Second, data anonymisation and obfuscation schemes for sharing personal health data. Finally, consent management for obtaining and handling informed consent.
2.4.4 Security and Privacy-Enhancing Mechanisms for MDCSs

In the context of primary care, CHWs equipped with smartphones use the MDCSs in their daily activities while visiting the families in their catchment area. This application scenario has some distinct characteristics regarding security. As illustrated in Figure 7, the main aspects that need to be considered are:

- **Secure data exchange**: the data at-rest and in-transit needs to be securely stored and transmitted through the mobile network operator.

- **Light weight and low cost**: implementations should consider the limited computing power of smartphones and employ lightweight cryptography.

- **Device theft or loss**: smartphones are can be lost or stolen and the data stored in the memory cards could be compromised.

- **Network delay and disconnection**: network coverage is not always possible, so the mobile application should continue working while offline.

- **Device sharing**: smartphones are shared among CHWs, but the system should allow access only the data of the families that they are responsible for.

- **Usability**: security functions should be intuitive (or seamless) and not get in the way of the CHWs’ main tasks.

Given these aspects, some cryptographic mechanisms and protocols are particularly useful for MDCSs. Essentially, MDCSs need a Key Management Mechanism (KMM) to provide Authentication and Key Exchange (AKE)
between parties (user’s mobile and application server). Authentication protocols and key derivation schemes for such applications usually rely on symmetric cryptography using passwords. These protocols should also give support for online and offline user authentication. In addition, encryption schemes should ensure the confidentiality of at-rest an in-transit data, allowing secure communication and storage. These security concepts and building blocks are introduced in the following sections.

**Authentication and Key Derivation**

Authenticating users in the system remains a challenge in modern computer security. It is well-known that authentication can be based on a combination of factors, such as [92]: biometrics (“what the user is”), security tokens (“what the user has”), or passwords (“what the user knows”). Although there is an increasing move towards (at least) two-factor authentication schemes, secret passwords are still the most widespread strategy.

Password-based authentication is everywhere, because it is a simple, cost-effective and efficient method of maintaining a *shared secret* between a human being and a computer system. Furthermore, the advantages of using passwords tend to out shadow the disadvantages, i.e. problems of choosing strong but easy-to-remember passwords. For these reasons, it is likely that passwords will continue to be used for quite some time into the future [96]; by itself or as part of multi-factor authentication schemes.

Password-based systems normally employ Key Derivation Functions (KDFs), cryptographic algorithms for generating a pseudo-random string of bits from the password itself [99, sec. 2.4]. KDFs internally employ a one-way function (e.g. hash), so that recovering the password from the KDF’s output turns out to be computationally infeasible [27]. The output generated from a password is usually used in two ways: (a) it can be locally stored in the form of a *token* for future verification; or (b) it can be used as the *secret key* for data encryption and/or authentication. However, attackers can still attempt a dictionary attack [99, sec. 8.1] and test many different passwords combinations until a match is found (i.e. brute-force).

KDFs usually rely on two strategies for preventing such brute-force attacks. The first strategy is to deliberately raise the cost of every password guess in terms of computational resources, such as: processing time and/or memory usage. Some examples of password hashing functions that do that are: the *bcrypt*, that uses an iteration count as a cost parameter to make it slower [94]; and, the *scrypt*, that uses large amount of memory to limit large-scale custom hardware attacks [88]. The second strategy is to take as input not only the user-memorisable password, but also a sequence of random bits known as *salt* [99, sec. 3.2]. The presence of such random variable thwarts several attacks based on pre-built tables of common passwords, i.e. forces the attacker to create a new table from the scratch for every different *salt*. The *salt* can, thus,

---

*Salt* is a random string that is concatenated with a password (*salt || password*), added to the input before being passed as argument to the one-way function.
be seen as an index into a large set of possible keys derived from the password, that does not have to be memorised by users or kept in secret.

**Password-based Remote Authentication and Key Exchange**

In principle, KDFs can be used for data communication. That is, if the local and remote systems share the same password, they could exchange data by revealing to each other the salt employed for generating the key that protects such data. However, since this would allow attackers to use the same salt in an offline dictionary attack, KDFs are usually employed only for local data storage, establishing a secure channel between the human user and the local system.

Data transmission to remote locations usually employs Password Authenticated Key Exchange (PAKE) protocols. Such schemes allow two or more parties who share a password to authenticate each other and create a secure channel to protect their communication (e.g. \[8, 7\]). To be considered secure, a PAKE solutions must ensure that an unauthorised party (that fully controls the communication channel but does not know the password) is unable to learn the resulting key and is, as much as possible, unable to guess the password using offline brute force attacks. Some examples of secure PAKE are the Authenticated Key Exchange [7] and the Secure Remote Password protocol [93], which mHealth developers can leverage from and use for creating secure channels for communication.

**Secure Data Storage**

Once user and server have agreed on a common shared key by means of a PAKE protocol, this key can then be used to protect the data stored in the mobile phone. To do so, a secure storage mechanism should encrypt all the sensitive information, including: (a) configuration files and users’ data; (b) sensitive data collected from the families and temporarily stored in the smartphone; and, (c) transmitted data that is sent to the server. Hence, encryption assures data confidentiality letting only authorised parties to read data. This mechanism should employ lightweight encryption algorithms owing to constrained computing power of low-end smartphones [39], even though today many smartphones can actually perform a typical AES with key length 256-bit.

However, encryption carries the risk of rendering your data useless if anything goes wrong with the key management process (e.g. losing keys). In other words, developers should be aware that the key management adds complexity since at least on the server-side its necessary to store partial values to rebuild users’ keys in order to decrypt and to consolidate received data.

**Data Anonymisation and Obfuscation**

In the context of MDCSs (see Figure 8), data is shared among the primary care team for the primary purpose of medical treatment. Health managers also have access to the data, to perform data analysis for public health surveillance (district- or national-level). Such activities can have direct impact in the
community, by prioritising care and providing target campaigns for health promotion. However, MDCSs can also be leveraged to create rich statistical databases for researchers in health-related areas. In such cases, the collected data is being used for secondary purposes.

When using data for primary purposes, access control is the most common strategy to enforce confidentiality of patient’s information. However, we believe that besides the typical binary decision of revealing or not a data value, access control can be further improved with data obfuscation. Obfuscation is used to lower the accuracy of a specific data item in a systematic, controlled, and statistically rigorous way [5] to enhance patient’s privacy while retaining its usefulness. For instance, instead of revealing the patient’s age one can reveal a range of values. Or even, replace a medical condition or disease by a more general term (e.g. “Human Immunodeficiency Virus (HIV) infection” replaced by “Infectious Disease”).

Besides that, for secondary use of health data, anonymisation can be employed, i.e. to protect privacy by making a number of data transformations so that individuals whom the data describe remain anonymous. The anonymisation process can have variable degrees of robustness [118], depending on how likely is to: 1) single out an individual in the dataset; 2) link records concerning the same individual; or, 3) infer the value of one attribute based on other values. In essence, all these circumstances should be avoided, resulting in an anonymised dataset. Therefore, anonymised data is not considered personal data, so that, privacy and data protection laws would no longer apply.

Obtaining and Handling Informed Consent

Obtaining informed consent is common practice in health care. It is used for getting patient’s permission before conducting a clinical interventions, disclosing personal information, or enrolling a person in a clinical trial. Regardless the case, an informed consent implies that the person can give consent based on a clear appreciation and understanding of the facts. A valid informed con-
sent values the person’s autonomy and the right to be treated ethically [121]. Likewise, informed consent is also enshrined in privacy laws over the decades, because it becomes more clear that processing of personal data can also negatively impact one’s dignity. Data controllers can use informed consent as a lawful basis for processing personal data. As mentioned earlier, the EU GDPR sets a high standard for consent, stressing that it ought to be a freely given, specific, informed and unambiguous indication of the data subject’s wishes [119]. Again, emphasising the person’s autonomy, but also stressing the need to inform. Nowadays, traditional “terms & conditions” for privacy fall short in providing clear and understandable information for users. In addition, making users understand is difficult and today the consent seems to be written as much as to lawyers and policy-makers as it is to users [17].

In the context of MDCSs, data of entire communities are used to support various health care providers and government agencies. Areas of public health are exempted of obtaining consent of data subjects (Recital 54, [34]), but not everything in MDCSs is public health, e.g. research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects (Recital 33, [34]). Besides, consent also empowers data subjects, not only by raising awareness about privacy, but also allowing them to withdraw consent and stop the data processing – to some extent. Engineers have proposed new interfaces to convey adequate information to users in order to obtain informed consent, in areas such as identity management [60] and participation in mobile-mediated research [114]. On the other hand, for handling the consent data structures are being designed to capture Consent Receipts (similar to a customer receipt) in human- and machine-readable formats [65], as well as more sophisticated architectures enabling the provision of user consent as a service [105].

3 Research Question

This thesis addresses the following research question.

*How to design secure and privacy-preserving systems for Mobile Health Data Collection Systems?*

By *design* we mean the whole system design process of analysis, specification, modeling, implementation, test, deployment and evaluation of a solution. Mobile systems bring various challenges associated to limited computing capacities, vulnerabilities of wireless communication channels, and human-computer interaction. In the context of mHealth technologies, privacy and information security are paramount due to the sensitivity of health information. Although there is no one-size-fits-all solution, this thesis helps to close the security and privacy gap in MDCSs for health surveys and surveillance. Altogether, the papers included in this thesis compose a serie of steps towards answering this research question.
4 Research Method

This research can be categorised as applied science: a discipline of science that applies existing scientific knowledge to develop more (knowledge and) practical applications, such as technology and inventions. Technology is of course supported by scientific knowledge, but also “other organised knowledge to practical tasks by social systems involving people and machines” [24]. In order to answer the research question, “How to design secure and privacy-preserving systems for Mobile Health Data Collection Systems?”, the engineering Design Method [26, sec. 1.3.2] was adopted as predominant research strategy, as an alternative approach to the Scientific Method [24]. In short, the Design Method comprises eight main steps, that cyclically iterate from one to another every time assumptions should be redefined.

The Design Method starts with the (1) problem definition, basically questioning: [Who] need(s) [what] because [why]?. What is the problem? Who has the problem? Why is it important to solve? So, rather than scientific curiosity, the design is actually driven by needs of society [26].

The second step is the (2) background research, i.e. the state-of-the-art that includes scientific knowledge, but it also includes devices, components, market and economic conditions [26]. These steps correspond to the problem characterisation and observations in the scientific method.

After understanding the problem and existing solutions, the third step is to (3) specify requirements, i.e. the characteristics that your solution should meet. Requirements specification is based on what is known from other existing solutions, as well as by consulting users that need it. Again, this step would correspond to the formulation of a hypothesis, or proposing an explanation.

The fourth step is to (4) propose a solution. Researchers should brainstorm solutions within their group and choose the one that better satisfies the project goals, meeting all (or most) requirements. The fifth step, the solution is then further (5) detailed and modeled, by means of modeling languages, drawings, and so forth. This need of a model is important to predict the beha-
vior of the solution before prototyping [26]. Similarly, the scientific method needs to define the experiment and its procedures.

Once researchers have a clear idea of what to do, they start the sixth step, to (6) build a prototype. In the scientific method this is the test of the hypothesis by experimentation. The seventh step is the (7) test and redesign of the prototype by multiple iterations. Similar to the evaluation and improvement steps in the scientific method. And at last, in both methods, researchers should (8) communicate results, through technical reports, publications and documentation.

The Design Method was applied throughout the research with varying levels of completeness, as described below:

**Paper I** is a survey or review paper. A survey deals with the problem of identification, analysis and synthesis of the state-of-the-art in and specific area. In this case, to understand the current situation of mHealth initiatives in Brazil. The research is relevant to mHealth developers, academy, industry and government agencies that could benefit from mHealth solutions in similar settings. By carrying out the survey we initiated the background research process, which resulted in a well-structured review of the state-of-the-art. This background research was performed through a “ad hoc literature review”, searching various publicly available electronic documents, including scholar databases, general online literature, and scientific venues specialized on health-oriented technology. The results were communicated by means of a scientific publication. (The remaining steps of the Design Method are not required for survey papers.)

**Paper II** describes a security framework for MDCS, named SecourHealth. The initial problem referred to the design of a security framework for a MDCS that was to be implemented in the city of São Paulo (Brazil). Relevant publications were found during our ad hoc background research, from which solutions could be however improved and/or adapted to our settings. A new solution was proposed, modeled, prototyped and tested in order to demonstrate its feasibility. The results were communicated by means of a scientific publication.

**Paper III** describes a georeferenced and secure MDCS, named GeoHealth. The problem refers to the design of a MDCS that could support public primary health care in the city of São Paulo. A specific background research revealed that existing solution could not meet all desired requirements (e.g. health data quality, security, georeferencing). A new MDCS was therefore proposed, modeled, prototyped and tested in order to demonstrate its feasibility. The results were communicated by means of a scientific publication.

**Papers IV and V** present a Privacy Impact Assessment (PIA) for GeoHealth. PIA as a methodology, supports a in-depth analysis of the problem of privacy in MDCSs. It helped to not only identify the privacy threats
but also identify the technical and organisational controls to mitigate privacy risks. This PI A for GeoHealth also partially contributes to the background research of the thesis, but it is also a proposed solution for relieving privacy by design. The results are communicated in the form of scientific publications.

**Paper VI** outlines a preliminary ontology-based data sharing system (O-DSS) for medical information. The problem refers to a solution to transfer and share individuals’ health information in a privacy-preserving manner, by exploiting ontology-based obfuscation and anonymisation functions. The specific background research shows that although there exist viable solutions to be used, they should be linked to realistic use cases and adapted accordingly. An O-DSS was therefore proposed and exemplified with use cases, yet it should be still further developed. Partial results were communicated by means of a scientific publication.

**Paper VII** addresses the problem of electronic consent (e-Consent) for MDCSs. The background research reveals many approaches for obtaining and handling informed consent, but they need to be adapted and integrated to MDCSs. A user interface is therefore proposed, modeled, prototyped and tested to evaluate its usability; and a data structure is also adopted for storing and managing consent. For the usability evaluation of the proposed e-Consent tool, a mixed-methods approach of cognitive walk-through, questionnaire and interviews with experts on the field was used. Thus, enabling a qualitative evaluation on the principles of informed consent (i.e. information disclosure, comprehension, voluntariness and agreement). The results are communicated in a scientific publication.

## 5 Contributions

As a result of the research, this thesis adds to the body of knowledge of security and privacy for mHealth systems with three overarching contributions:

- Analysis of mHealth ecosystem in Brazil providing a comprehensive review of existing initiatives and evidence about the lack of security in most projects (Paper I).

- Sharing experience in developing and deploying a secure MDCS in different communities and in large-scale research experiment (Paper III).

- Analyses and proposals of security and privacy-enhancing mechanisms for MDCSs (Papers II-IV-V-VI-VII).

These overarching contributions comprise multiple partial contributions made in Papers I-VII, which are illustrated in Figure 10 and listed as follows.

1. *Analysis of the Mobile Health Ecosystem in Brazil*. An in-depth analysis about mHealth initiatives in Brazil is provided (Paper I). Review papers
help researchers to understand current front-runners, target users, types of health applications, adopted devices, and security problems in existing proposals. In addition, it helps to understand the impact of mHealth solutions in primary care settings, potential nation-wide projects, business opportunities and potential research areas. And in particular, it evidences the security gap in mHealth, whereas many solutions (in Paper I) have shown little or no concern about protecting collected and processed data. Such results allow better understanding of the mHealth scenario, and most importantly, further motivate the main research question of the thesis.

2. **Sharing Experience in Developing and Deploying MDCSs for Primary Care.** GeoHealth is proposed as a secure, low-cost and high-impact MDCS (Paper III). In this research the experience of designing and deploying a MDCS is shared, showing that such systems can significantly improve the efficiency and quality of the entire process of health surveys and surveillance in the primary care. Furthermore, GeoHealth stands out from other systems for having strong security features, which were crucial for deploying it in a large-scale. Designing and deploying GeoHealth, partly answers the research question, by describing the entire software development process of a MDCS and the integration of security features, as well as enabling firsthand experience with many stakeholders (e.g. CHWs, health managers, developers, families).

3. **Privacy Analysis of MDCS and Proposal of Security and Privacy-Enhancing Mechanisms.** In order to further understand the privacy issues in MDCSs, a PIA is carried out using the GeoHealth as a study case (Papers IV and V). The PIA comprises the characterisation of the system, data flows of personal information, threat analysis and identification of controls. A range of technical mechanisms for privacy and data protection are introduced, namely: (a) the SecourHealth, a delay-tolerant security framework designed specifically for MDCSs (Paper II); (b) the Ontology-
based Data Sharing System (O-DSS) for enforcing anonymisation and obfuscation of health data (Paper VI); and, (c) an application for obtaining and handling informed consent within the context of CBPHC (Paper VII). These results, also partly answer the research question, toward further understanding the privacy and security issues and providing technical controls.

5.1 Limitations

In spite of the contributions, we would like to remark some limitations of our work. Firstly, the research question presents a complex multi-faceted problem, and even though the problem can be seen as solved, some dimensions or aspects of it remain to be addressed. The vast majority of privacy aspects relevant to MDCSs were identified, but not all of them were addressed in detail. The focus of this thesis was put on technical privacy controls and thus non-technical privacy controls were not discussed in the thesis with the same emphasis and rigour. In fact, the proposed controls, which we elaborated in more detail, are essentially technical, addressing security, anonymisation, and consent management. Non-technical measures still to complement these technical controls and need to be further detailed in future.

Secondly, regarding privacy law, this thesis considers mainly the European GDPR [34] and Brazilian LGPD [10]. Researchers and developers working with other MDCSs used in different countries should be attentive to the regulations and privacy laws that apply in their jurisdiction. In addition, although MDCSs share various functions, every country has it is particular programme for primary health care, and stakeholders may change as well as the purpose specification for data processing. Thus, potentially adding new assumptions and considerations to the privacy analysis.

Finally, among the proposals, the O-DSS was not prototyped and tested, owing that just its conceptual architecture has been presented. The e-Consent prototype should preferable be further developed and tested with health workers and data subjects, in order to evaluate its usability and effectiveness in the field.

6 Related Work

This section reviews the state-of-the-art associated to the thesis. Related literature and existing solutions are discussed and (if possible) compared to the publications listed in Section 7.

6.1 Mobile Health for LMICs

Along the years, various reports have been published on eHealth and mHealth initiatives in LMICs. Some remarkable publications were made by international organisations, such as: United Nations Foundation [21], Earth Institute
[74], and World Health Organization [86]. At the same time, a few nation-specific reviews were performed for countries such as India [38] and China [64]. Taken together, global and local efforts were consolidating “mHealth for LMICs” as an emerging field of research.

However, there was still no comprehensive surveys for Brazil, so that the mHealth ecosystem remained to be investigated. The work presented in Paper I contributes with an in-depth analysis of the Brazilian initiatives on mHealth. As a result, the survey provides a critical analysis of mHealth projects and stakeholders, and also sharing knowledge that was originally only available in Brazilian Portuguese.

Apart from that, researchers have made other valuable contributions in the past years. A systematic review of mHealth projects provides a critical analysis about what works (or not) and why in Africa [3]. And other authors also made a more general review of the state-of-the-art in mHealth [15]. Altogether, both are essential readings for researchers, policy makers, project managers and developers working with mHealth in LMICs.

### 6.2 Security for MDCSs

Although much has been published about mHealth security, there are not many papers that address MDCSs specifically. Requirements such as offline authentication and delay-tolerant transmission of data make MDCSs particularly challenging. Classic approaches such as HTTP over TLS require persistent connectivity, and handshakes for establishing keys are often onerous for low-speed 2G and 3G mobile phone networks. In addition, developers also tend to avoid solutions that rely on public-keys and certificates, because of the additional infrastructure costs.

These requirements of MDCSs led researchers to propose tailored security solutions. Prominent contributions were made by a research group at Bergen University [41, 40, 69, 70]. In this series of works, the authors proposed a security framework for MDCSs covering: user and server authentication, secure data storage, and secure communication. Their solution was integrated in open-source projects, such as openXdata⁷ and Open Data Kit (ODK)⁸. However, the differences between their proposal and the SecourHealth framework (Paper II), are: (a) forward secrecy is added to stored data, and (b) the key management for mutual authentication and data exchange is simpler. SecourHealth was also implemented within the GeoHealth MDCS, deployed in large-scale in a pioneering project in the city of São Paulo.

### 6.3 Design and Deployment of MDCSs

As already mentioned, Brazil’s Department of Primary Care (DAB) controls the Health Information System for Primary Care (SISAB), aggregating data from all municipalities in the country. One of DAB’s strategy is the e-SUS

---


⁸[https://opendatatool.org/](https://opendatatool.org/)
Atenção Básica (e-SUS AB, Primary Care) as part of the SISAB. Today, all forms used in the Brazil’s CBPHC are specified by the DAB, both in paper and electronic formats, streamlining communication with SISAB.

However, various MDCSs were developed before the e-SUS AB became a reality. Most systems were developed by researchers in projects like Colibri [91] and Borboleta [23, 30]. Such solutions were mainly focused on replacing paper-based forms with electronic ones and small-scale prototypes, leading to the following limitations: (1) they lack support for remote data communication, obliging users to synchronise the collected data only when inside the health units; (2) they do not provide strong security mechanisms for protecting the data stored in the device; and (3) they have no feature for dealing with families having no formal address. These are probably among the reasons why, to the best of our knowledge, they have never been broadly adopted in practice.

The proposed MDCS (GeoHealth, Paper III) copes with these limitations, by providing: support for offline and online data collection; protection of data in-transit and at-rest; and georeferencing data (i.e. GPS for locating families). Moreover, GeoHealth was deployed and tested in a large-scale (total of 28,324 families/96,061 people), proving to be a feasible and low-cost solution (approx. monthly cost of USD 0.04 per inhabitant). Now the DAB itself and other private companies have released their MDCSs. At last corroborating previous work on MDCSs conducted not only in Brazil but globally with positive results for CBPHC.

6.4 Privacy Impact Assessment for MDCSs

Although MDCSs are essential for PHC, they are also inherently privacy-invasive technologies that allow surveillance of entire communities. If the general public and health care providers cannot trust the system they will not use it. Privacy incidents in relation to health data can have severe implications for both, data subjects (e.g. discrimination, reputation damage) and health professional (e.g. prosecution and dismissal). If organisations fail to address the privacy issues they might face lawsuits, financial penalties, public backlash and damage to their reputation. It is therefore critical to conduct Privacy Impact Assessments (PIA) for such systems.

For this reason, a PIA was carried out using GeoHealth as main study case (Paper IV and V). To the best of our knowledge, there is no related work on PIAs for MDCSs yet. In fact, it is difficult to find PIA reports even for mHealth systems in general, possibly because such reports yield negative results, i.e. highlighting privacy issues in the system.

GeoHealth’s PIA is based on the PIA RFID framework [79]. The privacy threat analysis and identification of controls were derived from relevant literature on mHealth privacy [61] as well as the original threats and controls from the PIA RFID [80, 81]. As main contributions, this PIA compiles a exhaustive list of privacy threats and controls for MDCSs. Furthermore, it offers a concrete example for mHealth project managers and developers regarding the
engineering of privacy in their systems.

6.5 Obfuscation and Anonymisation of Health Data

This thesis also investigates how to use ontologies for handling textual data values (e.g. diseases, medical procedures, drugs) to decrease semantic loss along the obfuscation and anonymisation process. Ontology-based obfuscation and anonymisation can be used in wide range of systems to enforce fine-grained access control and data minimisation. However, in this thesis only use cases in health care were investigated (in Paper VI). MDCSs is one example, in which health data used for secondary purposes can be obfuscated or anonymised before sharing it (e.g. research and statistics).

The conceptual architecture for the Ontology-based Data Sharing System (O-DSS) was inspired by two main proposals: the ontology-based anonymisation for electronic health records (EHRs) [72, 73]; and, the data obfuscation based on user context [117]. As a result, we exemplified how to use obfuscation functions in the Peer Manager\(^9\) platform [50], defining them as obligations in a privacy policy. In addition, this mechanism could be implemented in health systems (e.g. E/PHRs) leveraging from real medical ontologies (e.g. SNOMED-CT\(^{10}\)).

6.6 Consent Management Systems

As aforementioned, consent is critical for personal privacy, but obtaining truly informed consent is not trivial. Making users understand the nature of processing of personal data is difficult. Worse still, current practices in the industry with ill-defined purposes and excessive collection of personal data just make this task harder. Thus, asking the user’s informed consent before collection helps to balance this problem of information asymmetry. To the best of our knowledge, there are no “e-Consent” solutions for MDCSs nor for other mHealth systems in primary care or public health surveillance yet.

Considering this, we contribute with the design of an e-Consent mechanism for obtaining and handling consent, inspired in the Participant-Centered Consent (PCC) toolkit and Kantara’s Consent Receipt specification. The PCC toolkit defines the process and the interfaces for electronic informed consent (e-Consent), used for app-mediated research studies. [28, 114]. Apple’s ResearchKit\(^{11}\) adopted this toolkit in 2015 and it has been tested in multiple projects [29, 76]. Apart from obtaining consent, organisations also have the responsibility of managing and archiving it. The burden of proof falls under

\(^9\)The Peer Manager works as an user-centered identity management platform that keeps user’s information private. This framework was built upon the privacy policy language PPL (PrimeLife Policy Language), with which every user can control his personal information by imposing access and usage control restrictions. The Peer Manager is part of the SmartSociety research project (http://smart-society-project.eu/).

\(^{10}\)Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT).

\(^{11}\)ResearchKit is an open source framework introduced by Apple that allows researchers and developers to create powerful apps for medical research. (http://researchkit.org/)
the data controller. With that in mind, the Kantara initiative has proposed the Consent Receipt [65], as a data structure to capture the minimum viable data for a consent. This receipt can be then stored in human- and machine-readable formats.

The proposed e-Consent tool was also evaluated in regards to its usability and adherence to principles of informed consent [36]. Namely, evaluating user perceptions on information disclosure, comprehension, voluntariness and agreement; in the pursuit of obtaining truly informed and valid consent.

Nonetheless, more sophisticated systems have been proposed for handling and enforcing consent (i.e. authorisation). Most notably, the MyData initiative [105] and the ForgeRock’s Identity Platform12, both using User-Managed Access (UMA) [68], a protocol standard based on OAuth13. Authorisation problems are however outside the scope of this thesis.

7 Summary of Appended Papers

Paper I – Mobile Health in Emerging Countries: A Survey of Research Initiatives in Brazil

Mobile health consists basically in the application of mobile devices and communication capabilities for expanding the coverage and improving the effectiveness of health care programs. This technology is particularly promising for developing countries, in which health authorities can take advantage of the flourishing mobile technology market to bring adequate health care to underserved or underserved communities. Specifically, mHealth can effectively improve basic care and help combating endemic diseases not so often encountered in developed countries. This huge potential has lead to intensive research efforts not only in emerging countries and also around the world, creating a number of innovative solutions. In this paper we provide a comprehensive survey of mHealth research initiatives developed specifically for tackling health challenges in Brazil, an emerging country with a flourishing mobile market. This study identifies the main providers of solution, the areas of deployment, the health conditions that are focus of attention, the types of devices used, the target users, the (lack of) attention to data security issues, among others. Our goal is to discuss gaps, opportunities and tendencies observed in the country, giving some insight on the challenges faced by the mHealth technology in similar scenarios.

Paper II – SecourHealth: A Delay-Tolerant Security Framework for Mobile Health Data Collection

Security is one of the most imperative requirements for the success of systems that deal with highly sensitive data, such as medical information. However,

12ForgeRock’s Identity Platform (https://www.forgerock.com/platform)
13OAuth, is an open protocol to allow secure authorisation in a simple and standard method from web, mobile and desktop applications. (https://oauth.net)
many existing mobile health solutions focused on collecting patients’ data at their homes that do not include security among their main requirements. Aiming to tackle this issue, this paper presents SecourHealth, a lightweight security framework focused on highly sensitive data collection applications. SecourHealth provides many security services for both stored and in-transit data, displaying interesting features such as tolerance to lack of connectivity (a common issue when promoting health in remote locations) and the ability to protect data even if the device is lost/stolen or shared by different data collection agents. Together with the system’s description and analysis, we also show how SecourHealth can be integrated into a real data collection solution currently deployed in the city of São Paulo, Brazil.

**Paper III – Georeferenced and Secure Mobile Health System for Large Scale Data Collection in Primary Care**

The Primary Care Information System (SIAB) concentrates basic health care information from all regions of Brazil, providing a rich database for health-related action planning. This data is collected by Family Health Teams (FHTs) in periodical visits to enrolled families in targeted areas. The fact that this procedure relies on paper forms, however, degrades the quality of the information provided to health care authorities and slows down the process of decision making. Aiming to overcome such issues, this article describes GeoHealth, a data gathering application that allows FHTs to use a 3G- and GPS-enabled smartphone for collecting the families’ data. Besides quick data validation and delivery, GeoHealth provides strong security features and allows more data to be collected (e.g., the precise location of families having no formal address and extra fields not present in standardized paper forms). We discuss the system’s deployment at 6 primary care units in the city of São Paulo, where a total of 33,675 families are regularly surveyed. The results obtained show that the process is a low-cost and interesting approach for primary care data collection and analysis.

**Paper IV – mHealth: A Privacy Threat Analysis for Public Health Surveillance Systems**

Community Health Workers (CHWs) have been using Mobile Health Data Collection Systems (MDCSs) for supporting the delivery of primary healthcare and carrying out public health surveys, feeding national-level databases with families’ personal data. Such systems are used for public surveillance and to manage sensitive data (i.e., health data), so addressing the privacy issues is crucial for successfully deploying MDCSs. In this paper we present a comprehensive privacy threat analysis for MDCSs, discuss the privacy challenges and provide recommendations that are specially useful to health managers and developers. We ground our analysis on a large-scale MDCS used for primary care (GeoHealth) and a well-known Privacy Impact Assessment (PIA) methodology. The threat analysis is based on a compilation of relevant privacy threats from the literature as well as brain-storming sessions with privacy and security
experts. Among the main findings, we observe that existing MDCSs do not employ adequate controls for achieving transparency and interevenability. Thus, threatening fundamental privacy principles regarded as data quality, right to access and right to object. Furthermore, it is noticeable that although there has been significant research to deal with data security issues, the attention with privacy in its multiple dimensions is prominently lacking.

**Paper V – Mobile Health Systems for Community-Based Primary Care: Identifying Controls and Mitigating Privacy Threats**

This research expands the PIA presented in Paper IV. It describes the entire process of the GeoHealth’s PIA, but emphasizing controls and mitigation strategies to handle negative privacy impacts. Extensive documentation is also provided in the form of appendices, in order to ensure the research reproducibility. All the PIA steps were based on discussions among the researchers (privacy and security experts), and in particular, the identification of threats and controls was based on literature reviews and working group meetings among the group. Moreover, we also received feedback from specialists in primary care and software developers of other similar MDCSs in Brazil. The GeoHealth PIA is based on 8 Privacy Principles and 26 Privacy Targets derived from the European General Data Protection Regulation (EU GDPR). Associated with that, 22 threat groups with a total of 97 sub-threats and 41 recommended controls were identified. Among the main findings, we observe that privacy principles can be enhanced on existing MDCSs with controls for managing consent, transparency, intervenability and data minimisation.

**Paper VI – Ontology-based Obfuscation and Anonymisation for Privacy: A Case Study on Healthcare**

Healthcare Information Systems typically fall into the group of systems in which the need of data sharing conflicts with the privacy. A myriad of these systems have to, however, constantly communicate among each other. One of the ways to address the dilemma between data sharing and privacy is to use data obfuscation by lowering data accuracy to guarantee patient’s privacy while retaining its usefulness. Even though many obfuscation methods are able to handle numerical values, the obfuscation of non-numerical values (e.g., textual information) is not as trivial, yet extremely important to preserve data utility along the process. In this paper, we preliminary investigate how to exploit ontologies to create obfuscation mechanism for releasing personal and electronic health records (PHR and EHR) to selected audiences with different degrees of obfuscation. Data minimisation and access control should be supported to enforce different actors, e.g., doctors, nurses and managers, will get access to no more information than needed for their tasks. Besides that, ontology-based obfuscation can also be used for the particular case of data anonymisation. In such case, the obfuscation has to comply with a specific criteria to provide anonymity, so that the data set could be safely released. This research contributes to: state the problems in the area; review related
privacy and data protection legal requirements; discuss ontology-based obfuscation and anonymisation methods; and define relevant healthcare use cases. As a result, we present the early concept of our Ontology-based Data Sharing Service (O-DSS) that enforces patient’s privacy by means of obfuscation and anonymisation functions.

Paper VII – E-Consent for Data Privacy: Consent Management for Mobile Health Technologies in Public Health Surveys and Disease Surveillance

Community health workers in primary care programs increasingly use Mobile Health Data Collection Systems (MDCSs) to report their activities and conduct health surveys, replacing paper-based approaches. The mHealth systems are inherently privacy invasive, thus informing individuals and obtaining their consent is important to protect their right to privacy. In this paper, we introduce an e-Consent tool tailored for MDCSs. It is developed based on the requirement analysis of consent management for data privacy and built upon the solutions of Participant-Centered Consent toolkit and Consent Receipt specification. The e-Consent solution has been evaluated in a usability study. The study results show that the design is useful for informing individuals on the nature of data processing, privacy and protection and allowing them to make informed decisions.

8 Conclusions and Future Work

Initiatives of mHealth can truly revolutionise the delivery of health care in LMICs. Particularly in middle-income countries, researchers and government can benefit from favorable economic conditions and reliable network infrastructure, unlocking mHealth to its full potential. This thesis contributed for describing the mHealth ecosystem in Brazil as well as tackling the privacy and security problems, health care imperatives for achieving trust. Nation-wide MDCSs were the main target our research, considering that such systems have tremendous impact in Brazil’s PHC by empowering CHWs.

Notwithstanding, dealing with privacy as a broad legal concept requires careful analysis of existing laws. Based on the privacy principles enshrined in legal frameworks, it is required from practitioners to interpret juridical concepts into the system’s design. There is however a considerable knowledge gap in the mHealth community regarding how to actually engineer privacy into the systems. Given that, this research offers concrete examples of privacy-preserving mechanisms to solve identified issues, and thus, helps project managers and developers with such translations from policy to engineering. As a result, mHealth initiatives have a higher probability of being widely deployed and scaling-up.

Privacy engineering needs to be broadly adopted within the Mobile for Development (M4D) community. Various documents indeed already state principles for digital development that put privacy and security as critical pil-
lars for sustainable solutions. Project leaders and practitioners, in consultation with privacy engineers, should further examine the existing solutions and adapt them to their reality. Privacy issues are specially pressing in the context of mHealth, because most of these systems in LMICs extend the health care to vulnerable parts of the population, in which the violations of privacy rights can cause embarrassment and discrimination, at greater detriment of one’s dignity.

In summary, this thesis confirms previous research, supporting the claim that MDCSs are simple and effective solutions for improving the delivery of CBPHC. Moreover, it further analyses privacy and security in MDCSs, stressing the need and introducing new feasible solutions. Here we explained how to: (a) carry out a PIA for MDCSs, enabling stakeholders to surface all the privacy-invasive characteristics of the systems, derive privacy principles from relevant legal frameworks, identify privacy threats and propose controls to mitigate risks; (b) design security mechanisms for password-based authentication and key exchange, encrypted storage and communication, that are sufficiently lightweight for low-budget resource settings; (c) design obfuscation and anonymisation functions to enforce data minimisation, allowing fine-grained access control during disclosure of data for primary purposes as well as complete de-identification when sharing data for secondary purposes; and (d) design consent management systems to obtain and handle individuals’ informed consent, emphasising the need of making users understand the nature of data processing, and utilising simple data structures for capturing the consent.

Security and privacy do not have to be complicated black-boxes, and besides, various controls are actually not even technical but organisational measures for enhancing system’s transparency and openness. New regulations, such as the EU GDPR and BR LGPD, provide us the state-of-the-art on privacy law and should be recognised widely. It is an opportunity to call for Privacy for Development (P4D) in the ICT4D and M4D communities, whereas the respect for one’s privacy is indispensable for real development.

In this regard, other research challenges on privacy were identified. Future work on privacy for MDCSs concerns transparency and intervenability. Transparency of processing activities can be achieved by providing data subjects adequate information about the systems, and making this information easy to access and to understand. A range of transparency-enhancing tools could be also be adopted, giving users insights on privacy policies, data processing activities, disclosures to third-parties, tracking algorithms, or data controller’s reputation regarding privacy.

Additionally, intervenability refers to multiple privacy rights, such as the right to access, to object, to be forgotten, to correct, to withdraw consent and to data portability. Regarding the right to access, data subjects should have access to their data, and to see what data was collected, how it is processed and to whom it is shared; and also providing electronic copies of the one’s data and supporting data portability. On the right to object, data subjects should have the ability to object to the processing of personal data, as well as ensuring the
right to delete or block data; and including the right to be forgotten. Finally, when consent is used as a lawful basis for data processing, data subjects should also have the right to withdraw consent, as easily as it was to give consent.

Transparency and intervenability are challenging because currently MDCSs are designed to be used only by health care providers. Adding an interface for all the data subjects incurs in higher costs for developing and maintaining the system. Nonetheless, patient’s access is something that is already reality for more traditional electronic health records systems – at least in high-income nations – and could be implemented in MDCSs in a similar fashion.

Finally, some broader implications for privacy in mHealth applications remain to be understood. Countries of all income levels have reported initiatives and projects on mHealth during the past decades. The digitisation of health care already entails serious privacy considerations, but for some new technologies it is still hard to predict the possible impacts of privacy and data protection. For instance, areas such as big data and artificial intelligence, in which research can greatly benefit public health and clinical practice, at the expense of increasing people’s profiling, automated-decision making, and indiscriminate or excessive processing of personal data. Therefore, innovative mHealth systems that leverage from such new technologies should carefully consider the impacts on privacy, or for that matter respect to people’s rights and freedoms.

References


[90] Andreas Pfitzmann and Marit Hansen. A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management. http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf, August 2010. v0.34.


Engineering Privacy for Mobile Health Data Collection Systems in the Primary Care

Mobile health (mHealth) systems empower Community Health Workers (CHWs) around the world, by supporting the provisioning of Community-Based Primary Health Care (CBPHC). In particular, Mobile Health Data Collection Systems (MDCSs) are used by CHWs to collect health-related data about the families that they treat, replacing paper-based approaches. Although MDCSs improve the efficiency of CBPHC, existing solutions lack adequate privacy and security safeguards.

To bridge this knowledge gap between the research areas of mHealth and privacy, we start by asking: How to design secure and privacy-preserving systems for Mobile Health Data Collection Systems? To answer this question, an engineering approach is chosen to analyse and design privacy and security mechanisms for MDCSs.

Among the main contributions, a comprehensive literature review of the Brazilian mHealth ecosystem is presented. On the privacy engineering side, the contributions are a Privacy Impact Assessment (PIA) for the GeoHealth MDCS and three mechanisms: SecourHealth, a security framework for data encryption and user authentication; an Ontology-based Data Sharing System (O-DSS) that provides obfuscation and anonymisation functions; and, an electronic consent (e-Consent) tool for obtaining and handling informed consent.
ENGINEERING PRIVACY FOR MOBILE HEALTH DATA COLLECTION SYSTEMS IN THE PRIMARY CARE

Mobile health (mHealth) systems empower Community Health Workers (CHWs) around the world, by supporting the provisioning of Community-Based Primary Health Care (CBPHC). In particular, Mobile Health Data Collection Systems (MDCSs) are used by CHWs to collect health-related data about the families that they treat, replacing paper-based approaches. Although MDCSs improve the efficiency of CBPHC, existing solutions lack adequate privacy and security safeguards.

To bridge this knowledge gap between the research areas of mHealth and privacy, we start by asking: How to design secure and privacy-preserving systems for Mobile Health Data Collection Systems? To answer this question, an engineering approach is chosen to analyse and design privacy and security mechanisms for MDCSs.

Among the main contributions, a comprehensive literature review of the Brazilian mHealth ecosystem is presented. On the privacy engineering side, the contributions are a Privacy Impact Assessment (PIA) for the GeoHealth MDCS and three mechanisms: SecourHealth, a security framework for data encryption and user authentication; an Ontology-based Data Sharing System (O-DSS) that provides obfuscation and anonymisation functions; and, an electronic consent (e-Consent) tool for obtaining and handling informed consent.